Design of Nuclear-Targeting Peptides for Macromolecule Delivery via Machine Learning

by

Carly Katherine Schissel

B.A., Williams College (2016)

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 2022

Signature of Author……
Department of Chemistry
January 15, 2022

Certified by……
Bradley L. Pentelute
Professor of Chemistry
Thesis Supervisor

Accepted by …….
Adam P. Willard
Associate Professor of Chemistry
Graduate Officer

This doctoral thesis has been examined by a committee of the
Department of Chemistry as follows:

Ronald T. Raines……
Thesis Committee Chair
Roger and Georges Firmenich Professor of Chemistry

Bradley L. Pentelute……...
Thesis Supervisor
Professor of Chemistry

Laura L. Kiessling……..
Thesis Committee Member
Novartis Professor of Chemistry

Design of Nuclear-Targeting Peptides for Macromolecule Delivery via Machine Learning

by

Carly Katherine Schissel

Submitted to the Department of Chemistry on January 15, 2022
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Chemistry

Abstract

The effective design of functional peptide sequences remains a fundamental challenge in biomedicine. For example, cell-penetrating peptides (CPPs) are capable of delivering macromolecular cargo to intracellular targets that are otherwise inaccessible. However, design of novel CPPs with high activity and unique structure remains challenging. In this thesis, methods to design and characterize highly active CPPs for antisense oligonucleotide delivery were explored.

Machine learning is a promising method for de novo design of functional peptide sequences. A deep learning model inspired by directed evolution was used to optimize abiotic sequences that traffic antisense oligomers to the nucleus of cells. The model was able to predict activities beyond those in the training dataset, and simultaneously decipher and visualize sequence-activity predictions. The validated miniproteins (40-80 residues) were more effective than any previously known variant in cells. By augmenting the machine learning model to over-represent shorter sequence space, the model also predicted a short peptide (18-residues) with comparable activity to a positive control peptide. Empirical sequence-activity studies demonstrated reliance on the cationic residues as well as the C-terminal cysteine residue. These sequences were nontoxic, able to deliver other biomacromolecules to the cytosol, and efficiently delivered antisense cargo in mice.

A different approach to discover and characterize CPP sequences was also taken, by extracting peptides taken up into cells and analyzing their relative quantities or identifying their sequences by mass spectrometry. First, several mirror-image D-peptides had similar delivery activity to their native forms, while demonstrating complete proteolytic stability. Mixtures of fully intact antisense-peptide conjugates could be recovered from whole cell and cytosolic lysates, and relative concentrations were quantified by MALDI-TOF. This method was then extended to the discovery of de novo sequences from a combinatorial library of antisense-peptide conjugates containing unnatural residues. Following cell treatment with the biotinylated antisense-peptide library, the cytosol of cells was extracted and internalized peptides recovered via affinity capture. De novo sequencing was achieved by Orbitrap tandem mass spectrometry, and several unique, unnatural sequences were identified that could effectively deliver the antisense oligomer to the nucleus.

In summary, machine learning and mass spectrometry-based strategies to discover and characterize novel CPP sequences for antisense delivery were explored. In the future, we envision combining these methods in order to use lists of library hits to train a machine learning model to design sequences composed of fully unnatural amino acids.

Thesis Supervisor: Bradley L. Pentelute
Professor of Chemistry

Acknowledgements

Graduate school has been a unique odyssey. While I am preparing to see Massachusetts off in the rear-view mirror, I'm fortunate to fondly reminisce about all for which I feel grateful.

First and foremost I would like to thank my advisor Prof. Brad Pentelute for his encouragement and support over these past five years. His boundless enthusiasm for science was a great motivation to me in the early days, and his say-yes attitude always encouraged me to try new things and think in new ways. Additional thanks to Dr. Andrei Loas, who has reviewed countless pieces of writing and has been a great support and friend in our lab. Many thanks as well to Emily Wensberg, Rachael Fuller, Dr. Jennifer Weisman, and Kathy Sweeney who through the years have kept the gears of our lab and the department turning.

I also would like to thank Prof. Ronald Raines for providing guidance as my Thesis Chair. His feedback encouraged me to think about my work from a different perspective, and I'm immensely grateful for his career support. Many thanks as well to Prof. Laura Kiessling for serving on my committee and for her always thoughtful questions and guidance during my qualifying exams and defense. I would also like to thank my Professors at Williams College for their boundless encouragement and belief in me: Prof. Tom Smith, Prof. David Richardson, Prof. Jay Thoman, Prof. Patrick Barber, Prof. Gehring, I am so truly luck to have been taught by you in the classroom, research lab, and glass lab.

I have had the pleasure of working with many amazing scientists during my time here, and am incredibly grateful for the number of high-quality collaborations I've engaged in. I would like to thank Somesh Mohapatra and Prof. Rafa Gomez-Bombarelli, whose computational expertise made possible the machine learning-guided design described herein. Special thanks to Charlotte Farquhar who, in addition to being a great friend, made the mad dash to the finish line with me to put together the final story of this thesis. Also many thanks to the Sarepta team: Dr. Eva Maria Lopez Vidal, Dr. Xuyu Tan, Dr. Nicholas Vecchiarello, Dr. Rebecca Holden, Dr. Justin Wolfe, Dr. Colin Fadzen, Dr. Annika Malmberg, and Dr. Gunnar Hanson; as well as my other excellent collaborators in the lab: Dr. Sebastian Pomplun, Dr. Muhammad Jbara, Dr. Chris Shugrue, and Dr. Nina Hartrampf.

My labmates and collaborators have become my dear friends. Dr. Alex Loftis, Dr. Anthony Quartararo, Dr. Aaron Mallek, and Dr. Suan Tuang—Your humor, laughter, and energy made each day a joy, and I've so missed our daily lunches and coffee hangouts in the last year or so since you've graduated. However I wasn't without company, as newer members including Charlotte Farquhar, Dr. Coralie Backlund, Dr. Joe Brown, Azin Saebi, Dio Dieppa-Matos, Sarah Antilla, Michael Lee, Amanda Cowfer, and Xiyun Ye have been wonderful friends. The Pentelute Lab summer volleyball team are the true champions, regardless of this last year's result. Our dynamic on the court mirrored our interactions in the lab: communicative, fun, and with excellent teamwork and sportsmanship, and I was lucky getting to play with everyone: Sarah, Dio, Azin, Ed, Cameron, Aaron, Hannah, Rebecca, Coralie, Lee, Anthony, Ethan, and Justin. I am so grateful to have been able to work with others in the Pentelute Lab, past and present: Dr. John Albin, Alex Callahan, Dr. Nathalie Grob, Dr. Heemal Dhanjee, Dr. Chengxi Li, Jacob Rodriguez, Dr. Jason Tao, Dr. Genwei Zhang, Xiyun Ye, Dr. Mette Ishoey, Dr. Alex Mijalis, Mackenzie Poskus, Stephanie Hanna, Bente Somsen, Jessica Wilson, Dr. Chi Zhang, and Vanessa Zuger. Shoutout also to my friend Dr. Mason Wu—can't wait to see you in the Bay. Extra special thanks to my favorite collaborator, Dr. Kyan D'Angelo—Here's to more collaborations!

Sending love and thanks to my high school and college friends, who I've miraculously stayed in close contact with after all these years; Katie & Megan, I'm sorry to be leaving Boston so soon but I still can't believe we lived out our promise of living here together. Sonia, Kimberly, Cassy, Lydia, Sibo, and Jessica, the crew stands strong 10 years later and I can't wait to live a bit closer to most of you (and see the others along our road trip!). Kathy, Irene, Dylan, I'm so grateful for your friendship and feel lucky to have endured graduate school at the same time as you three.

Perhaps more than any other element of graduate school, the MIT Women's Volleyball Club has been a true blessing. Although I know he would deny it, Dr. Tony Lee, aided by Dr. Max Wolf, has facilitated the creation of the most welcoming, hardworking, and supportive group of women volleyball players that I've ever met. This group helped me grow on and off the court, learning teamwork, leadership, and how to have unconditional fun. I've also made amazing lifelong friends with those who I played with the most: Cathy Melnikow, Hannah Jacobs, Laura Li, Laura Huang, Amy Leung, Lauren Clarke, Colet Eggermont-te Grotenhuis, Melinda Szabo, Sheri Grill, Heidi Li, Sarah Antilla, Abby Bertics, Brooke Huisman, Darcy Duke, Vera Schroeder—can't wait to get on the court with you again soon!

Finally, I would like to thank my family, whose love and support keeps lifting me up higher than I imagine. To my Dad, ("*No matter where you go, there you are*") thank you for being my steadfast supporter and for your invaluable life advice. Thank you Mimi for your love and support and providing plentiful Maya pictures. To my Mom—thank you for inspiring curiosity of science by literally being my 8th grade science teacher, and introducing me to my childhood hero Dr. Sally Ride. Thank you Brian for being so welcoming and for introducing us to skiing. Lots of love to my East Coast family: Grammie, Holly, Kelsey, John, Carl, Sue Ellen, Chris, and Kate.

Thank you, thank you, thank you.

Table of Contents

List of Figures

List of Tables

1 Background and Overview

1.1 Introduction

A major challenge in realizing the full potential of large biomolecular therapeutics is the efficient delivery of these membrane-impermeable species into the cell. While the U.S. Food and Drug Administration (FDA) approval of biologics has boomed in the past decade,[1] the therapeutic targets of a vast majority of these drugs remain extracellular. The reason for this limitation is that the cell membrane is an exceptional barrier against unwanted transport.

Several approaches have been explored to tackle the problem of intracellular delivery of macromolecules, including liposomes, nanoparticles, and viral envelopes.[2] Although extensively studied in the lab, these methods contain a massive ratio of delivery vehicle to drug, and can suffer from unwanted toxicity. However, such methods of intracellular delivery are making large impacts on public health and passing through the FDA. For example, the mRNA vaccines recently approved for inoculation against SARS-CoV-2, turning the tide against the 2020 coronavirus pandemic, take advantage of liposomes to deliver synthetic RNA to the nuclei of cells.[3]

Conjugation to peptides may also be a promising strategy for the intracellular delivery of macromolecules. Sarepta Therapeutics, whose 2016 antisense oligonucleotide drug eteplirsen was the first U.S. approved treatment for Duchenne muscular dystrophy, is currently investigating a version of this drug with a covalently attached peptide portion to assist with delivery to the nucleus.[4] By modifying the oligonucleotide drug, the conjugate showed 10-fold improvement over the bare drug. Other promising therapeutic advancements using this type of peptide attachment, termed cell-penetrating peptides (CPPs), have emerged in recent years after more than two decades of in vitro research.[5] Therefore, the effective design of such sequences can greatly impact future therapies.

This chapter will discuss the promises and pitfalls of cell-penetrating peptides for the delivery of antisense oligonucleotides. While this chapter is not meant to be an exhaustive review of the current art, it will contextualize the topics presented in this thesis.

1.2 Antisense Oligonucleotides

A deeper understanding of the genetic and molecular basis of numerous human diseases has paved the way for the development of targeted therapies. Control of gene expression is a powerful means of rectifying a particular disease state but is challenging to implement as a therapeutic. Antisense oligonucleotides (ASOs), capable of binding sense-strand DNA or RNA via Watson-

Crick base pairing and modifying gene expression, have emerged as valuable tools for functional genomics, target validation, and more recently as therapeutics.[6] Oligonucleotides have developed through three generations.[7] Firstly, modifying the phosphodiester linkage has improved stability against nucleases. Secondly, additions of O-alkyl groups to the 2' position of the ribose moiety have lowered systemic toxicity. Most recently, major structural modifications to the backbone have enhanced hybridization affinity and stability, taking form as the third generation: phosphorodiamidate morpholino oligonucleotides (PMO) and peptide nucleic acids (PNA). These charge-neutral compounds act by steric blocking and can correct aberrant mRNA splicing by forcing the spliceosome to "skip" exons.

The first of these agents to be approved by the Food and Drug Administration (FDA) is Exondys 51 (eteplirsen) in 2016. This PMO therapy developed by Sarepta therapeutics was a breakthrough therapy for treatment of Duchenne muscular dystrophy (DMD) and targets exon 51.[8] Duchenne muscular dystrophy is a degenerative neuromuscular disorder that affects one in ten male births. A mutation in the dystrophin gene, the largest gene in the human body, causes a nonfunctional truncation ultimately leading to muscle degeneration.[9,10] Progressive muscle atrophy ultimately leads to loss of ambulation by teen years, and loss of life by late twenties.[11] However, the frontier in the race for therapies is bright, and just in the past four year, several more PMO therapies have been approved by the FDA, targeting exons 53 and 45.[12,13] These drugs may help up to 30% of individuals with DMD.

Although the recent approvals of PMO analogs to treat DMD is exciting, major obstacles remain to feasible treatment and further development. The primary hurdle for clinical advancement of these synthetic biopolymers is their poor cell permeability.[14] As might be expected for these macromolecular therapies (>6,500 Da), delivery to its genetic target in the nucleus remains a major challenge. As a consequence, clinical use of PMOs without delivery vehicles requires large and frequent administered doses.[15] However, studies have shown that two-thirds of eteplirsen is cleared renally within 24 h of administration.[8,16]

Traditional approaches to deliver PMO using liposomes and nanoparticles rarely advance to clinical work, often suffering from poor endosomal escape or significant toxicity.[17] Other methods of macromolecule delivery exist with varying efficiencies.[14,18–20] However, delivery of PMO by covalent attachment to cell-penetrating peptides (CPPs) has been studied broadly and has recently shown promise in clinical trials. Phase II clinical trial results have shown that once-monthly dosing

of a PMO drug covalently linked to a peptide (SRP-5051) resulted in higher tissue exposure, exon-skipping, and dystrophin production in patients taking a monthly dose compared to patients taking weekly doses of Exondys 51.[4]

1.3 Cell-Penetrating Peptides

The use of cell-penetrating peptides (CPPs) is a promising, general mode of macromolecule drug delivery. CPPs are short peptides of cationic, amphipathic or hydrophobic nature that facilitate intracellular delivery of cargoes that are otherwise non-cell penetrant, such as large and charged hydrophilic biomolecules.[21–24] Since the discovery that truncated HIV-1 Tat protein can rapidly translocate through the plasma membrane[25], several successful classes of CPPs have been discovered including cationic and amphipathic variants.[26] D-peptides have also been explored as CPPs, and have been found to display at times greater activities than their L-counterparts, although studied to a much lesser extent.[27] These sequences can deliver covalently bound cargo, offering therapeutic potential to macromolecules otherwise restricted to extracellular targets.

Although CPPs have been widely studied since their discovery, the field lacks robust methodology to quantify cell entry and penetration efficacy. This dearth of knowledge is due to the complicated mechanisms of CPP cell entry and the many variables that affect CPP efficacy in any given assay—such as peptide concentration, cell type, temperature, treatment time, and cargo.[26] For example, for the well-studied CPP penetratin (sequence: RQIKIWFQNRRMKWKK), the reported ratio between intracellular and extracellular concentration ranges from 0.6:1.0 to 95.0:1.0.[28,29] Even more challenging is determining the subcellular localization once a peptide is internalized, despite advances in fluorescence, immunoblot, and mass spectrometry detection.[30]

In addition to the varying experimental conditions, a CPP's translocation ability is affected by covalently bound cargo.[30] For example, penetratin was found to have a three-fold difference in intercellular concentration when attached to biotin versus carboxyfluorescein (CF).[30] Studies in our laboratory have demonstrated that the cell-penetrating ability of more than ten common CPPs differs when bound to a cyanine dye versus a PMO drug, with no discernable trend.[21] Therefore, if a CPP is intended to eventually deliver a therapeutic cargo, that cargo must be included in the initial development and study of the peptide.

The known sequence-activity motifs for CPPs largely rely on the guanidinium group of arginine (Arg).[31] While polyarginine peptides have been shown to promote cell uptake with greater

efficiency than other cationic residues due to membrane affinity and difference in effective protonation,[32,33] these sequences often remain trapped in the endosomes.[34] A promising polyarginine peptide, Bpep, triggers endosomal escape by interspacing arginine residues with long alkyl 6-aminohexanoic acid and beta-alanine residues.[35] Still, one of the key challenges to clinical translation of polyarginine PMO-CPPs is their demonstrated *in vivo* toxicity caused by the peptide portion.[36,37] While a linear relationship has been shown between number of arginine residues and LD50 in mice, there is a stark difference between observed toxicity in vitro and in vivo.[36,38] It is thought that the systemic toxicity induced by nonaarginine at 5 μmol/kg doses may be due to mast cell degranulation caused by the positive charges.[39,40]

Even so, recent clinical trial results show promise for the future of PMO-CPPs in patients.[4] Advances in strategies for the design and discovery of effective CPPs for the delivery of specific cargo would greatly impact the drug delivery field.

1.4 Methods to Characterize Delivery

There are several common methods to quantify uptake into cells. Even with these methods, it is still challenging to distinguish between endosomal and cytosolic localization. Currently, there are three widely used methods to assay CPPs, including fluorescence, immunoblot, and mass spectrometry.

Firstly, quantitative fluorescence is a technique in which a fluorophore is conjugated to a CPP and either monitored by fluorescence spectroscopy or quantified by flow cytometry.[30,41,42] These methods require a fluorescent reporter tag to be attached and limit analysis to one peptide per assay. Direct quantification of the cytosolic concentration of a fluorescently-labeled construct is possible using fluorescence correlation spectroscopy.[43] This method is currently the most robust method for directly quantifying construct concentration in the cytosol of live cells.

Next, phenotypic readout assays are methods in which successful delivery of a cargo causes a phenotypic change that can be measured. Several of these assays can differentiate between endosomal and cytosolic localization using indirect quantification via a readout generated by a delivered cargo, including the chloroalkane penetration assay (CAPA),[44] GFP complementation assays,[45] and more recently the NanoClick[46] assay and SLEEQ[47] assay. Our laboratory has studied CPPs capable of delivering an exon-skipping PMO that when delivered to the nucleus of engineered HeLa cells causes green fluorescent protein (GFP) expression, quantified by flow

cytrometry.[48] This method is particularly useful in that it requires no additional modifications besides the model therapeutic cargo, and it provides a measurement of the desired activity. However, this method does not provide information on the amount of conjugate inside the cell.

Another technique is mass spectrometry, which is used less frequently but has the advantage of direct quantification. Past studies have illustrated how matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry is a practical tool for absolute and relative quantification of peptides and proteins. For example, biotinylated L-CPPs were extracted from live cells and quantitated using heavy-labeled isotope analogs.[49,50] While this assay provided information regarding whole cell uptake of CPPs and CPP-peptide conjugates, it is limited by the need for heavy-atom labeling and the rapid degradation of L-peptides.[49] Other studies showed that an internal standard of a similar molecular weight is sufficient for generation of a calibration curve,[51] and a similar heavy isotope-free method was developed for the relative quantification of phosphopeptides.[52]

1.5 Empirical Methods to De Novo Discover Peptides

The first CPPs were discovered through truncations of full length proteins, including penetratin and TAT.[25,53] Motifs from these sequences informed empirical design of the first fully synthetic CPPs, including octaarginine.[54] As more CPP sequences were used, more synthetic sequences were designed using prediction programs, rational design, and even trial and error. Besides these approaches, a few profiling platforms have been developed to discover cell-penetrating peptide sequences de novo.

One study identified "phylomer" CPPs derived from bacterial and viral genomes using a phage-based screening platform.[55] Peptide-decorated phages contained an avitag sequence, which is biotinylated once inside the cytosol of engineered mammalian cells expressing biotin ligase (BirA). Phage that successfully enter the cytosol are then identified by sequencing. Sequences discovered by this method were able to deliver a variety of biological cargoes with micromolar activity.

Another screening platform discovered an unusual peptide via a plasmid display-based functional selection platform. After four rounds of screening a 14-mer peptide library, several peptides were identified and tested for their ability to deliver GFP into cells. While the tested peptides were unique in that they were non-cationic, the top-performing discovered peptide

performed at reduced activity compared to TAT peptide when the cationic lipid Lipofectamine was not added.[56]

On-bead library screening also has resulted in de novo discovered membrane-active sequences.[57] One-bead-one-compound (OBOC) combinatorial libraries were screened for binding against synthetic liposomes. Compounds with strong binding were identified as membrane active peptides and were synthesized and validated. Again, the discovered peptide showed uptake into cells but with reduced activity compared to a standard polyarginine peptide. Additional studies have screened combinatorial peptide libraries for translocation ability through synthetic membranes and preformed vesicles and identified effective sequences using mass spectrometry.[58,59]

While these methods are able to access broad chemical diversity, the profiling platforms typically screen peptide sequences out of the context of their desired delivery activity. As mentioned previously, experimental conditions and attached cargo can have significant impacts on the outcomes of activity assays. Therefore, screening platforms such as these described here would benefit greatly from inclusion of the desired cargo and targeted activity assays.

1.6 Machine Learning Methods to De Novo Discover Peptides

Besides sequence discovery by means of experimental selection, de novo design of cell penetrating peptide sequences can also be achieved using computational approaches. It is hypothesized that machine learning can enable interpolation in high-dimensional search spaces by bridging the gaps between experimental training data points.[60,61]

Machine learning has demonstrated promise for the design of functional peptides. Recent works have shown promise using a variety of input representations and quantitative activity prediction for design of new antimicrobial peptides and antibody CDR3 loops.[62–64] Similar studies have attempted to design novel CPP sequences using machine learning, including the use of binary classifiers to optimize activity.[65–68] The majority of such studies predict peptide sequences as being "CPPs" or "not CPPs," without a measure of range of activity.[61,66,69–71] The use of this strategy is a result of compiled public data, such as CPPsite 2.0, which hosts a large (>1000) repository of CPP sequences that have been reported in the literature. While this large dataset appears suitable for machine learning, which requires a large amount of data for reliable training, the lack of experimental consistency in the data prevents meaningful activity predictions.

In addition, inclusion of chemically diverse unnatural moieties using this strategy is challenging because such physicochemical descriptors may not be readily available, although it is possible to use one-hot encodings to represent unnatural residues in the training data if such descriptors are available. The ability to predict the effect of unnatural residues expands the chemical search space, and may lead to enhanced macromolecule delivery.[72]

In an attempt to predict novel sequences using experimentally consistent training data, our lab made 64 PMO-CPP sequences using CPP sequences found in the literature, and tested them in a consistent activity-based phenotypic assay. A random forest model was able to judge whether a randomly generated peptide sequence would or would not result in activity greater than 3-fold over PMO alone.[21] While meeting the benchmark of success, the study noted that a larger library composed of even more active sequences may make possible a machine learning framework able to produce optimized sequences with activities exceeding 50-fold over PMO.

1.7 Thesis Overview

This thesis focuses on the challenge of discovering and characterizing CPPs for PMO delivery. Chapters 2 and 3 of this thesis address this challenge by using machine learning to design de novo sequences using machine learning. Chapter 2 investigates a deep learning framework to design nuclear-targeting miniproteins, novel sequences that are far longer (30-80 residues) and more active than standard CPPs (<20 residues). The directed evolution-inspired model was composed of a generator-predictor-optimizer triad, in which the generator generates sequences, the predictor predicts a quantitative activity value, and the optimizer mutates the sequence against the predictor in order to increase activity while decreasing similarity and arginine content. The training of the predictor was primarily achieved using the standardized testing of a PMO-CPP library composed of modular chimeric peptides based on previously studied sequences. The resulting "Mach" miniproteins were unique compared to the training set, reached 50-fold activity over PMO, ranged from 30-80 residues in length, and contained less than 20% arginine content. A few sequences had a wide concentration window between PMO delivery activity and toxicity, were able to deliver other macromolecules including a 20 kDa anionic enzyme, and were effective in mice up to 30 mg/kg.

Chapter 3 builds upon the work in Chapter 2 in order to discover short peptides (<20 residues) by an augmented machine learning model. The lead peptide, P6, was 18 residues in length,

contained a single arginine residue, and rivaled the PMO activity of our positive control, polyarginine peptide Bpep. Sequence-activity studies of the P6 sequence demonstrated the importance of the C-terminal cysteine residue as well as the cationic lysine residues. A derivative of P6, lacking the cysteine-containing C-terminal motif, had slightly reduced activity but improved toxicity. Animal studies of P6 showed that the sequence effectively delivered PMO in mice up to 60 mg/kg without observed renal toxicity.

Chapter 4 steps back from sequence discovery and instead investigates the consequences of chirality of CPPs as well as a quantitative measure of delivery efficiency. We found that several widely studied CPPs retain their PMO activity in their all D-amino acid form, but the D-form remained completely stable to proteolysis while the native form was rapidly degraded. These biotinylated sequences were able to be recovered from inside cells following treatment, and analyzed by MALDI-TOF. By using an equimolar standard of a mixture of the CPPs, relative internalization into the whole cell and into the cytosol could be measured. Comparing relative internalization of PMO-CPPs to their delivery activity yields a measure of PMO delivery efficiency that could be a useful metric of judging future sequences.

Finally, Chapter 5 extends this information to the realm of peptide libraries and sequence identification. A combinatorial peptide library containing both unnatural residues and D-residues was screened for internalization into HeLa cells. Extraction of biotinylated PMO-peptides followed by sequencing by tandem mass spectrometry revealed several novel sequences localized in the cytosol. Validation of these sequences in the PMO activity assay demonstrated that the identified sequences showed higher delivery activity compared to the library in which they were found. This platform represents a new approach to discovery of CPP sequences for delivery of specific, therapeutically relevant cargos.

Taken together, the strategies demonstrated in this thesis provide new insights into how to design effective functional peptide sequences.

1.8 References

(1) Brown, D. G.; Wobst, H. J. A Decade of FDA-Approved Drugs (2010–2019): Trends and Future Directions. *J. Med. Chem.* **2021**, *64* (5), 2312–2338. https://doi.org/10.1021/acs.jmedchem.0c01516.

(2) Slastnikova, T. A.; Ulasov, A. V.; Rosenkranz, A. A.; Sobolev, A. S. Targeted Intracellular Delivery of Antibodies: The State of the Art. *Front. Pharmacol.* **2018**, *9*. https://doi.org/10.3389/fphar.2018.01208.

(3) Xia, X. Detailed Dissection and Critical Evaluation of the Pfizer/BioNTech and Moderna MRNA Vaccines. *Vaccines* **2021**, *9* (7), 734. https://doi.org/10.3390/vaccines9070734.

(4) Inc, S. T. Sarepta Therapeutics Announces Positive Clinical Results from MOMENTUM, a Phase 2 Clinical Trial of SRP-5051 in Patients with Duchenne Muscular Dystrophy Amenable to Skipping Exon 51 http://www.globenewswire.com/news-release/2020/12/07/2140613/0/en/Sarepta-Therapeutics-Announces-Positive-Clinical-Results-from-MOMENTUM-a-Phase-2-Clinical-Trial-of-SRP-5051-in-Patients-with-Duchenne-Muscular-Dystrophy-Amenable-to-Skipping-Exon-5.html (accessed 2020 -12 -07).

(5) Lulla, R. R.; Goldman, S.; Yamada, T.; Beattie, C. W.; Bressler, L.; Pacini, M.; Pollack, I. F.; Fisher, P. G.; Packer, R. J.; Dunkel, I. J.; Dhall, G.; Wu, S.; Onar, A.; Boyett, J. M.; Fouladi, M. Phase I Trial of P28 (NSC745104), a Non-HDM2-Mediated Peptide Inhibitor of P53 Ubiquitination in Pediatric Patients with Recurrent or Progressive Central Nervous System Tumors: A Pediatric Brain Tumor Consortium Study. *Neuro-Oncol.* **2016**, *18* (9), 1319–1325. https://doi.org/10.1093/neuonc/now047.

(6) Chery, J. RNA Therapeutics: RNAi and Antisense Mechanisms and Clinical Applications. *Postdoc J. J. Postdr. Res. Postdr. Aff.* **2016**, *4* (7), 35–50.

(7) Kurreck, J. Antisense Technologies. Improvement through Novel Chemical Modifications. *Eur. J. Biochem.* **2003**, *270* (8), 1628–1644. https://doi.org/10.1046/j.1432-1033.2003.03555.x.

(8) Lim, K. R. Q.; Maruyama, R.; Yokota, T. Eteplirsen in the Treatment of Duchenne Muscular Dystrophy. *Drug Des. Devel. Ther.* **2017**, *11*, 533–545. https://doi.org/10.2147/DDDT.S97635.

(9) Shimizu-Motohashi, Y.; Komaki, H.; Motohashi, N.; Takeda, S.; Yokota, T.; Aoki, Y. Restoring Dystrophin Expression in Duchenne Muscular Dystrophy: Current Status of Therapeutic Approaches. *J. Pers. Med.* **2019**, *9* (1), 1. https://doi.org/10.3390/jpm9010001.

(10) Abdel-Salam, E.; Abdel-Meguid, I.; Korraa, S. Markers of Degeneration and Regeneration in Duchenne Muscular Dystrophy. *Acta Myol.* **2009**, *28* (3), 94–100.

(11) Duan, D.; Goemans, N.; Takeda, S.; Mercuri, E.; Aartsma-Rus, A. Duchenne Muscular Dystrophy. *Nat. Rev. Dis. Primer* **2021**, *7* (1), 1–19. https://doi.org/10.1038/s41572-021-00248-3.

(12) Aartsma-Rus, A.; Corey, D. R. The 10th Oligonucleotide Therapy Approved: Golodirsen for Duchenne Muscular Dystrophy. *Nucleic Acid Ther.* **2020**, *30* (2), 67–70. https://doi.org/10.1089/nat.2020.0845.

(13) Iftikhar, M.; Frey, J.; Shohan, Md. J.; Malek, S.; Mousa, S. A. Current and Emerging Therapies for Duchenne Muscular Dystrophy and Spinal Muscular Atrophy. *Pharmacol. Ther.* **2021**, *220*, 107719. https://doi.org/10.1016/j.pharmthera.2020.107719.

(14) Juliano, R. L. The Delivery of Therapeutic Oligonucleotides. *Nucleic Acids Res.* **2016**, *44* (14), 6518–6548. https://doi.org/10.1093/nar/gkw236.
(15) Moulton, H. M.; Moulton, J. D. Morpholinos and Their Peptide Conjugates: Therapeutic Promise and Challenge for Duchenne Muscular Dystrophy. *Biochim. Biophys. Acta BBA - Biomembr.* **2010**, *1798* (12), 2296–2303. https://doi.org/10.1016/j.bbamem.2010.02.012.
(16) Baker, D. E. Eteplirsen. *Hosp. Pharm.* **2017**, *52* (4), 302–305. https://doi.org/10.1310/hpj5204-302.
(17) Malam, Y.; Loizidou, M.; Seifalian, A. M. Liposomes and Nanoparticles: Nanosized Vehicles for Drug Delivery in Cancer. *Trends Pharmacol. Sci.* **2009**, *30* (11), 592–599. https://doi.org/10.1016/j.tips.2009.08.004.
(18) Angelov, B.; Garamus, V. M.; Drechsler, M.; Angelova, A. Structural Analysis of Nanoparticulate Carriers for Encapsulation of Macromolecular Drugs. *J. Mol. Liq.* **2017**, *235*, 83–89. https://doi.org/10.1016/j.molliq.2016.11.064.
(19) Feger, G.; Angelov, B.; Angelova, A. Prediction of Amphiphilic Cell-Penetrating Peptide Building Blocks from Protein-Derived Amino Acid Sequences for Engineering of Drug Delivery Nanoassemblies. *J. Phys. Chem. B* **2020**, *124* (20), 4069–4078. https://doi.org/10.1021/acs.jpcb.0c01618.
(20) Fu, A.; Tang, R.; Hardie, J.; Farkas, M. E.; Rotello, V. M. Promises and Pitfalls of Intracellular Delivery of Proteins. *Bioconjug. Chem.* **2014**, *25* (9), 1602–1608. https://doi.org/10.1021/bc500320j.
(21) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS Cent. Sci.* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.
(22) Fadzen, C. M.; Holden, R. L.; Wolfe, J. M.; Choo, Z.-N.; Schissel, C. K.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. *Biochemistry* **2019**. https://doi.org/10.1021/acs.biochem.9b00413.
(23) Bechara, C.; Sagan, S. Cell-Penetrating Peptides: 20years Later, Where Do We Stand? *FEBS Lett.* **2013**, *587* (12), 1693–1702. https://doi.org/10.1016/j.febslet.2013.04.031.
(24) Schissel, C. K.; Mohapatra, S.; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep Learning to Design Nuclear-Targeting Abiotic Miniproteins. *Nat. Chem.* **2021**, 1–9. https://doi.org/10.1038/s41557-021-00766-3.
(25) Vivès, E.; Brodin, P.; Lebleu, B. A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus. *J. Biol. Chem.* **1997**, *272* (25), 16010–16017. https://doi.org/10.1074/jbc.272.25.16010.
(26) Reissmann Siegmund. Cell Penetration: Scope and Limitations by the Application of Cell-penetrating Peptides. *J. Pept. Sci.* **2014**, *20* (10), 760–784. https://doi.org/10.1002/psc.2672.
(27) Ma, Y.; Gong, C.; Ma, Y.; Fan, F.; Luo, M.; Yang, F.; Zhang, Y.-H. Direct Cytosolic Delivery of Cargoes in Vivo by a Chimera Consisting of D- and L-Arginine Residues. *J. Controlled Release* **2012**, *162* (2), 286–294. https://doi.org/10.1016/j.jconrel.2012.07.022.
(28) Fischer, R.; Waizenegger, T.; Köhler, K.; Brock, R. A Quantitative Validation of Fluorophore-Labelled Cell-Permeable Peptide Conjugates: Fluorophore and Cargo

Dependence of Import. *Biochim. Biophys. Acta BBA - Biomembr.* **2002**, *1564* (2), 365–374. https://doi.org/10.1016/S0005-2736(02)00471-6.
(29) Lindgren, M. E.; Hällbrink, M. M.; Elmquist, A. M.; Langel, U. Passage of Cell-Penetrating Peptides across a Human Epithelial Cell Layer in Vitro. *Biochem. J.* **2004**, *377* (Pt 1), 69–76. https://doi.org/10.1042/BJ20030760.
(30) Illien, F.; Rodriguez, N.; Amoura, M.; Joliot, A.; Pallerla, M.; Cribier, S.; Burlina, F.; Sagan, S. Quantitative Fluorescence Spectroscopy and Flow Cytometry Analyses of Cell-Penetrating Peptides Internalization Pathways: Optimization, Pitfalls, Comparison with Mass Spectrometry Quantification. *Sci. Rep.* **2016**, *6*, 36938. https://doi.org/10.1038/srep36938.
(31) Brock, R. The Uptake of Arginine-Rich Cell-Penetrating Peptides: Putting the Puzzle Together. *Bioconjug. Chem.* **2014**, *25* (5), 863–868. https://doi.org/10.1021/bc500017t.
(32) Åmand, H. L.; Rydberg, H. A.; Fornander, L. H.; Lincoln, P.; Nordén, B.; Esbjörner, E. K. Cell Surface Binding and Uptake of Arginine- and Lysine-Rich Penetratin Peptides in Absence and Presence of Proteoglycans. *Biochim. Biophys. Acta BBA - Biomembr.* **2012**, *1818* (11), 2669–2678. https://doi.org/10.1016/j.bbamem.2012.06.006.
(33) Chuard, N.; Fujisawa, K.; Morelli, P.; Saarbach, J.; Winssinger, N.; Metrangolo, P.; Resnati, G.; Sakai, N.; Matile, S. Activation of Cell-Penetrating Peptides with Ionpair−π Interactions and Fluorophiles. *J. Am. Chem. Soc.* **2016**, *138* (35), 11264–11271. https://doi.org/10.1021/jacs.6b06253.
(34) Erazo-Oliveras, A.; Muthukrishnan, N.; Baker, R.; Wang, T.-Y.; Pellois, J.-P. Improving the Endosomal Escape of Cell-Penetrating Peptides and Their Cargos: Strategies and Challenges. *Pharmaceuticals* **2012**, *5* (11), 1177–1209. https://doi.org/10.3390/ph5111177.
(35) Jearawiriyapaisarn, N.; Moulton, H. M.; Buckley, B.; Roberts, J.; Sazani, P.; Fucharoen, S.; Iversen, P. L.; Kole, R. Sustained Dystrophin Expression Induced by Peptide-Conjugated Morpholino Oligomers in the Muscles of Mdx Mice. *Mol. Ther.* **2008**, *16* (9), 1624–1629. https://doi.org/10.1038/mt.2008.120.
(36) Li, Q.; Xu, M.; Cui, Y.; Huang, C.; Sun, M. Arginine-rich Membrane-permeable Peptides Are Seriously Toxic. *Pharmacol. Res. Perspect.* **2017**, *5* (5). https://doi.org/10.1002/prp2.334.
(37) Amantana, A.; Moulton, H. M.; Cate, M. L.; Reddy, M. T.; Whitehead, T.; Hassinger, J. N.; Youngblood, D. S.; Iversen, P. L. Pharmacokinetics, Biodistribution, Stability and Toxicity of a Cell-Penetrating Peptide−Morpholino Oligomer Conjugate. *Bioconjug. Chem.* **2007**, *18* (4), 1325–1331. https://doi.org/10.1021/bc070060v.
(38) Grogg, M.; Hilvert, D.; Ebert, M.-O.; Beck, A. K.; Seebach, D.; Kurth, F.; Dittrich, P. S.; Sparr, C.; Wittlin, S.; Rottmann, M.; Mäser, P. Cell Penetration, Herbicidal Activity, and *in-Vivo* -Toxicity of Oligo-Arginine Derivatives and of Novel Guanidinium-Rich Compounds Derived from the Biopolymer Cyanophycin. *Helv. Chim. Acta* **2018**, *101* (10), e1800112. https://doi.org/10.1002/hlca.201800112.
(39) Aguilera, T. A.; Timmers, M. M.; Olson, E. S.; Jiang, T.; Tsien, R. Y. Systemic in Vivo Distribution of Activatable Cell Penetrating Peptides Is Superior to Cell Penetrating Peptides. *Integr. Biol. Quant. Biosci. Nano Macro* **2009**, *1* (5–6), 371–381. https://doi.org/10.1039/b904878b.
(40) Albright, C. F.; Graciani, N.; Han, W.; Yue, E.; Stein, R.; Lai, Z.; Diamond, M.; Dowling, R.; Grimminger, L.; Zhang, S.-Y.; Behrens, D.; Musselman, A.; Bruckner, R.; Zhang, M.;

Jiang, X.; Hu, D.; Higley, A.; Dimeo, S.; Rafalski, M.; Mandlekar, S.; Car, B.; Yeleswaram, S.; Stern, A.; Copeland, R. A.; Combs, A.; Seitz, S. P.; Trainor, G. L.; Taub, R.; Huang, P.; Oliff, A. Matrix Metalloproteinase-Activated Doxorubicin Prodrugs Inhibit HT1080 Xenograft Growth Better than Doxorubicin with Less Toxicity. *Mol. Cancer Ther.* **2005**, *4* (5), 751–760. https://doi.org/10.1158/1535-7163.MCT-05-0006.

(41) Rezgui, R.; Blumer, K.; Yeoh-Tan, G.; Trexler, A. J.; Magzoub, M. Precise Quantification of Cellular Uptake of Cell-Penetrating Peptides Using Fluorescence-Activated Cell Sorting and Fluorescence Correlation Spectroscopy. *Biochim. Biophys. Acta BBA - Biomembr.* **2016**, *1858* (7, Part A), 1499–1506. https://doi.org/10.1016/j.bbamem.2016.03.023.

(42) Qian, Z.; Dougherty, P. G.; Pei, D. Monitoring the Cytosolic Entry of Cell-Penetrating Peptides Using a PH-Sensitive Fluorophore. *Chem. Commun.* **2015**, *51* (11), 2162–2165. https://doi.org/10.1039/C4CC09441G.

(43) Wissner, R. F.; Steinauer, A.; Knox, S. L.; Thompson, A. D.; Schepartz, A. Fluorescence Correlation Spectroscopy Reveals Efficient Cytosolic Delivery of Protein Cargo by Cell-Permeant Miniature Proteins. *ACS Cent. Sci.* **2018**, *4* (10), 1379–1393. https://doi.org/10.1021/acscentsci.8b00446.

(44) Peraro, L.; Deprey, K. L.; Moser, M. K.; Zou, Z.; Ball, H. L.; Levine, B.; Kritzer, J. A. Cell Penetration Profiling Using the Chloroalkane Penetration Assay. *J. Am. Chem. Soc.* **2018**, *140* (36), 11360–11369. https://doi.org/10.1021/jacs.8b06144.

(45) Schmidt, S.; Adjobo-Hermans, M. J. W.; Wallbrecher, R.; Verdurmen, W. P. R.; Bovée-Geurts, P. H. M.; van Oostrum, J.; Milletti, F.; Enderle, T.; Brock, R. Detecting Cytosolic Peptide Delivery with the GFP Complementation Assay in the Low Micromolar Range. *Angew. Chem. Int. Ed.* **2015**, *54* (50), 15105–15108. https://doi.org/10.1002/anie.201505913.

(46) Peier, A.; Ge, L.; Boyer, N.; Frost, J.; Duggal, R.; Biswas, K.; Edmondson, S.; Hermes, J. D.; Yan, L.; Zimprich, C.; Sadruddin, A.; Kristal Kaan, H. Y.; Chandramohan, A.; Brown, C. J.; Thean, D.; Lee, X. E.; Yuen, T. Y.; Ferrer-Gago, F. J.; Johannes, C. W.; Lane, D. P.; Sherborne, B.; Corona, C.; Robers, M. B.; Sawyer, T. K.; Partridge, A. W. NanoClick: A High Throughput, Target-Agnostic Peptide Cell Permeability Assay. *ACS Chem. Biol.* **2021**, *16* (2), 293–309. https://doi.org/10.1021/acschembio.0c00804.

(47) Teo, S. L. Y.; Rennick, J. J.; Yuen, D.; Al-Wassiti, H.; Johnston, A. P. R.; Pouton, C. W. Unravelling Cytosolic Delivery of Cell Penetrating Peptides with a Quantitative Endosomal Escape Assay. *Nat. Commun.* **2021**, *12* (1), 3721. https://doi.org/10.1038/s41467-021-23997-x.

(48) Wolfe Justin M.; Fadzen Colin M.; Holden Rebecca L.; Yao Monica; Hanson Gunnar J.; Pentelute Bradley L. Perfluoroaryl Bicyclic Cell-Penetrating Peptides for Delivery of Antisense Oligonucleotides. *Angew. Chem. Int. Ed.* **2018**, *0* (0). https://doi.org/10.1002/anie.201801167.

(49) Aubry, S.; Aussedat, B.; Delaroche, D.; Jiao, C.-Y.; Bolbach, G.; Lavielle, S.; Chassaing, G.; Sagan, S.; Burlina, F. MALDI-TOF Mass Spectrometry: A Powerful Tool to Study the Internalization of Cell-Penetrating Peptides. *Biochim. Biophys. Acta BBA - Biomembr.* **2010**, *1798* (12), 2182–2189. https://doi.org/10.1016/j.bbamem.2009.11.011.

(50) Burlina, F.; Sagan, S.; Bolbach, G.; Chassaing, G. Quantification of the Cellular Uptake of Cell-Penetrating Peptides by MALDI-TOF Mass Spectrometry. *Angew. Chem. Int. Ed.* **2005**, *44* (27), 4244–4247. https://doi.org/10.1002/anie.200500477.

(51) Duncan, M. W. Practical Quantitative Biomedical Applications of MALDI-TOF Mass Spectrometry.
(52) Ho, H.-P.; Rathod, P.; Louis, M.; Tada, C. K.; Rahaman, S.; Mark, K. J.; Leng, J.; Dana, D.; Kumar, S.; Lichterfeld, M.; Chang, E. J. Studies on Quantitative Phosphopeptide Analysis by MALDI Mass Spectrometry Without Label, Chromatography or Calibration Curves. *Rapid Commun. Mass Spectrom. RCM* **2014**, *28* (24), 2681–2689. https://doi.org/10.1002/rcm.7063.
(53) Derossi, D.; Calvet, S.; Trembleau, A.; Brunissen, A.; Chassaing, G.; Prochiantz, A. Cell Internalization of the Third Helix of the Antennapedia Homeodomain Is Receptor-Independent. *J. Biol. Chem.* **1996**, *271* (30), 18188–18193. https://doi.org/10.1074/jbc.271.30.18188.
(54) Lundberg, P.; El-Andaloussi, S.; Sütlü, T.; Johansson, H. Delivery of Short Interfering RNA Using Endosomolytic Cell-Penetrating Peptides. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2007**, *21* (11), 2664–2671. https://doi.org/10.1096/fj.06-6502com.
(55) Hoffmann, K.; Milech, N.; Juraja, S. M.; Cunningham, P. T.; Stone, S. R.; Francis, R. W.; Anastasas, M.; Hall, C. M.; Heinrich, T.; Bogdawa, H. M.; Winslow, S.; Scobie, M. N.; Dewhurst, R. E.; Florez, L.; Ong, F.; Kerfoot, M.; Champain, D.; Adams, A. M.; Fletcher, S.; Viola, H. M.; Hool, L. C.; Connor, T.; Longville, B. A. C.; Tan, Y.-F.; Kroeger, K.; Morath, V.; Weiss, G. A.; Skerra, A.; Hopkins, R. M.; Watt, P. M. A Platform for Discovery of Functional Cell-Penetrating Peptides for Efficient Multi-Cargo Intracellular Delivery. *Sci. Rep.* **2018**, *8* (1), 12538. https://doi.org/10.1038/s41598-018-30790-2.
(56) Gao, S.; Simon, M. J.; Hue, C. D.; Morrison, B.; Banta, S. An Unusual Cell Penetrating Peptide Identified Using a Plasmid Display-Based Functional Selection Platform. *ACS Chem. Biol.* **2011**, *6* (5), 484–491. https://doi.org/10.1021/cb100423u.
(57) Carney, R. P.; Thillier, Y.; Kiss, Z.; Sahabi, A.; Heleno Campos, J. C.; Knudson, A.; Liu, R.; Olivos, D.; Saunders, M.; Tian, L.; Lam, K. S. Combinatorial Library Screening with Liposomes for Discovery of Membrane Active Peptides. *ACS Comb. Sci.* **2017**, *19* (5), 299–307. https://doi.org/10.1021/acscombsci.6b00182.
(58) Guixer B.; Arroyo X.; Belda I.; Sabidó E.; Teixidó M.; Giralt E. Chemically Synthesized Peptide Libraries as a New Source of BBB Shuttles. Use of Mass Spectrometry for Peptide Identification. *J. Pept. Sci.* **2016**, *22* (9), 577–591. https://doi.org/10.1002/psc.2900.
(59) Marks, J. R.; Placone, J.; Hristova, K.; Wimley, W. C. Spontaneous Membrane-Translocating Peptides by Orthogonal High-Throughput Screening. *J. Am. Chem. Soc.* **2011**, *133* (23), 8995–9004. https://doi.org/10.1021/ja2017416.
(60) Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; Volkov, Y.; Zholus, A.; Shayakhmetov, R. R.; Zhebrak, A.; Minaeva, L. I.; Zagribelnyy, B. A.; Lee, L. H.; Soll, R.; Madge, D.; Xing, L. Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors. https://doi.org/10.1038/s41587-019-0224-x.
(61) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackerman, Z. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180* (4), 688–702.

(62) Spänig, S.; Heider, D. Encodings and Models for Antimicrobial Peptide Classification for Multi-Resistant Pathogens. *BioData Min.* **2019**, *12*, 7. https://doi.org/10.1186/s13040-019-0196-x.
(63) Witten, J.; Witten, Z. Deep Learning Regression Model for Antimicrobial Peptide Design. *bioRxiv* **2019**, 692681. https://doi.org/10.1101/692681.
(64) Liu, G.; Zeng, H.; Mueller, J.; Carter, B.; Wang, Z.; Schilz, J.; Horny, G.; Birnbaum, M. E.; Ewert, S.; Gifford, D. K. Antibody Complementarity Determining Region Design Using High-Capacity Machine Learning. *Bioinformatics* **2020**, *36* (7), 2126–2133. https://doi.org/10.1093/bioinformatics/btz895.
(65) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS Cent. Sci.* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.
(66) Su, R.; Hu, J.; Zou, Q.; Manavalan, B.; Wei, L. Empirical Comparison and Analysis of Web-Based Cell-Penetrating Peptide Prediction Tools. *Brief Bioinform* **2019**.
(67) Sanders, W. S.; Johnston, C. I.; Bridges, S. M.; Burgess, S. C.; Willeford, K. O. Prediction of Cell Penetrating Peptides by Support Vector Machines. *PLoS Comput. Biol.* **2011**, *7* (7), e1002101.
(68) Manavalan, B.; Subramaniyam, S.; Shin, T. H.; Kim, M. O.; Lee, G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *J. Proteome Res.* **2018**, *17* (8), 2715–2726. https://doi.org/10.1021/acs.jproteome.8b00148.
(69) Wei, L.; Tang, J.; Zou, Q. SkipCPP-Pred: An Improved and Promising Sequence-Based Predictor for Predicting Cell-Penetrating Peptides. *BMC Genomics* **2017**, *18* (7), 742.
(70) Pandey, P.; Patel, V.; George, N. V; Mallajosyula, S. S. KELM-CPPpred: Kernel Extreme Learning Machine Based Prediction Model for Cell-Penetrating Peptides. *J. Proteome Res.* **2018**, *17* (9), 3214–3222.
(71) Chen, B.; Khodadoust, M. S.; Olsson, N.; Wagar, L. E.; Fast, E.; Liu, C. L.; Muftuoglu, Y.; Sworder, B. J.; Diehn, M.; Levy, R.; Davis, M. M.; Elias, J. E.; Altman, R. B.; Alizadeh, A. A. Predicting HLA Class II Antigen Presentation through Integrated Deep Learning. *Nat. Biotechnol.* **2019**, *37* (November). https://doi.org/10.1038/s41587-019-0280-2.
(72) Jearawiriyapaisarn, N.; Moulton, H. M.; Buckley, B.; Roberts, J.; Sazani, P.; Fucharoen, S.; Iversen, P. L.; Kole, R. Sustained Dystrophin Expression Induced by Peptide-Conjugated Morpholino Oligomers in the Muscles of Mdx Mice. *Mol. Ther.* **2008**, *16* (9), 1624–1629.

2 Deep Learning to Design Nuclear-Targeting Abiotic Miniproteins

The work presented in this chapter has been reproduced from the following publication with the permission of Springer Nature:

Schissel, C. K.*; Mohapatra, S.*; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep Learning to Design Nuclear-Targeting Abiotic Miniproteins. *Nat. Chem.* **2021**, 13, 992-1000.

**These authors contributed equally to this work.*

2.1 Introduction

The vast chemical search space hinders design of functional macromolecules by empirical approaches alone.[1] It is hypothesized that machine learning can enable interpolation in high-dimensional search spaces by bridging the gaps between experimental training data points.[2,3] Recent works have shown promise using a variety of input representations and quantitative activity prediction for design of new antimicrobial peptides and antibody CDR3 loops.[4–6] For cell-penetrating peptides (CPPs), similar strategies involving binary classifiers have been used to optimize activity.[7–10] We sought to further address this challenge by using a large standardized dataset and an advanced input representation combined with deep learning to simultaneously design new functional miniproteins and quantitatively predict their activity.

Successful design of functional polymers can have considerable implications for medicine. For example, anticancer miniproteins have been shown to access intracellular targets.[11,12] Similarly, CPPs are short (5-20 residue) sequences that can enhance intracellular delivery of biomolecules, such as oligonucleotides and proteins, that otherwise cannot efficiently cross the cell membrane.[13–17] While promising, variation in experimental design has resulted in inconsistent and sometimes contradictory datasets. For example, penetratin has different efficacy as a CPP depending on the assay and the cargo.[18] These inconsistent results preclude the development of sequence-activity relationships and complicate the use of machine learning models to design analogs de novo.[19–21]

We overcome these challenges by de novo design of abiotic miniproteins that deliver an active cargo, antisense phosphorodiamidate morpholino oligomer (PMO), to the nucleus of cells. The miniproteins described here are distinct in that they have a defined function (PMO delivery) and are significantly longer (30-80 residues) than CPPs (5-20 residues). While PMO has recently been approved for the treatment of Duchenne muscular dystrophy, a major challenge remains with their poor cellular permeability.[13–17,22,23] High doses of PMO of up to 50 mg/kg are required for in vivo efficacy.[24] It has been shown that nuclear delivery can be improved by attaching PMO to CPPs, and the first clinical success of this strategy has been demonstrated just this year.[25,26] Development of advanced, novel sequences for antisense delivery would rapidly accelerate the development of these gene therapies.

Here we report a deep learning-based design strategy with predictive power fueled by robust input data containing unnatural residues and structures. Our framework includes generation of starting sequences, a predictor to predict the activity of a sequence, and an optimizer to improve

the activity of the sequence. A library containing 600 unique antisense-miniprotein conjugates was constructed using linear combinations of three peptides, or "modules" (Fig. 2.1a). A quantitative activity readout was achieved using an in vitro assay in which nuclear delivery of PMO results in enhanced green fluorescent protein (EGFP) fluorescence (Fig. 2.11b-c). Residues were encoded as fingerprints to provide chemical structure information, labeled with corresponding activity data, and used to train a predictor neural network (Fig. 1d). A "CPP thesaurus" dataset was used to train a generator neural network to produce novel sequences that are "CPP-like" to be used as seeds for optimization. These novel sequences were then optimized in the predictor-optimizer loop to increase predicted activity while minimizing similarity to the library, and minimizing length and arginine content to mitigate toxicity.[27] The output is hundreds of de novo designed sequences with a broad spectrum of predicted activity (Fig. 2.1a).

The model is also interpretable: we can visualize the decision-making process and identify structure-activity relationships that are consistent with empirical observations. From these predictions, we discovered best-in-class abiotic "Mach" (Machine Learning) nuclear-targeting miniproteins that improve PMO delivery by 50-fold and are effective in animals. Mach miniproteins are nontoxic and noninflammatory, and are able to deliver macromolecules other than PMO to the cytosol. Our approach has the potential to be extended to the design of peptides with other functions, although further work is required in these directions.

2.2 Results

2.2.1 Assembly of a Standardized Dataset

Recently, we demonstrated that linear combinations of known CPP sequences into chimeric miniproteins can synergistically improve delivery of PMO compared to each CPP alone.[28] We hypothesized that expanding this approach to a larger, more diverse library of linear combinations of CPPs would access a wide range of sequences and activities. We designed a synthetic method to assemble this library via bioconjugation of peptide "modules" into hundreds of novel PMO-miniproteins. Our rationale is that such a library would enable a broad sequence diversity and spectrum of activities and would be ideal to train machine learning models (Fig. 2.2).

Our synthesis strategy employs four modules: one for PMO and three for distinct pools of peptide sequences containing diverse structure and function, including nuclear-targeting peptides and peptides containing unnatural residues and cysteine-linked macrocycles.[29] The constructs were

synthesized in a series of bioconjugation reactions that are chemoselective and irreversible, yielding products of sufficient crude purity for direct testing in vitro. The resulting library contained 600 miniproteins, composed of combinations of 57 total peptides. Complete details for the synthesis and study of this combinatorial library were previously discussed in Dr. Justin Wolfe's thesis.[30]

The resulting dataset was broad in terms of both peptide sequences and range of activity, quantified by a high-throughput nuclear-targeting assay.[19] The activity-based assay that is used to acquire the training data provides a direct, quantitative readout of the activity characteristic we want to enhance—specifically nuclear delivery. In this assay, HeLa cells stably transfected with an EGFP gene interrupted by a mutated intron of β-globin (IVS2-654) produce a non-fluorescent EGFP protein. Successful delivery of PMO IVS2-654 to the nucleus results in corrective splicing and EGFP synthesis. The amount of PMO delivered to the nucleus is therefore correlated with EGFP fluorescence, quantified by flow cytometry. Activity is reported as mean fluorescence intensity (MFI) relative to PMO alone (Fig. 2.1b,c, Appendix I). The most active construct improved PMO delivery by nearly 20-fold, while the median activity was 3-fold.

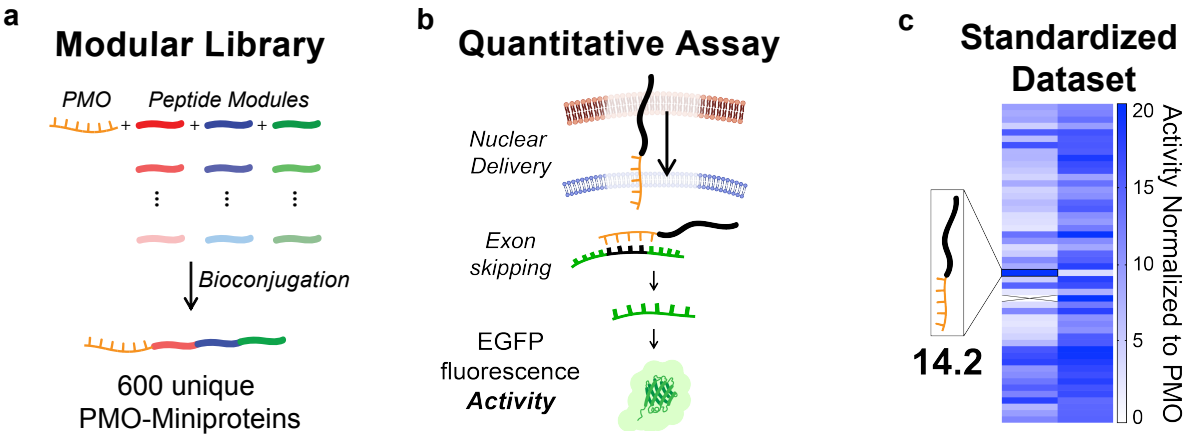


Figure 2.1 Modular library enabled 600-member standardized dataset (a) A 600-membered library of PMO-miniprotein conjugates was synthesized using linear combinations of abiotic peptide modules. (b) A standardized in vitro activity assay tests for nuclear delivery using a quantitative fluorescence readout. (c) Members of the modular library exhibit a broad spectrum of activities. Each bit corresponds to a PMO-peptide in the library and its corresponding activity.

2.2.2 Developing the Deep Learning Model

Inspired by directed evolution, we leveraged fingerprint sequence representations to develop a machine learning-based generator-predictor-optimizer triad. In this framework, the generator produces novel cell-penetrating sequences, the predictor quantitatively estimates the activity for a given sequence, and the optimizer evolves towards the most optimal miniprotein sequence.

The standardized dataset of activity-labeled sequences from the modular library allowed for development and training of a quantitative regressor algorithm. This approach enabled us to overcome the limitations of other efforts in the computational CPP literature which employed binary classifiers of active versus inactive sequences.[2,3,8,31–33] These previous predictors were mostly trained using physicochemical descriptors, with datasets obtained from non-standardized experiments and containing only natural residues.[7,10,34,35] Inclusion of chemically diverse unnatural moieties using this strategy is challenging because such physicochemical descriptors may not be readily available. The ability to predict the effect of unnatural residues expands the chemical search space, and may lead to enhanced macromolecule delivery.[36] One-hot residue encodings can be extended to represent unnatural residues in the training data. We were interested, however, in encoding the molecular structure of each residue. Therefore, to predict activity of de novo-designed nuclear-targeting abiotic miniproteins, we evaluated a topological representation based on stacking traditional cheminformatics fingerprints for each residue along the sequence.[37] This representation extends the approach of using one-hot encodings for quantitative structure activity relationship predictors in the peptide literature,[4,5] provides chemical structure information for unnatural residues, and may leverage weight sharing across structurally similar residues. Combined with quantitative experimental readouts, this polymer representation allows us to access the diverse pool of unnatural residues and structures and quantitatively predict delivery activity.

We represented peptide structure by encoding molecular construction. Peptide sequences are represented as matrices comprised of residue fingerprints in the columns, padded with zeros until each sequence matrix is the same length. Individual residue fingerprints are bit-vectors based on the molecular graph of the whole monomer, including backbone and side-chain. We used 2048-bit ECFP6 fingerprints generated by RDKit (Fig. 2.2b, Appendix III) but other structural descriptors may be used.[38] For analysis and visualization of fingerprints, we removed all indices which are inactive across all residues, resulting in a condensed 191-bit fingerprint. Each bit in the vector corresponds to a substructure, and is active/inactive depending on the presence/absence of the

particular substructure. Representing residues as chemical structures, rather than discrete choices, eases the use of both natural and unnatural residues and leverages chemical similarity between residues. The fingerprints are then compiled into a row matrix to encode the amide backbone of the peptide sequence (Fig. 2.2c).

The predictor neural network quantitatively estimated normalized MFI for a given sequence. Pairs of sequence representations and corresponding experimental activities were used to train a convolutional neural network (CNN). The training dataset consisted of PMO-miniproteins from the modular library as well as other conjugates previously tested in the same assay.[7] A randomly-selected 20% of the dataset was saved for validation of the predictive accuracy of the algorithm. The root mean squared error (RMSE) on the validation set was 0.4 of the standard deviation of the training data. The prediction relative error was found to be 11% as long as the predicted activity fell within the range of training values (normalized activity of 0.32-19.5) (Fig. 2.2d).

We developed a generator based on a recurrent neural network (RNN) that captured the ontology of CPPs and generated "CPP-like" starter sequences. We trained the generator using a nested long short-term memory (LSTM) neural network architecture, which is better able to capture long-range correlations in sequence data.[39] We trained the algorithm using a "CPP thesaurus," a collection of sequences from both our modular library and the literature.[40] Because the model is learning sequence grammar and has no role in activity predictions, no quantitative labels are necessary and we can use a large dataset of available sequences.

The optimizer completed the loop based on directed evolution. Sequences from the generator were randomly mutated and evaluated against an objective function, which maximized activity as predicted by the CNN model, and minimized length, arginine content, and similarity to the library while retaining water solubility estimated with net charge of the sequence (Table S10). After 1000 iterations over each sequence, the model delivered hundreds of unique sequences with a wide range of predicted activity values. Along with highly active sequences, we predicted inactive sequences as negative control. By directing the evolution of the optimizer in the opposite direction, i.e., minimizing MFI, but keeping other constraints the same, we were able to generate an inactive sequence (Mach11) that appeared similar in amino acid composition to the active predictions. After synthesis, the Mach11 conjugate displayed low experimental activity, demonstrating the robustness of the model in predicting the activity of a unique sequence (Fig. 2.2d).

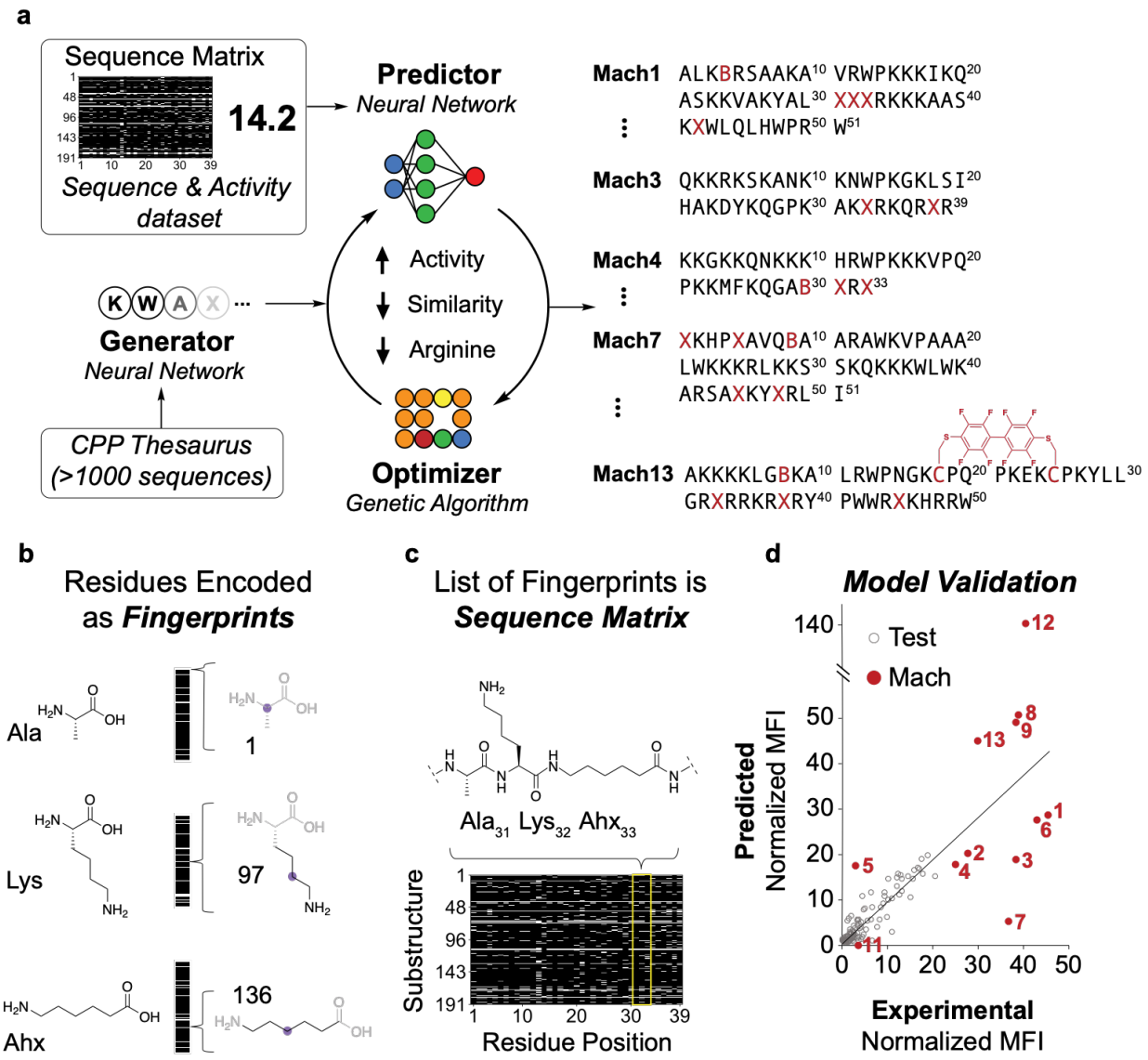


Figure 2.2 Machine learning-based generator-predictor-optimizer loop predicts nuclear-targeting abiotic miniproteins (a) Sequences are encoded into a fingerprint matrix, labeled with experimental activity, and used to train a machine learning model. The model designs novel sequences in a loop based on directed evolution. X = aminohexanoic acid, B = β-alanine, C = cysteine macrocycles linked through decafluorobiphenyl. (b) Each amino acid residue is represented as a unique fingerprint, constructed as a bit-vector encoding for the presence or absence of 191 possible substructures in the residue. (c) Sequences are represented as residue fingerprints stacked in a row matrix. (d) Performance of machine learning model, comparing the predicted and experimental activity values for the holdout test set and novel Mach sequences.

2.2.3 Evaluating the Deep Learning Model

When developing the machine learning model, we conducted tests to compare the CNN model against other model architectures, using both fingerprint and one-hot encodings in both regression and classification tasks (Methods Section; Tables 2.8-2.12, Figs. 2.26, 2.27). We observed that most of these models were limited by the range of the training data, and that only the CNN-FP

model was able to extrapolate in the codomain and generate predicted activity values (validated by experimental activity values) that were greater than any in the training set. This ability to extrapolate, however, came at a cost in average accuracy because of the increased statistical noise of extrapolated predictions. Models based on the topological representations added only minimal increase in performance over one-hot encodings on the validation dataset, and performed similarly or worse on the Mach dataset, due to outliers with extreme predicted activity values. However, a CNN model using one-hot encodings, despite its lowest overall average error, was not able to extrapolate in the codomain space, unlike when using fingerprint representations. To investigate the role of outliers that impact model performance, we used model ensembling and found the ensembled CNN one-hot model is superior for the validation dataset, whereas the ensembled CNN-FP model is superior for the Mach dataset, likely due to its ability to extrapolate in the codomain. These parameters and metrics are reported fully in the methods section. Further efforts should be focused on how to accurately predict activity values that reach beyond that of the training set.

As a direct comparison to existing machine learning models, we tested whether reported models (hosted as webservers) were able to predict activity of the Mach miniproteins accurately. We compared our model to currently available CPP prediction tools by evaluating predictions for Mach peptides (accessed on September 3, 2020). (doi: 10.1186/1479-5876-11-74, doi: 10.1186/s12864-017-4128-1, doi: 10.1021/acs.jproteome.7b00019, doi: 10.1021/acs.jproteome.8b00148, doi: 10.1093/bib/bby091) Of note, these prediction tools do not allow for unnatural residues, therefore when testing the Mach sequences, B (β-alanine) and X (aminohexanoic acid) were replaced by A (alanine) and L (leucine) respectively. Macrocyclic peptides were treated as linear peptides (Table 2.1). All the webservers are generic, as in they do not differentiate between different cargo, binary classifiers and provide the classification probability of the sequence being a CPP, and uptake probability. Most webservers, with the exception of CellPPD, classified all Mach peptides, including the negative control Mach11, as CPP. This result indicates that the webservers are not robust enough to differentiate between highly active and poorly active CPPs. The current work of training a quantitative model (regressor) over a standard dataset with consistent cargo and experimentation is necessary to achieve this distinction.

Table 2.1 Online webservers *

Model →	CellPPD		SkipCPP-Pred		CPPred-RF				MLCPP				CPPred-FL		Current Work
Summary →	2013 - SVM - 1-50		2017 - RF - >10		2017 - RF - N/A				2018 - ERT and RF - 5-30				2018 - RF - N/A*		2020 - CNN - N/A
Sequence ↓	Pred	SVM Score	Pred	Pred Conf	Pred	Pred Conf	Upt	Upt Conf	Pred	Pred Conf	Upt	Upt Conf	Pred	Pred Conf	Activity
Mach 1	Non-CPP	-0.03	CPP	0.93	CPP	0.73	High	0.50	CPP	0.86	High	0.51	CPP	0.69	29.68
Mach 2	Non-CPP	-0.13	CPP	0.81	CPP	0.67	High	0.52	CPP	0.77	Low	0.31	CPP	0.85	20.82
Mach 3	CPP	0.43	CPP	0.79	CPP	0.73	High	0.52	CPP	0.85	Low	0.45	CPP	0.69	19.52
Mach 4	CPP	0.20	CPP	0.81	CPP	0.79	High	0.56	CPP	0.87	Low	0.39	CPP	0.61	18.48
Mach 5	CPP	0.07	CPP	0.71	CPP	0.69	High	0.59	CPP	0.58	Low	0.41	CPP	0.76	17.47
Mach 6	Non-CPP	-0.03	CPP	0.94	CPP	0.72	High	0.53	CPP	0.86	Low	0.50	CPP	0.82	27.25
Mach 7	Non-CPP	-0.09	CPP	0.93	CPP	0.73	High	0.61	CPP	0.86	High	0.51	CPP	0.77	5.30
Mach 8	Non-CPP	-0.02	CPP	0.91	CPP	0.68	High	0.53	CPP	0.74	High	0.50	CPP	0.78	50.55
Mach 9	CPP	0.18	CPP	0.90	CPP	0.78	High	0.56	CPP	0.89	High	0.51	CPP	0.85	48.90
Mach 10	Non-CPP	-0.56	CPP	0.92	CPP	0.61	High	0.52	CPP	0.68	High	0.52	CPP	0.67	63.43
Mach 11	Non-CPP	-0.40	CPP	0.70	CPP	0.56	High	0.57	CPP	0.56	Low	0.35	CPP	0.77	-11.75
Mach 12	CPP	0.07	CPP	0.90	CPP	0.73	High	0.55	CPP	0.79	High	0.52	CPP	0.75	140.24
Mach 13	CPP	0.48	CPP	0.84	CPP	0.78	High	0.55	CPP	0.86	High	0.62	CPP	0.83	44.68

*that were accessible (as of September 3, 2020) were used to benchmark the Mach peptides. The row labeled 'Model' notes the name of the webserver, and a brief summary has been given in the row 'Summary' (format: year of publication - model architecture - sequence size limitation). The unabbreviated forms of the model architectures are – SVM: Support Vector Machine, RF: Random Forest, and ERT: Extremely Randomized Trees. For CPPred-FL (*), there is no limitation on sequence size, however the prediction is done for a window of 40 residues. Pred denotes prediction, upt denotes uptake and conf denotes confidence.

We then investigated the CNN model's ability to extrapolate from the training dataset and found that experimental activity above a threshold of ~8 is necessary to accurately predict peptides with activity beyond that of the training dataset (Fig. 2.3, Table 2.2). In order to determine which factors of the training set are required for accurate prediction, we tested the predictor's accuracy when trained on datasets with increasing activity. We trained CNN models with the same architecture as the optimized model, with activity thresholds increasing by 0.5, starting at 0.5 until 19. Trained models are able to extrapolate activities beyond the training data only once the training data reaches activity around 8-fold over PMO. The performance then continues to increase with subsequently increasing activity thresholds. RMSE for held out test dataset consisting of sequences having higher activity than threshold increases at first, and then decreases, indicating a barrier for learning. RMSE for Mach sequences continues to decrease as the number of training data points and threshold increases, indicating that extrapolation for high activity sequences (such as Mach

sequences) requires a wide range of training data. Including even higher activity sequences in future rounds of model training would potentially lead to more accurate predictions in future work. The experiment notes that activity has a sequence dependence that the model is able to learn, once meeting an activity threshold in the training data.

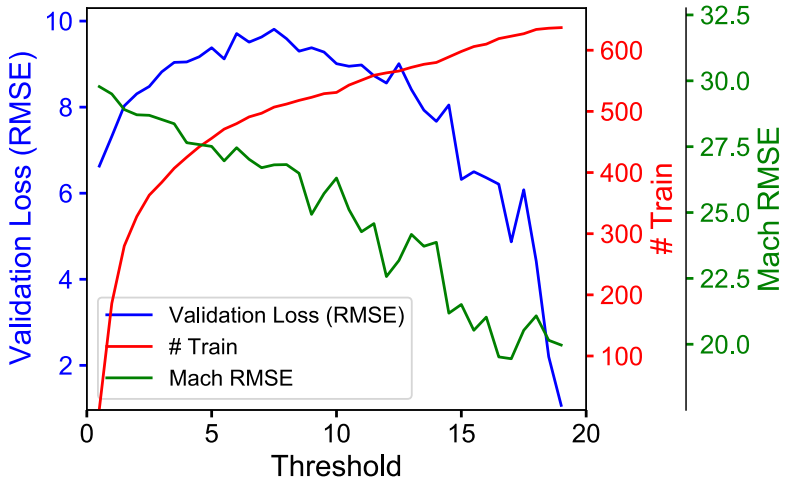


Figure 2.3 Threshold for extrapolation. The validation loss (RMSE), in blue, for the held-out dataset of sequences having higher activity than the ones used for training the model goes through a barrier. RMSE for Mach sequences, in green, is decreasing with increasing threshold. The number of training datapoints is indicated in red.

Table 2.2 Threshold for extrapolation*

Threshold	#Train	#Test	uRMSE	RMSE	MachRMSE	% Error	R^2	Pearson
0.5	13	627	1.34	6.63	29.78	34.53	-0.79	-0.08
1	186	454	1.48	7.32	29.5	38.08	-1	-0.09
1.5	280	360	1.62	8.03	28.91	41.82	-1.43	-0.33
2	328	312	1.68	8.31	28.71	43.28	-1.72	0.03
2.5	363	277	1.71	8.48	28.69	44.14	-1.98	-0.04
3	384	256	1.78	8.82	28.53	45.91	-2.38	-0.26
3.5	407	233	1.83	9.04	28.37	47.08	-2.8	-0.19
4	425	215	1.83	9.05	27.65	47.11	-3.07	-0.08
4.5	442	198	1.85	9.17	27.58	47.73	-3.53	0.02
5	456	184	1.9	9.38	27.51	48.84	-4.16	-0.03
5.5	471	169	1.84	9.12	26.96	47.48	-4.46	0.11
6	480	160	1.96	9.71	27.46	50.56	-5.72	0.04

6.5	491	149	1.92	9.51	27.02	49.49	-6.29	-0.04
7	497	143	1.95	9.63	26.7	50.15	-7.08	0.1
7.5	507	133	1.98	9.81	26.81	51.05	-8.61	-0.07
8	512	128	1.94	9.59	26.82	49.95	-8.99	-0.04
8.5	518	122	1.88	9.3	26.49	48.43	-9.34	0.11
9	523	117	1.9	9.38	24.93	48.85	-10.38	0.14
9.5	529	111	1.88	9.28	25.72	48.32	-11.27	0.12
10	531	109	1.82	9.01	26.31	46.92	-10.92	0.14
10.5	543	97	1.81	8.95	25.11	46.6	-13.17	0.22
11	551	89	1.82	8.98	24.27	46.73	-16	0.17
11.5	559	81	1.76	8.73	24.58	45.42	-19	0.11
12	563	77	1.73	8.56	22.57	44.57	-21.05	0.13
12.5	566	74	1.82	9.01	23.18	46.9	-25.47	0
13	572	68	1.7	8.41	24.17	43.78	-26.83	0.03
13.5	577	63	1.6	7.93	23.72	41.26	-27.35	0.23
14	580	60	1.55	7.67	23.87	39.92	-27.54	0.27
14.5	589	51	1.63	8.05	21.18	41.92	-36.2	0.24
15	598	42	1.28	6.32	21.51	32.89	-27.91	0.26
15.5	606	34	1.31	6.5	20.53	33.83	-38.67	0.27
16	610	30	1.29	6.36	21.03	33.11	-41.49	0.32
16.5	619	21	1.26	6.21	19.52	32.33	-57.81	-0.04
17	623	17	0.99	4.87	19.45	25.38	-45.61	0.65
17.5	627	13	1.23	6.08	20.53	31.66	-76.13	0.59
18	634	6	0.89	4.42	21.08	23.03	-75.21	0.66
18.5	636	4	0.45	2.2	20.15	11.47	-39.37	0.88
19	637	3	0.22	1.07	19.97	5.57	-44.72	-0.99
CNN-FP	512*	128*	0.41	2.03	10.55	35.53	0.83	0.92

*The evaluation metrics for models trained with sequences filtered by activity threshold have been noted. The original CNN model has been highlighted in grey color.

Using fingerprint representations is more effective than typical one-hot encodings at using inherent chemistry to predict novel sequences and is able to predict activities of sequences containing a new residue not in the training set. In order to demonstrate the advantage of the fingerprint representations in giving the model flexibility in predictions, we trained CNN models by leaving one residue out then having it predict activity of all sequences (Table 2.3). Models trained with sequences leaving one residue/linker out are able to compensate for the missing residue to a certain extent. Models were trained with sequences without a particular residue/linker

and evaluated against the test dataset with sequences containing the residue/linker. This demonstrates that the model is able to infer chemical rules, such as the similarities between glutamic acid and glutamine, and aspartic acid and asparagine. It is hypothesized that the performance in some cases may be marred by a lack of sufficient training data points for the maximum similarity residue. The absence of similar sequence motifs, in the case of replacement by maximum similarity residue, may also be a contributing factor in the performance of the models. These results suggest that the model is able to use information learned from other amino acids.

Table 2.3 Evaluating predictions when removing one residue from training *

Residue/ Linker	Chem. Simil.	#Train	#Test	uRMSE	RMSE	Mach RMSE	% Error	R^2	Pearson
I	V, 0.70	114	526	1.41	6.95	20.79	36.17	-2	0.3
E	Q, 0.66	456	184	1.39	6.86	24.49	35.68	-0.35	0.54
A	S, 0.76	150	490	1.24	6.14	17.93	31.95	-0.99	0.42
R	Q, 0.51	12	628	1.16	5.76	28.62	29.96	-0.34	-0.12
L	I, 0.63	74	566	1.14	5.66	22.77	29.44	-0.23	0.12
K	L, 0.62	25	615	1.14	5.64	19.86	29.35	-0.26	0.24
Q	E, 0.66	221	419	1.14	5.64	21.74	29.34	-1.19	0.31
S	A, 0.76	186	454	1.1	5.44	27.79	28.32	-0.09	0.18
G	A, 0.62	172	468	1.1	5.42	25.35	28.21	-0.02	0.39
P	Q, 0.47	143	497	1.07	5.29	24.8	27.52	-0.09	0.28
H	F, 0.27	402	238	1.05	5.2	22.31	27.07	-1.2	0.18
3	W, 0.18	64	576	1.05	5.17	27.17	26.91	-0.01	0.33
V	I, 0.70	44	596	0.99	4.88	25.93	25.42	0.08	0.56
2	C, 0.55	64	576	0.97	4.79	26.96	24.94	0.13	0.44
B	A, 0.55	592	48	0.92	4.57	24.62	23.8	0.55	0.87
X	K, 0.61	592	48	0.89	4.4	25.02	22.92	0.59	0.87
Y	F, 0.82	387	253	0.83	4.13	18.74	21.48	0.15	0.5
N	D, 0.54	267	373	0.79	3.91	21.33	20.37	0.09	0.38
C	2, 0.55	495	145	0.7	3.47	22.95	18.07	0.63	0.8
F	Y, 0.82	245	395	0.67	3.31	25.38	17.23	0.36	0.62
W	F, 0.35	323	317	0.65	3.23	22.65	16.81	0.36	0.66
M	V, 0.48	445	195	0.63	3.14	24.05	16.34	0.27	0.55
T	S, 0.57	343	297	0.62	3.07	22.13	15.99	0.46	0.7
D	N, 0.54	399	241	0.6	2.96	22.58	15.42	0.42	0.71
CNN-FP	-	512*	128*	**0.41**	**2.03**	**10.55**	35.53	**0.83**	**0.92**

*The evaluation metrics for models trained with sequences without a particular residue/linker have been noted, in decreasing order of RMSE. The residue/linker in the dataset with maximum chemical similarity, evaluated using Tanimoto similarity over the fingerprints, has been noted in the column Chem. Simil.

Next, we evaluated the role and advantages of the generator in the model's ability to predict high predicted-activity sequences. We conducted five in silico experiments comparing the generator to other methods of generating seed sequences (Fig. S2.4, Table S2.4). The predictor-optimizer loop was seeded with 50 sequences sampled from random sequences from the predictor training dataset (CPP Library), the 50 most active sequences from the predictor training dataset (CPP Library Top50), the CPP thesaurus (CPPsite 2.0), randomly generated sequences with equal likelihood for all amino acids at all sites, and sequences sampled using the generator as reported in the main text.

Our three criteria for optimized sequences are: high predicted activity, low similarity, and low Arg content. The optimized sequences have varying ranges of these characteristics. The top 50 sequences from the predictor dataset receive a head start in terms of activity, resulting in the highest predicted activity, followed by the generator-sampled, CPP library, and CPP thesaurus and random sequences. On the other hand, the maximum and mean similarities for sequences optimized using the seeds from the Top50 and full CPP Library are higher than for the sequences optimized using generator-sampled, CPP thesaurus, and randomly generated seeds. Finally, the generator-sampled, CPP thesaurus, and random sequences resulted in optimized sequences with lower Arg content than sequences from Top50 and full CPP library. Taken together, sampling seeds from the generator is a more favorable option for meeting our three criteria. However, we note that with appropriate diversity constraints and predicted activity thresholds, it is possible to sample sequences using other routes and still predict sequences with the desired characteristics. It is possible for the other methods of seed selection and optimization to also produce optimal peptide sequences, but experimental validation is required to adequately compare these methods.

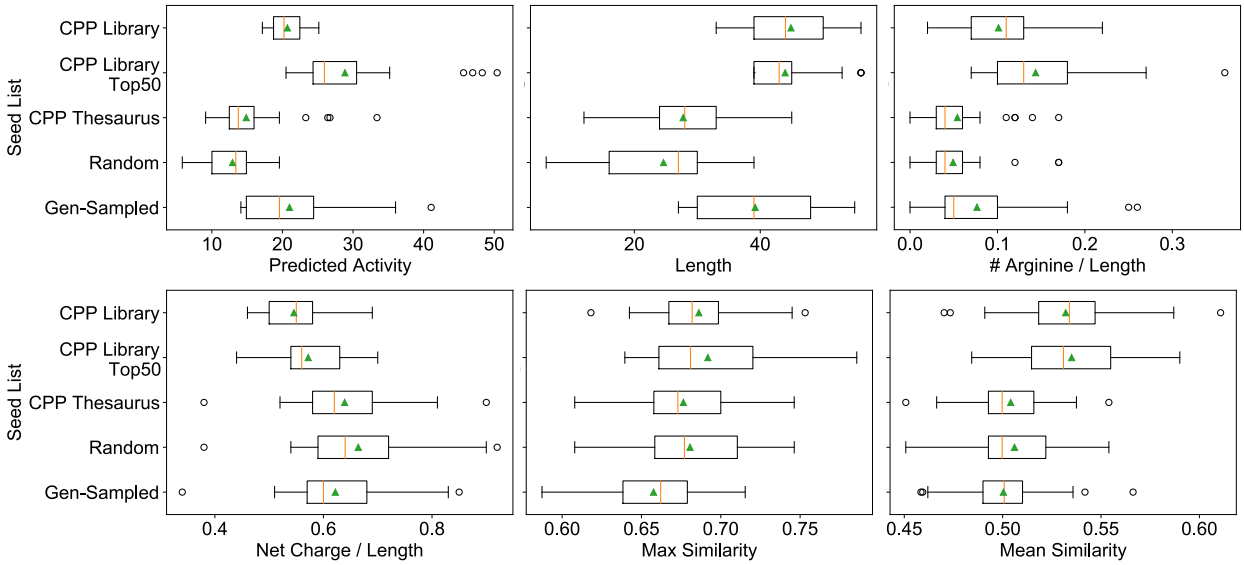


Figure 2.4 Box plots comparing seed sequence generation Box plots comparing optimized sequences comparing random seeds, randomly selected sequences from the predictor dataset, and seed sequences sampled from the generator dataset. For the box plot, the box marks the interquartile range (IQR), Q1 and Q3; the whiskers are at Q1-1.5*IQR and Q3+1.5*IQR; the orange line is the median; the green triangle is the mean, and outliers, if outside the whiskers, are marked as dots.

Table 2.4 Optimization of seed sequences from different generators

Statistic	Seed List	Predicted Activity	Length	# Arginine/ Length	Net Charge/ Length	Maximum Similarity	Mean Similarity
Mean	**CPP Library**	20.70	44.88	0.10	0.55	0.69	0.53
	CPP Library Top 50	28.86	43.95	0.14	0.57	0.69	0.54
	CPP Thesaurus	14.87	27.76	0.05	0.64	0.68	0.50
	Random	12.92	24.63	0.05	0.66	0.68	0.51
	Gen-Sampled	21.00	39.2	0.08	0.62	0.66	0.50
Median	**CPP Library**	20.23	44	0.11	0.55	0.68	0.53
	CPP Library Top 50	25.94	43	0.13	0.56	0.68	0.53
	CPP Thesaurus	13.74	28	0.04	0.62	0.67	0.50
	Random	13.40	27	0.04	0.64	0.68	0.50
	Gen-Sampled	19.58	39	0.05	0.60	0.66	0.50
Minimum	**CPP Library**	17.16	33	0.02	0.46	0.62	0.47
	CPP Library Top 50	20.51	39	0.07	0.44	0.64	0.48
	CPP Thesaurus	9.11	12	0.00	0.38	0.61	0.45
	Random	5.81	6	0.00	0.38	0.61	0.45
	Gen-Sampled	14.12	27	0.00	0.34	0.59	0.46
Maximum	**CPP Library**	25.15	56	0.22	0.69	0.75	0.61

	CPP Library Top 50	50.42	56	0.36	0.70	0.79	0.59
	CPP Thesaurus	33.40	45	0.17	0.90	0.75	0.55
	Random	19.57	39	0.17	0.92	0.75	0.55
	Gen-Sampled	41.07	55	0.26	0.85	0.72	0.57
Q1	**CPP Library**	18.74	39	0.07	0.50	0.67	0.52
	CPP Library Top 50	24.34	39	0.10	0.54	0.66	0.51
	CPP Thesaurus	12.47	24	0.03	0.58	0.66	0.49
	Random	10.01	16	0.03	0.59	0.66	0.49
	Gen-Sampled	14.89	30	0.04	0.57	0.64	0.49
Q3	**CPP Library**	22.47	50	0.13	0.58	0.70	0.55
	CPP Library Top 50	30.51	45	0.18	0.63	0.72	0.55
	CPP Thesaurus	15.98	33	0.06	0.69	0.70	0.52
	Random	14.89	30	0.06	0.72	0.71	0.52
	Gen-Sampled	24.42	48	0.1	0.68	0.68	0.51

We next evaluated the role of constraints within the optimizer-predictor loop. We tested the model with varying combinations of constraints to probe the role of each constraint in sequence optimization (Table S2.5). 5 seed sequences with variable length (10, 20, 30, 40, 50) were used to seed the predictor-optimizer loop, where the optimizer had none to all constraints – maximization of predicted activity, minimization of similarity, minimization of Arg content, minimization of length, and maintenance of net charge for water solubility. Removal of a constraint leads to clear change in the optimized sequences. For instance, when minimization of Arg content is not a constraint, the sequences have a high degree of net Arg. Such sequences have been shown to be toxic in vivo and are already a known cell-penetrating motif. A goal of this work was to generate unique high-activity sequences that do not rely on Arg for activity.

The constraints are also necessary to shift away from the bias in the training dataset, since without them, the predicted sequences would appear to be very similar to the sequences in the training dataset. This dataset was created using a combinatorial approach with the currently known cell-penetrating peptides. The peptides in the training dataset inherently have a net high Arg content and longer length than desired. The optimizer also makes mutations to the seed sequences using motifs from the training data, so without minimizing similarity, the end sequences would resemble the training sequences. The constraints help to reduce the bias present in the dataset, and

optimize sequences towards desired properties and away from those in the training dataset such that we are more likely to discover new sequences and motifs.

Table 2.5 Sequences optimized with varying constraints.

Constraints	Sequences	Intensity	Length	Relative Arg	Relative Charge
Seeds	KHAPRRESSW	2.50	10.00	0.20	0.28
	RWTAWTLRRIAKAVGPIVRR	2.71	20.00	0.25	0.30
	SCRRPQRKDVLTIAHRSRNRIRGAHARPNR	4.60	30.00	0.30	0.35
	GKEKQSWRRFQRKTPRSAAQMRAKRALARARLQLSRSQRR	3.89	40.00	0.28	0.35
	RSSHHGCARSPRLRRHKRRKPIKVRLRRRMKLELKKTARKRKSRRRGLHC	2.78	50.00	0.32	0.52
MFI	RRRRRQRRRRRR	11.53	12.00	0.92	0.92
	RKRRRQRKRRRRWPXRXIPQYDQXF	14.40	25.00	0.40	0.44
	KKKRPQLKRRRRGPMRXCSEFDFHFPRPTK	14.45	30.00	0.23	0.36
	GKKRRSRRRRRRGPKGGVPQPSQGYPKYSBNRXRRRRRX	28.31	39.00	0.36	0.46
	RRRRRLLKRRRRKGKKXLPKFREGYPLGLKPRKRRQRRRYRWGRGKHRTWW	26.90	51.00	0.37	0.53
MFI, Length	RRRRRQRRRRRR	11.53	12.00	0.92	0.92
	RRRRRQGKRRRRGPRGKVPEPPQHSPKY	15.53	28.00	0.36	0.46
	RKKRRQRKRRRRGPMGKRSRPSQGYALYLK	16.03	30.00	0.33	0.50
	BKKKNSBBKRRRWRGKNAPQPKAKYPLWILRRRRRQRGRYRR	21.28	42.00	0.31	0.48
	XRRRRLLRRLRRNPGRGRLRVIFGRKRGAANRXRRMRXRGPWARKRHXRW	33.74	50.00	0.42	0.48
MFI, Arg	RRRRRQRRRRRRWPMG	12.50	16.00	0.69	0.69
	RRRRRQRKRRRRWCKKGIPE	13.14	20.00	0.50	0.60
	RRKRPQERRRRRGLNRXCSEPPQHYAIYCK	14.16	30.00	0.30	0.32
	GKRKKSBLKKRRALKXRRBKAKAGQRQYALKRXRRQRXRLPRWR	25.83	44.00	0.30	0.50
	RRNRRENKRRRRGLBMALPRPAEGYLLRLINKRRLQRRRGPWARXRKXRW	32.84	50.00	0.36	0.38
MFI, Charge	RRRRRQRRRRRR	11.53	12.00	0.92	0.92
	RRRRRQRKRRRRGPMGXCPRP	14.19	21.00	0.52	0.57
	RRRRRQRRRRRRGPGGGNPRPEQHVPDFLBG	16.10	31.00	0.39	0.35
	GKKRRQRRRRRRGPNNKNPQFSQKYPQPPRXKRRRRRRR	24.81	39.00	0.41	0.54
	RRLRRLRLRRRRYLRGKLLQKKVKYKQGLRBRRRRQRXRAPRKPRRRRKRWCR	38.53	53.00	0.45	0.58
	RRKRRQ	6.66	6.00	0.67	0.83

MFI, Length, Arg	RRKRRQRKRRRRGPKKGVPQ	14.32	20.00	0.45	0.65
	RRRRRQRPKKRRGPLRGCPQFRQHFLQYL	15.43	29.00	0.34	0.44
	BRRRRQRKRRRRYRBKGIPQPREKYLQYLIRXXKRQRXRRRR	26.89	42.00	0.43	0.50
	RRGRRLRKLRRRWRGRRRAKPRLGYPRYADRRRRRERRRRRYWRQKHXRW	29.25	50.00	0.52	0.56
MFI, Arg, Charge	RRRRRQRRRRRRWP	12.32	14.00	0.79	0.79
	BKKRRQRRRRNRWRGKNCPQPSLSYAMY	14.31	28.00	0.29	0.39
	WRKRPQRKRRRRWPKKADPQPAQBVAQPLBGRX	17.67	33.00	0.24	0.33
	GRKKRQRLKRRRGPMRGKPQPSSKYPRYSKKXRRLQRRX	23.06	39.00	0.31	0.49
	RRRRRRLRRRRRRPGNALARADQDYLAYVLNRGRRRRRXACBCRXLHW	26.00	48.00	0.42	0.39
MFI, Length, Charge	RRRRRQRRRRRR	11.53	12.00	0.92	0.92
	RRRRRQEKLRRRGPNKGIPQPSQHYPIYLLG	16.94	31.00	0.26	0.32
	RRRRRQRKKRRRGPLGGGLQFKEGVPQYVQNRX	18.48	33.00	0.30	0.36
	RKRRKSRRRRRRRPNGGRSQPEQXYLLPTBXRRRRRKXXRRRW	27.35	43.00	0.44	0.49
	RRRRRKALKLLRYPKKINLQPREKQPQWLAKKRRRRRRXRRRWRXRRWRWXCGRXM	53.09	56.00	0.38	0.48
MFI, Length, Arg, Charge	HRKRRQ	6.70	6.00	0.50	0.80
	RRRRRQRKRRRRWRGGGVPRPSQBQPV	14.73	27.00	0.44	0.48
	RRKRRQRRRRRRGGKBGBPIPIQXVPQYLIRXXRRBR	20.80	37.00	0.38	0.43
	KKRRKQAKKRRRNPKKNNPQFDFHFPRPTLXRGRRKGRX	23.29	39.00	0.26	0.46
	RRLRRSGLKSRRGLLGKSSQPSKGRRLPSKKRGKLLKGXGLWGRGKRXTWWCM	30.39	53.00	0.21	0.36

Finally, inclusion of unnatural amino acids was required for high-activity predictions, as predicting canonical sequences using the same model resulted in a significant drop in predicted activity (Fig. 2.5, Table 2.6). We determined that sequences with comparable activity could not be achieved using only canonical residues. We optimized peptides containing only canonical residues by constraining the optimizer to use only canonical residues for mutations. 50 seed sequences sampled from generator were used to seed the predictor-optimizer loop. While we can predict fully canonical peptides, the predicted activities of these peptides are significantly lower than those containing noncanonical residues. Given the constraints in the optimizer (minimization of length

and Arg content), we observe diminishing returns of length versus Arg content, where shorter sequences have more Arg in order to have a high predicted MFI.

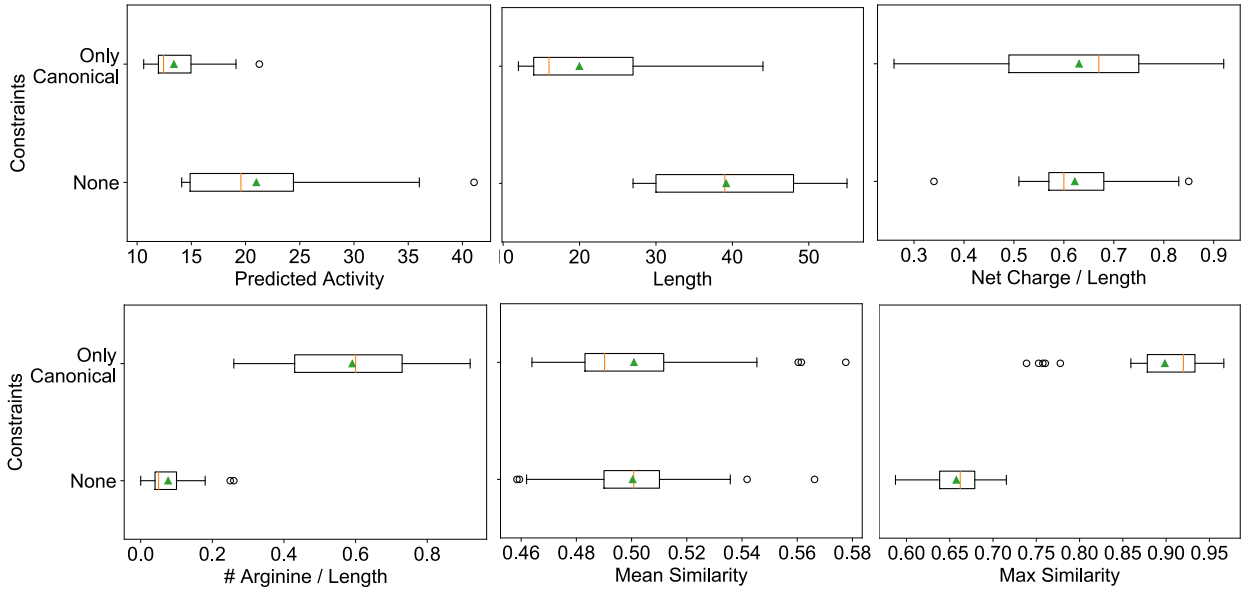


Figure 2.5 Box plots showing predicted activities with and without unnatural residues Box plots comparing optimized sequences with and without the constraint of only being able to use canonical residues for the genetic algorithm mutations. For the box plot, the box marks the interquartile range (IQR), Q1 and Q3; the whiskers are at Q1-1.5*IQR and Q3+1.5*IQR; the orange line is the median; the green triangle is the mean, and outliers, if outside the whiskers, are marked as dots.

Table 2.6 Optimized sequences containing only natural residues

Sequence	Predicted MFI	Length	Net Arg	Net Charge
PRKKRRSRRRRRRRLRGDPQPPQGRKIYVLG TRRLQRRRGPWRPRRRGRR	21.62	50	0.46	0.5
KRRRRQIRRRRRYRLRNVLQPEQMRKQGLL GRRRRQLRRYPYRR	21.27	44	0.45	0.48
ARKRRQRKRRRRWPMGANLVFSLHYAQYT KGRRRRRR	19.12	37	0.38	0.48
RRRRRQRKRRRRGPGGGDPEPAQGYPI	16.18	27	0.37	0.33
KAKRRQRRRRRRGPNGGDPRPSQHYPD	16	27	0.33	0.36
RRRRRQRRRRRRGPQKPCPQPSQKYA	15.85	26	0.42	0.5
RRRRRQGKRRRRWRNRGCPQPDQKYPDYC	15.51	29	0.38	0.38
RRRRRQRRRRRRWPQRPLPQPRQHILDYVN	15.36	30	0.43	0.43
RRRRREEKRRRRGPGGPCLQFSLSAPQYSK	15.2	30	0.3	0.3
RRRRRQRRRRRRWPMGKMPQPSQ	15.12	23	0.48	0.52
KKKRRQRRRRRRWRAKGIPEPSFKYKQPPH GR	15.04	32	0.31	0.49
RRRRRQRRRRRRWRGGPCPRPIQHIPQ	14.97	27	0.48	0.51
RRRRRQRRRRRRGRGGPRSQFSQHYPQ	14.6	27	0.48	0.51

RRRRRLGKRRTRGPLGPCPQFDEGILI	14.28	27	0.3	0.26
RRRRRQRPLRRRGPNKPCPEPDQ	14.17	23	0.39	0.35
KRRRREEKKKKKWRRGGCPRPRQHYPQYPKG	13.87	31	0.26	0.44
RRRRRLRKRRRRGPMGKCSD	13.02	20	0.5	0.55
RRRRRQRKRRRRGCNGNCPD	12.92	20	0.5	0.5
RRRRRQRKRRRRGPMGPC	12.79	18	0.56	0.61
RRRRRQRKKRRRGPMGPC	12.52	18	0.5	0.61
RRRRRQRRRRRRWPMG	12.5	16	0.69	0.69
RRRRRQRRRRRRGPM	12.44	15	0.73	0.73
RRRRRQRRRRRRWPM	12.41	15	0.73	0.73
RRRRRQRRRRRRWPG	12.36	15	0.73	0.73
RKRRRQRRRRRRGPM	12.35	15	0.67	0.73
RRRRRQRRRRRRGPG	12.35	15	0.73	0.73
RRKRRQRRRRRRWP	12.27	14	0.71	0.79
RRRRRQRRRRRRGP	12.27	14	0.79	0.79
RRRRRQRNRRRRWPM	12.23	15	0.67	0.67
RRRRRQRPRRRRWP	12.2	14	0.71	0.71
RRRRRQRKKRRRWPM	12.08	15	0.6	0.73
RRRRRERKRRRRWPMG	11.97	16	0.62	0.62
RRRRRQRKKRRRWP	11.96	15	0.6	0.73
RRRRRQRRRRRR	11.53	13	0.85	0.85
RRRRRQRRRRRR	11.53	12	0.92	0.92
RRKRRQRRRRRR	11.49	12	0.83	0.92
RRRRRQRKRRRR	11.48	12	0.83	0.92
RRKRRQRKRRRR	11.45	12	0.75	0.92
RKRRRQRRRRRR	11.44	12	0.83	0.92
RRRRRQRPRRRR	11.39	12	0.83	0.83
RRRRRQERRRRR	11.15	12	0.83	0.75
RRKRQQRRRRRR	10.61	12	0.75	0.83
RRRRRLRRRRRR	10.6	12	0.92	0.92
RRRRRQRRRKRR	10.56	12	0.83	0.92
RRKRRQRRRRHR	10.51	12	0.75	0.9
HRKRRQ	6.7	7	0.43	0.68
HRKRRQ	6.7	6	0.5	0.8
RRRRRQ	6.61	6	0.83	0.83
HRKRRE	5.91	6	0.5	0.63
RRKRRM	5.42	6	0.67	0.83

Once the predicted sequences were acquired, we evaluated how similar they were compared to the training library. Similarity among sequences in each training and validation dataset was analyzed using Jaro-Winkler distance metric (Fig. 2.6).[41] Each sequence was compared with the rest of the library to evaluate the string similarity. The sequences used to train the generator have a mean similarity of 47%, indicating that we capture a combinatorial chemical space of cell-penetrating peptide sequences. For the sequences used to train the predictor, composed mostly of the modular library, the mean similarity is 66%. The modularity of the sequences from the library can be seen clearly in the visualization of sequence similarity. The four highlighted squares along the diagonal correspond to module 2 of the sequences. Similarly, the four lighter colored boxes correspond to module 3. The non-modular sequences, which are dissimilar from one another, are on the bottom of the visualization.

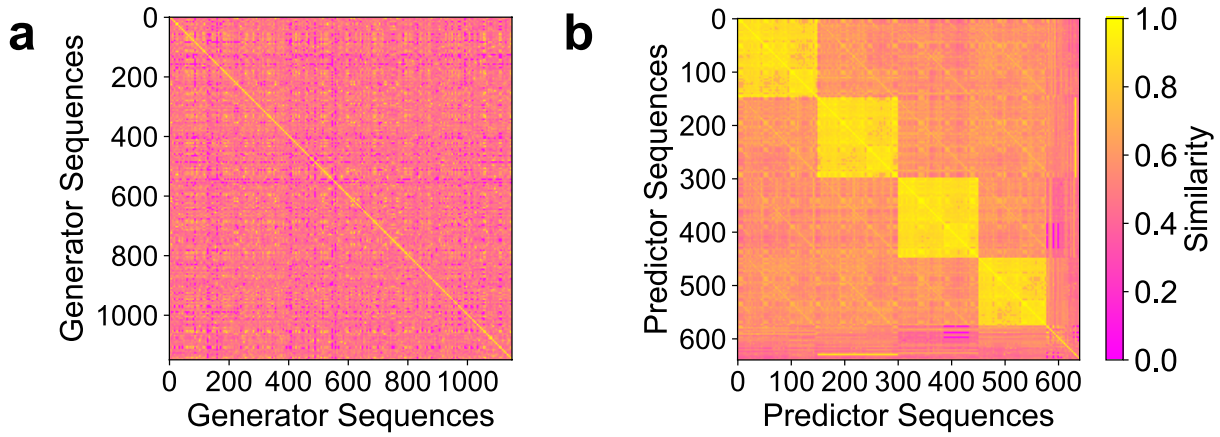


Figure 2.6 Similarity of sequences used in training of generator and predictor. Similarity of sequences used in the training of generator and predictor. Each sequence used in training of (a) generator (Nested LSTM) and (b) predictor (Convolutional Neural Network based model) is compared with the rest of respective training dataset. The mean similarities of the sequences are 47% and 66% for the generator and predictor respectively. The heatmap for the predictor sequences have a modular pattern owing to the combinatorial nature of the library. Jaro-Winkler distance was used as the metric to assess the similarity between two sequences

2.2.4 Interpreting the Predictor model

We interpreted the predictor CNN by visualizing the residue substructures that are important in its decision-making process. This type of visualization was a longstanding attribution challenge that was recently addressed for image classification and more recently with small molecule design.[42–44] We developed an analogous tool to correlate the input sequence representation with predicted activity. This process generated bit-wise positive and negative activation values for each chemical substructure in the sequence. Bits with higher activation indicated the features that most strongly influence the final activity prediction.

Using the conceptual attribution framework developed to understand activation of neural networks for image classification, we developed a toolkit to visualize the decision making process of the CNN model.[42] We chose the first convolution layer of the model to access the fingerprint indices. Taking the first derivative of the model output (normalized fluorescence intensity) with respect to the input representation (row matrix of fingerprints), produces a Jacobian matrix of partial derivatives. We performed element-wise multiplication of the Jacobian with the input representation to zero out the activation of absent chemical features, and clipped negative values, to focus on features that drive high MFI. We analyzed the role of individual activated fingerprints by visualizing the corresponding chemical substructure, and also obtained the average activation over the residue positions and fingerprint indices.

As an example, for the predicted Mach3 sequence the two C-terminal aminohexanoic acid (Ahx) residues were the most positively activated (Fig. 2.7a), followed by arginine (Arg). The alkyl backbone in Ahx was the most activated substructure (Fig. 2.7b). A similar trend was observed for active sequences and substructures in the training dataset (Fig. 2.8, 2.9). We used this visualization approach to better understand how the trained model designed sequences. We chose five random sequences of different lengths, seeded them in the predictor-optimizer loop to maximize activity contingent upon other design constraints, and visualized the activations for the best predictions. Again, a higher activation can be seen for C-terminal residues (Fig. 2.7c), most likely due to the attachment of PMO to the N-terminus. We also observed that the general composition of charged and hydrophobic residues remained unchanged across different sequence lengths (Fig. 2.7d). Particular residue fingerprints were activated irrespective of the sequence length, such as the side chains of Lys, Ser, and Asp (Fig. 2.7e-f). Consistent with earlier observations, a strong preference for polar and charged side chains as well as for Ahx was evident.

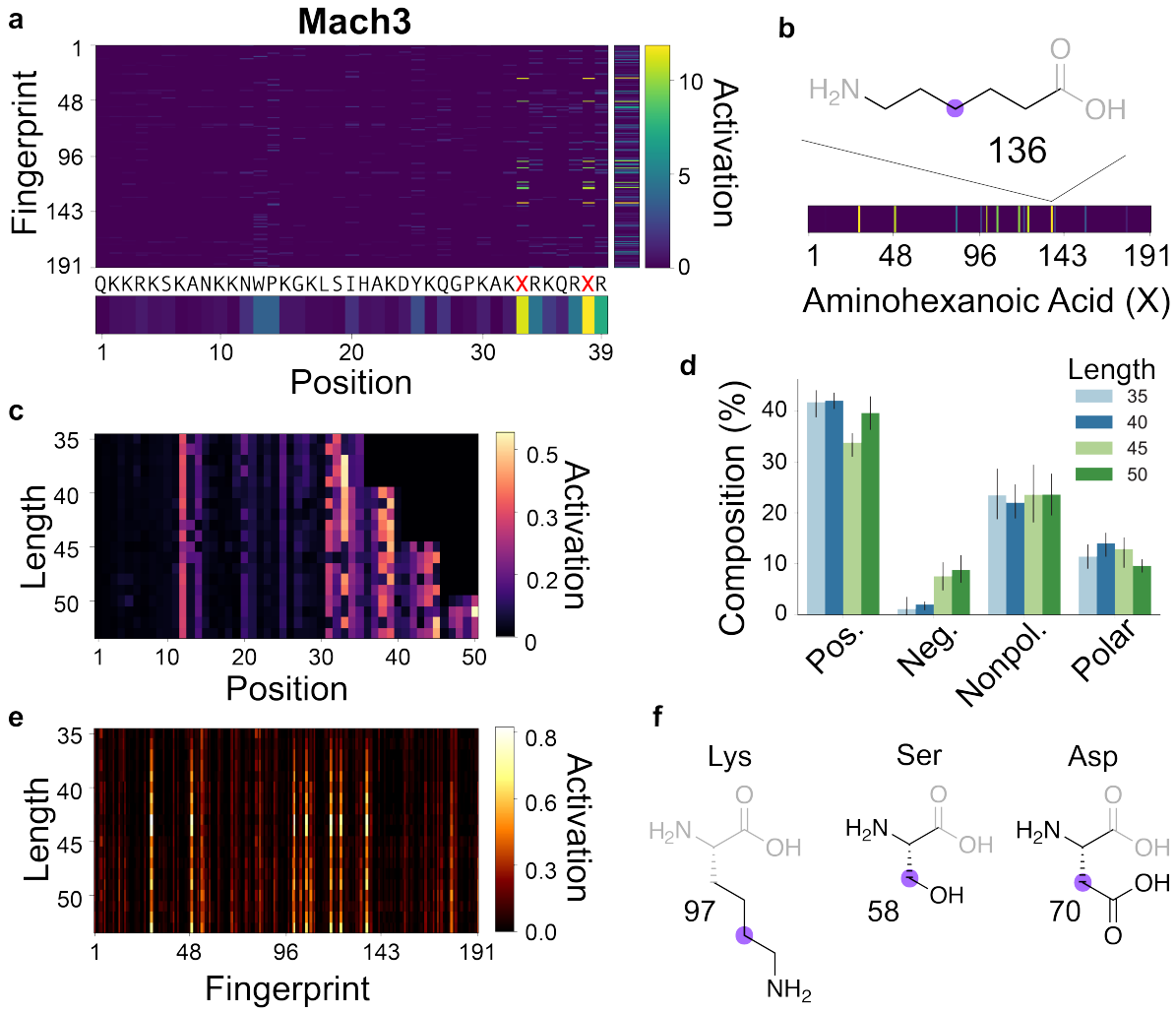


Figure 2.7 Interpretation of predictor CNN unveils activated substructures. (a) CNN positive activation gradient map was calculated for input sequence representation of Mach3. The averaged activation values over fingerprint indices and residue positions are shown. Fingerprint index represents a corresponding substructure. (b) The activation gradient map of aminohexanoic acid in Mach3 indicates the activated substructures of this residue. The alkyl backbone substructure (136) is shown. (c) Gradient maps of predicted sequences with lengths 35, 40, 45 and 50 are shown relative to residue position. (d) Percent composition of each type of residue (positive, negative, nonpolar, and polar) relative to predicted sequences with lengths 35, 40, 45 and 50 is shown. (e) Gradient maps of predicted sequences with lengths 35, 40, 45 and 50 are shown relative to substructure fingerprints. (f) Several residues and substructures that are consistently activated across all sequence lengths are shown, including the amine side chain of Lys, polar side chain of Ser, and the carboxylic acid side chain of Asp.

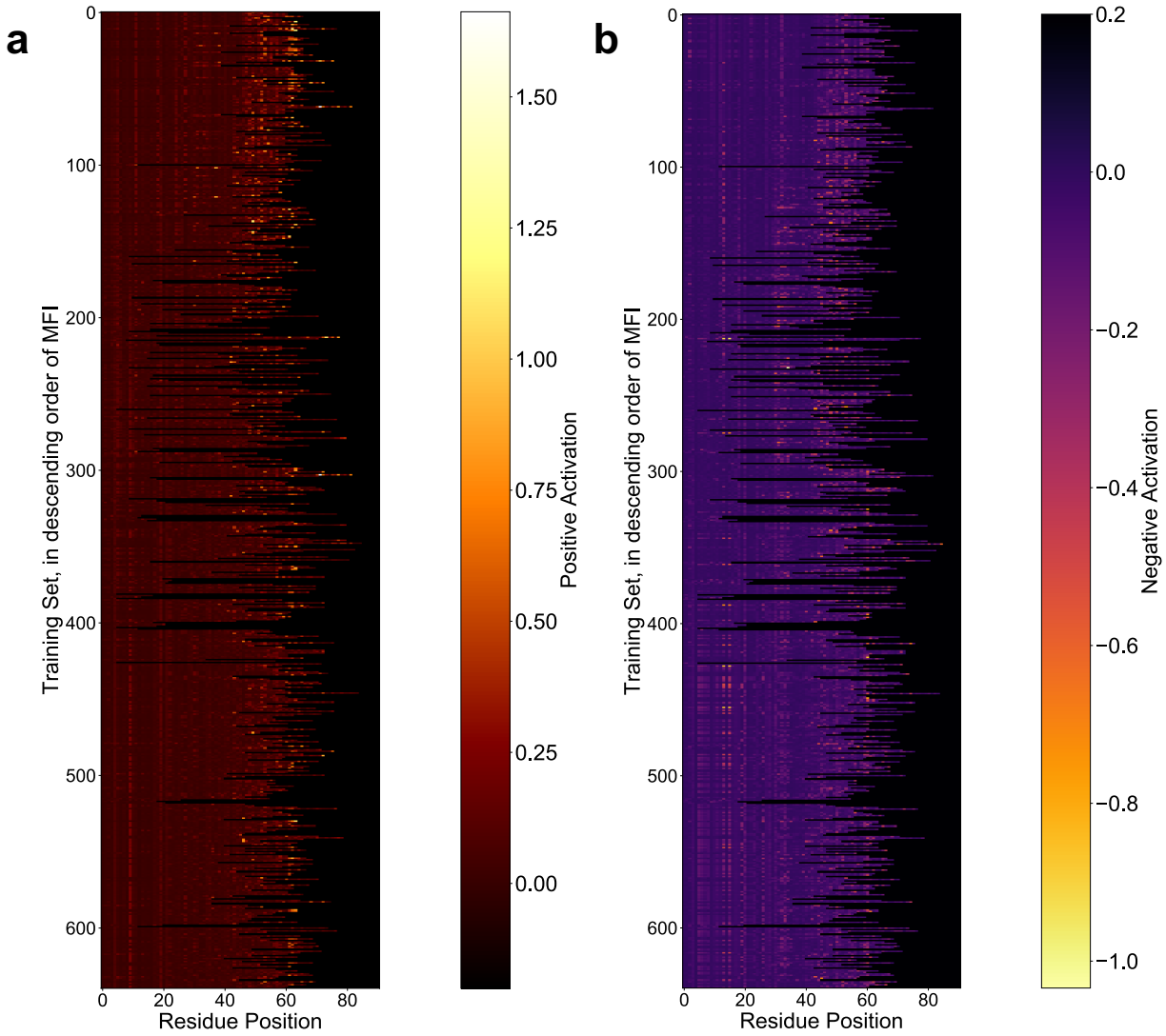


Figure 2.8 Activation map of predictor training set relative to amino acid position. Gradient activations for sequences are arranged in descending order of experimental normalized MFI for (a) positive and **(b)** negative activation averaged over residue position. The positive activation for C-terminal residues decreases with decrease in normalized MFI values. The most active sequences have a highly positively activated C-terminus and a sparsely negatively activated C-terminus.

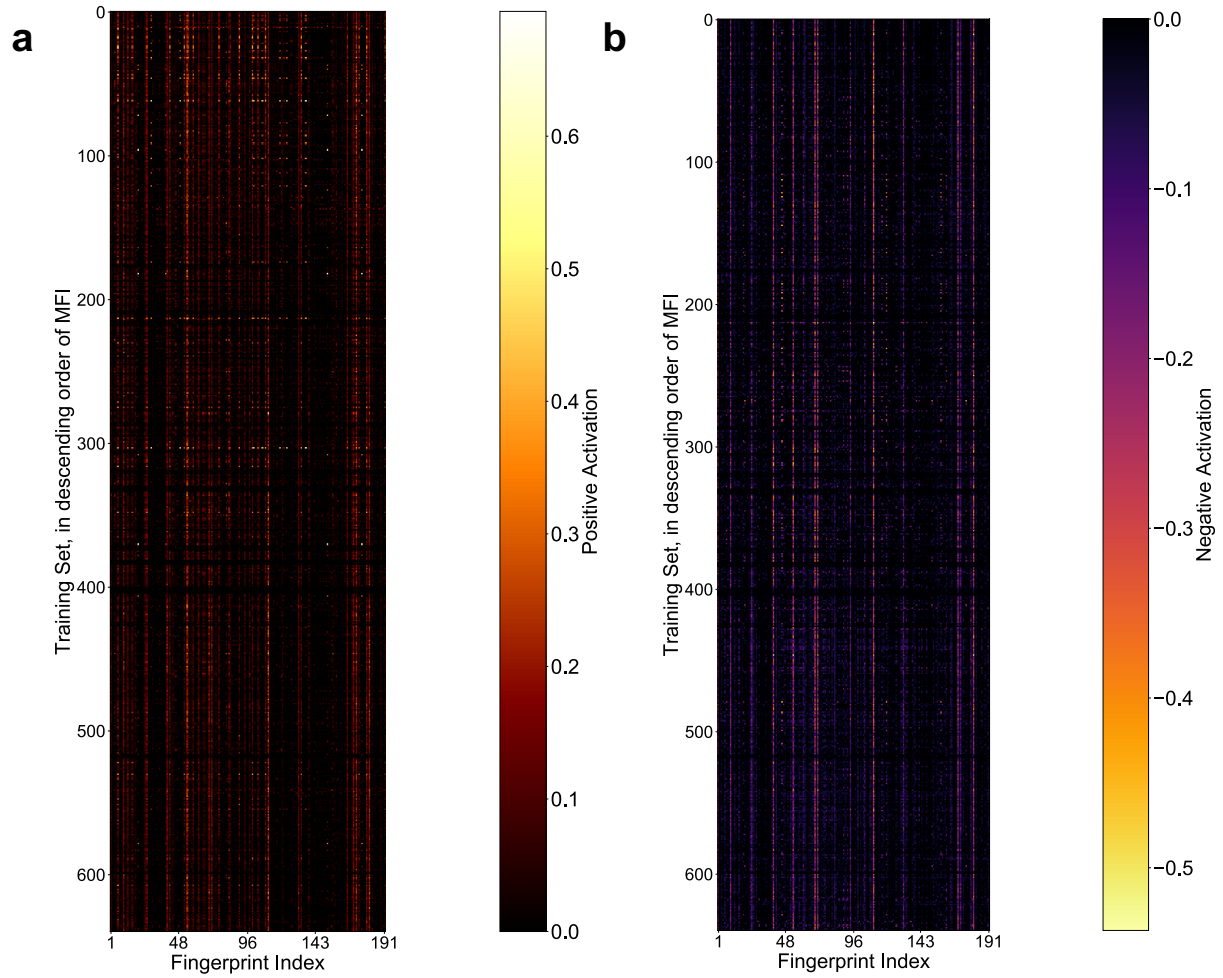


Figure 2.9 Activation map of predictor training dataset relative to fingerprint index. Gradient activations for sequences are arranged in descending order of normalized MFI for (a) positive and **(b)** negative activation averaged over fingerprint index. The most positively activated substructures by residue are for aminohexanoic acid, β-alanine, aspartic acid, threonine and serine.

2.2.5 Mach Miniproteins enhance PMO delivery

We synthesized and characterized twelve candidates from hundreds of miniproteins predicted by the model, selecting diverse sequences and predicted activities. Mach1, 2 and 6 were selected because they had high predicted activity among 50-mer sequences. Mach3 was selected as a mid-length peptide (39 residues), Mach4 was selected as a shorter sequence (33 residues) with only two Arg residues, and Mach5 was selected because it was predicted to have moderate activity while having the lowest net charge (10.5). Mach7 was initially designed to be a negative control—where the sequence of Mach1 was rearranged until the model predicted the lowest activity. Mach8 and 9 were selected from a list of much longer miniproteins (around 80 residues) and Mach12 and 13 were selected from sequences that contained Cys-linked macrocycles. Finally, Mach11 was

selected from a list of sequences for which the activity was optimized in the negative direction, to show that the algorithm could predict peptides of similar length, charge, and amino acid composition, but with no PMO delivery activity. Each candidate was synthesized using automated fast-flow solid-phase peptide synthesis, and when applicable, the two cysteine residues were connected with decafluorobiphenyl as previously reported (Fig. 2.10).[45,19] Conjugation of azido-Mach to PMO IVS2-654 was achieved in the same manner as in the library. The final PMO-Mach constructs are described in Table 2.7.

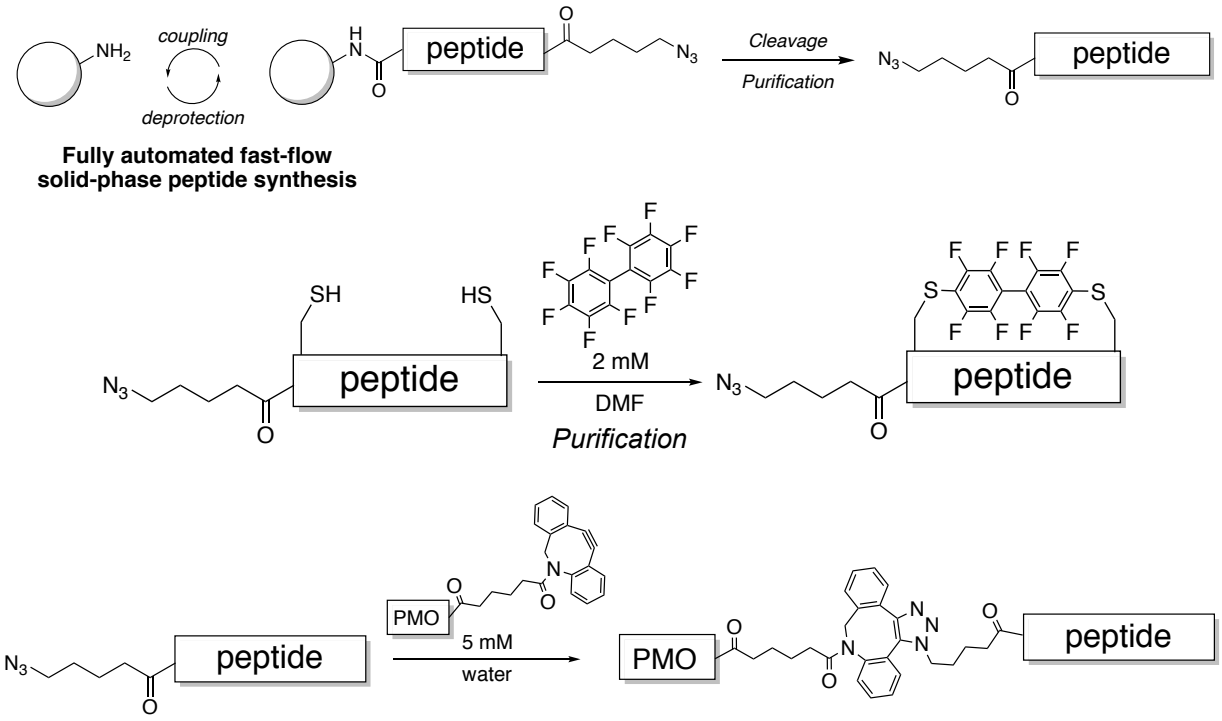


Figure 2.10 Synthesis route for Mach peptides. Synthesis route for Mach peptides. Predicted sequences were synthesized by fully automated SPPS, cyclized, and conjugated to PMO. Our synthesis technology can reliably synthesize long polycationic peptides in one-shot. If the predicted peptide contains a Cys macrocycle, cleaved and purified peptides were cyclized before being attached to PMO via copper-free click chemistry.

Table 2.7 List of Mach peptides*

		Fold over PMO	% Arg	PPMO MW	Net charge
Mach 1	ALKBRSAAKAVRWPKKKIKQASK KVAKYALXXXRKKKAASKXWLQ LHWPRW	45	8	12,645	18
Mach 2	PPLRNAKKKNLKNNLKMDPKFTK KVKQGALKLNRRKKNRGPKGPX KHWTT	27	8	12,499	18
Mach 3	QKKRKSKANKKNWPKGKLSIHAK DYKQGPKAKXRKQRXR	39	10	11,324	17
Mach 4	KKGKKQNKKKHRWPKKKVPQPK KMFKQGABXRX	25	6	10,622	16
Mach 5	AKKKIAKAKKHRGPNBGIHAPVS KIKDPLKXXX	3	8	10,222	11
Mach 6	ALKBRSAAKAVRWPKKAIKQASK KVAKYALKXXRKKKAASKXWLQ LHWPRW	43	8	12,603	18
Mach 7	XKHPXAVQBAARAWKVPAAALW KKKRLKKSSKQKKKWLWKARSA XKYXRLI	36	8	12,645	18
Mach 8	BKGKNLLAKIRRGPNGGNBQGSQ GYLLYLLXRXRRQRXXYPWWRX KHXRWXXRXRGHXRRRRQXLKP DRXRGGKGSVS	39	21	15,929	22
Mach 9	KKKKNLNBKSRRGPNGGALQPSQ GYLQPLNXRXRRQRXXYPWWRX KHXRWRXRYHXRRRRQXLKPG	38	21	14,845	22
Mach 11	TSNLKLHLAPPVKKKALKKPLYK AKKKKKVVSPTWXTDQEW	4	0	11,423	11
Mach 12	KGGKNLAKKIRRGPNGGALQPSQ GYLLYLBXRXRRQRXXGPXWRX KHXRWXXXXXRPTHXRRRRQXL **C**PGRXRP**C**RGSVS	40	20	16,285	22
Mach 13	AKKKKLGBKALRWPNGK**C**PQPK EK**C**PKYLLGRXRRKRXRYPWWR XKHRRW	30	18	13,228	20

*Peptide 10 was found to degrade in solution, so its analysis was discontinued. 'X' is 6-amino hexanoic acid, and 'B' is β-alanine. **C** residues are linked through decafluorobiphenyl.

The PMO-Mach constructs were then tested for PMO delivery in the HeLa 654 assay as was done with the library (Fig. 2.11). Nearly all sequences predicted to have activity greater than 20-fold did indeed surpass the highest performing modular library construct, with the exception of Mach5. Because the model is extrapolating outside the range of the training data, the predicted and experimental activity of PMO-Mach constructs shows greater % error than the test dataset (Fig. 2.12a). Physicochemical properties of validated predictions show little correlation with PMO activity. We compared the activities of Mach constructs to the training library in relation to various physicochemical properties (Fig. 2.12b-c). While library constructs clearly show an increase in activity with an increase in arginine content relative to length, and net charge relative to length, there is no obvious correlation between activity of Mach constructs and these same properties.

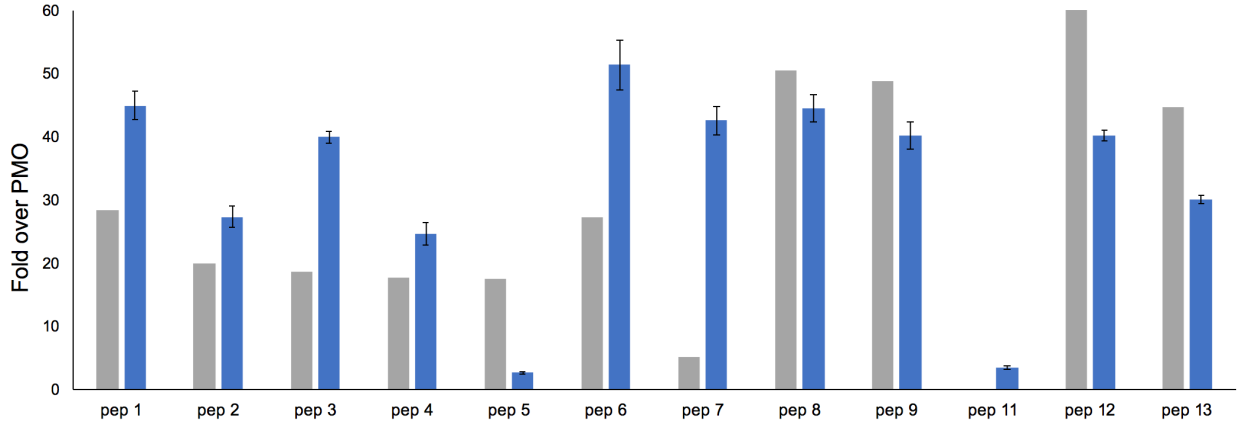


Figure 2.11 Experimental vs Predicted activity of Mach peptides. Mach peptides enhance delivery of PMO by 40-50 fold as determined by the HeLa 654 assay. Experimental activity (blue) is comparable to predicted activity (grey). Mach12 predicted activity is off the scale, at 140. Each bar represents group mean ± SD, N = 3.

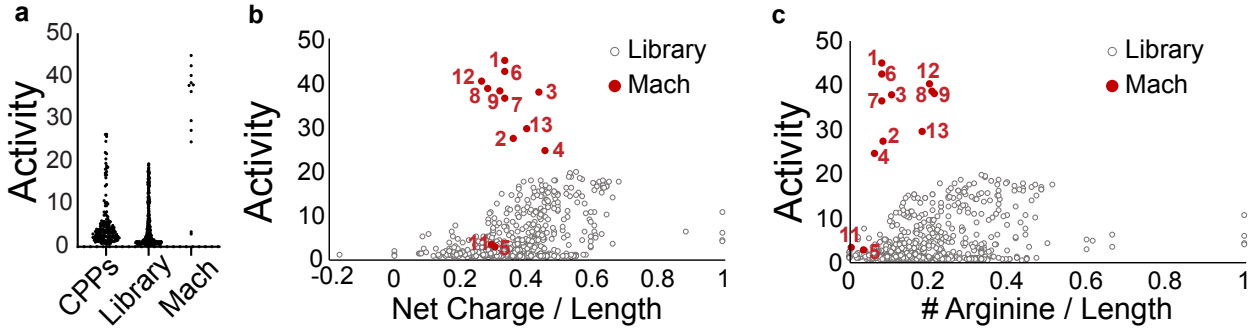


Figure 2.12 PMO-Mach miniproteins exhibit high activity with reduced Arg content compared to previously studied CPPs and the combinatorial library. (a) Dot plot showing PMO activity relative to PMO alone for CPPs, library, and Mach - 58 -iniproteins. (b) Dot plot showing PMO activity against net charge relative to length of PMO-Mach constructs (red) and library constructs (grey). (c) Dot plot showing PMO activity against number of arginine residues to length of PMO-Mach constructs (red) and library constructs (grey).

Because one of our goals was to optimize unique sequences, we compared the sequence similarity of Mach sequences to protein databases as well as to the training library. The Mach sequences were confirmed to be unique by comparing to library sequences as well as a protein database. Similarity of Mach sequences was first compared to the library using mean Jaro-Winkler distance (Fig. 2.13). All Mach sequences had a mean similarity less than 60% when compared to the training dataset. Then to compare Mach peptides to the existing proteome, we used BLASTp on the online server.[46] The search was done using default values to search the UniProt database. There was no sequence homology between Mach sequences and known proteins for significant E-values less than 0.01. For the unnatural residues, B (β-alanine) and X (aminohexanoic acid) were replaced by A (alanine) and L (leucine) respectively for the search operation. Mach sequences containing cysteine macrocycles were excluded from the search.

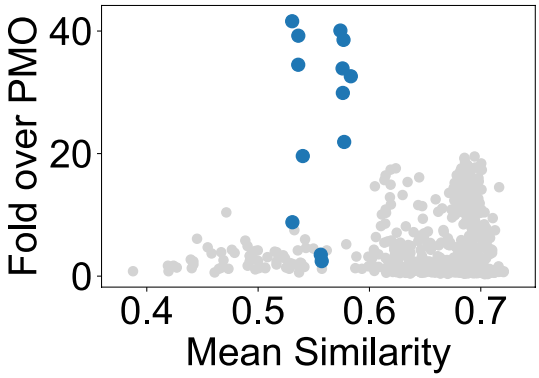


Figure 2.13 Similarity and experimental activity of Mach vs training sequences Similarity and experimental normalized MFI of Mach and training sequences. Mach sequences (blue) are novel and high-performing in comparison to the sequences used in the training of the predictor (grey). For each Mach sequence, the Jaro-Winkler distance with the predictor training dataset was averaged. For the rest of the training dataset, the mean similarity was calculated by averaging over the similarity with rest of the library. The mean similarities for all Mach sequences is less than 60%.

We then performed dose-response experiments to characterize activity in the EGFP assay and toxicity in a lactate dehydrogenase (LDH) release assay. PMO-Mach2, 3, 4, and 7 each had an EC_{50} value near 1 μM and were nontoxic at the concentrations tested, as determined by viability staining with propidium iodide (PI) and an LDH release assay (Fig. 2.14-16). Extending toxicity tests to higher concentrations in renal cells showed that no toxicity was observed at the highest concentration needed for maximum PMO activity in HeLa 654 cells. We compared these results to a previously well-performing CPP for PMO delivery, Bpep-Bpep.[28] This peptide has similar

activity, but is composed of mostly Arg residues and exhibits cytotoxicity above 10 µM (Fig 2.17). This contrast between Mach peptides and Bpep-Bpep indicates that there is no apparent direct connection between toxicity and cargo delivery efficacy. PMO-Mach constructs have high activity, low arginine content, and a wide therapeutic window, highlighting their suitability for cytosolic and nuclear delivery.

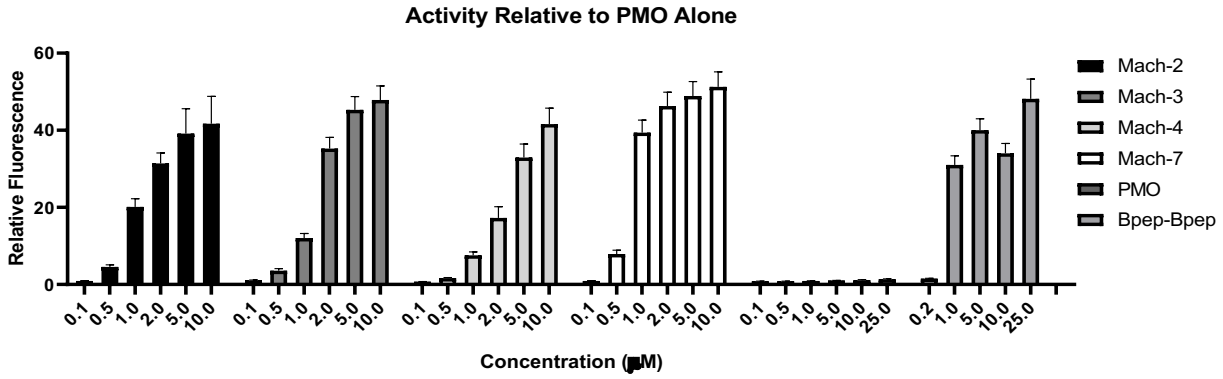


Figure 2.14 Dose-response in HeLa 654 cells (Activity). PMO-Mach constructs elicit a dose-dependent increase in EGFP fluorescence. Included here is chimera PMO-Bpep-Bpep, a previously reported high-performing PMO-peptide. Each bar represents group mean ± SD. For Mach2 and Bpep-Bpep n = 2; PMO, Mach3, Mach4, Mach7 n = 4.

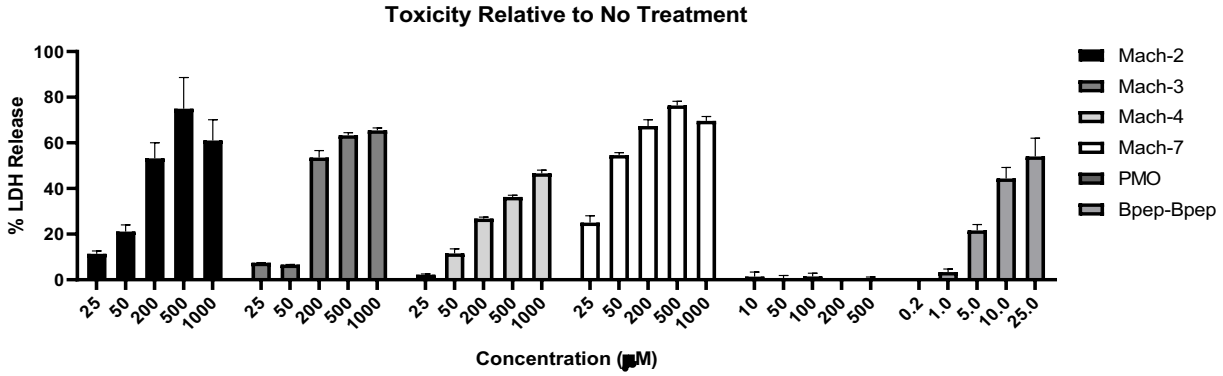


Figure 2.15 Dose-response in RPTEC (Toxicity). PMO-Mach constructs elicit a dose-dependent increase in membrane toxicity as measured by LDH release assay. LC50 of PMO-Mach constructs are between 100-200 µM, in contrast to PMO-Bpep-Bpep, which has a significantly lower LC50 near 10 µM. Each bar represents group mean ± SD. For Mach2 n = 3; Mach3 and Mach7 n = 5; Mach4, PMO, Bpep-Bpep n = 6.

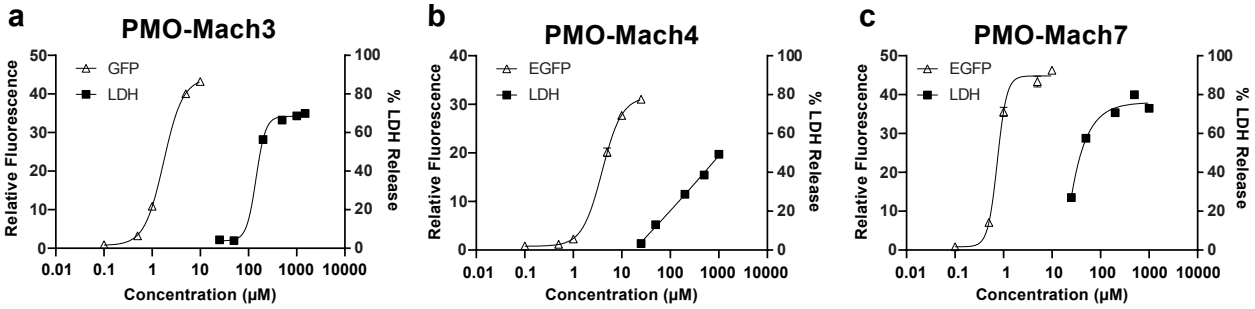


Figure 2.16 Dose response activity and toxicity for Mach3, 4, and 7 Certain PMO-Mach constructs show a gap in concentrations needed to elicit the highest activity and those needed to elicit toxicity. (a) PMO-Mach3 EC50 = 2 μM LD50 = 200 μM (b) PMO-Mach4 EC50 = 5 μM LD50 = 200 μM (c) PMO-Mach7 EC50 = 1 μM LD50 = 80 μM. Each point represents group mean ± SD.

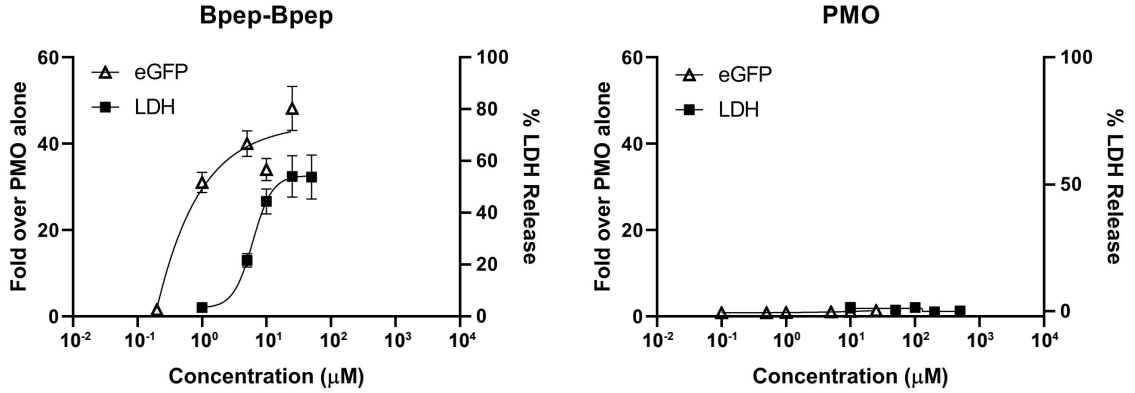
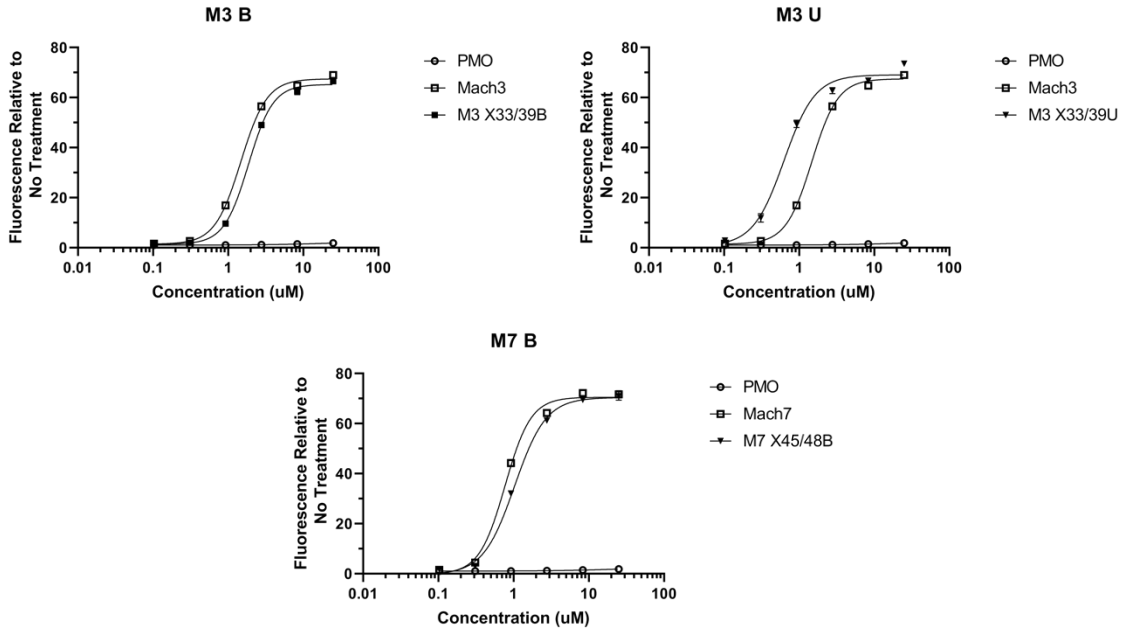


Figure 2.17 Dose response curves corresponding to activity and toxicity in HeLa 654. HeLa 654 cells were treated with varying concentrations of PMO-Bpep-Bpep or PMO alone for 22 h. Following treatment, cell supernatant was removed and tested for LDH release, reported as % LDH release relative to full lysis control (LDH, square). Toxicity is compared to activity (EGFP, triangle) in HeLa 654 cells. Bpep-Bpep EC50 = 1 μM, LC50 = 10 μM.

Knowing that the Mach sequences boosted activity, we used both the attribution analysis tool and empirical knowledge to further increase activity of Mach3. Attribution analysis provides opportunity for post-hoc experimentation with peptide sequences by highlighting certain residues that play a role in predicted activity. Given that Ahx is highly activated in Mach3, and reports that extended alkyl backbone chain amino acids have a large effect on CPP activity, we hypothesized that mutating these residues to residues with a longer chain may increase activity. We also observed that the C-terminus of Mach3 and Mach7 are highly activated regions along the sequences, and so hypothesized that the 10 C-terminal residues may retain some CPP activity. We made several point mutations and truncations to Mach3 and Mach7 to investigate our hypotheses. Mutating Ahx (6-carbon chain) to undecanoic acid (11-carbon chain) indeed enhanced the activity

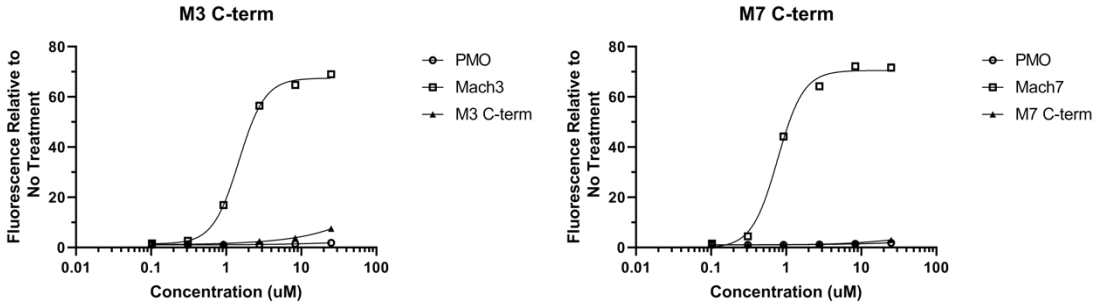
of Mach3, decreasing the EC50 to 0.6 μM from 1.5 μM. Mutation from Ahx to beta-Ala (3-carbon chain) decreased activity only slightly in both Mach3 and Mach7 (Fig 2.18). Finally, truncated versions of PMO-Mach constructs do not retain the activity of the parent sequences (Fig. 2.19). These observations suggest that the model is taking advantage of sequence-activity relationships that go beyond sequence length and charge.

At the same time, we evaluated the robustness of gradient-based attribution by analyzing residue-activations for these mutated Mach3 sequences (Fig. 2.20). We mutated each active Ahx residue with beta-alanine or aminoundecanoic acid. Consistent with our earlier findings, we observed that the most activated residue was Arg when both Ahx are mutated to β-Ala, a residue with a shorter alkyl backbone. The most activated residue reverts back to undecanoic acid, a residue with a longer alkyl backbone than Ahx, for the undecanoic acid mutation. This experiment validates the robustness of attribution analysis, both in terms of activated residues which conform to known biochemical principles and experimental validation of the mutations.



Mach3	QKKRKSKANKKNWPKGKLSIHAKDYKQGPKAKXRKQRXR
Mach3 X33,38B	QKKRKSKANKKNWPKGKLSIHAKDYKQGPKAKBRKQRBR
Mach3 X33,38U	QKKRKSKANKKNWPKGKLSIHAKDYKQGPKAKURKQRUR
Mach7	XKHPXAVQBAARAWKVPAAALWKKKRLKKSSKQKKKWLWKARSAXKYXRLI
Mach7 X45,48B	XKHPXAVQBAARAWKVPAAALWKKKRLKKSSKQKKKWLWKARSABKYBRLI

Figure 2.18 Mutations of Mach peptides can affect activity. Shown are dose-response curves in HeLa 654 after testing with PMO-Mach analogs, along with their sequences. B = beta-alanine, X = aminohexanoic acid, U = aminoundecanoic acid. Mutation to beta-alanine decreases activity slightly. Mutation to aminoundecanoic acid increases activity of Mach3 significantly. Activity is shown as fluorescence relative to untreated cells, with the curve corresponding to PMO alone also shown. Points and error bars represent mean and standard deviation, respectively. N = 3.



Mach3	QKKRKSKANKKNWPKGKLSIHAKDYKQGPKAKXRKQRXR
Mach7	XKHPXAVQBAARAWKVPAAALWKKKRLKKSSKQKKKWLWKARSAXKYXRLI
M3 C-term	KAKXRKQRXR
M7 C-term	RSAXKYXRLI

Figure 2.19 Truncated Mach peptides do not retain delivery activity Truncation of Mach peptides ablates PMO activity. Shown are dose-response curves in HeLa 654 after testing with PMO-Mach analogs, along with their sequences. B = beta-alanine, X = aminohexanoic acid. The 10 C-terminal residues of Mach3 and Mach7 do not retain the activity of the parent sequence. Activity is shown as fluorescence relative to untreated cells, with the curve corresponding to PMO alone also shown. Points and error bars represent mean and standard deviation, respectively. N = 3.

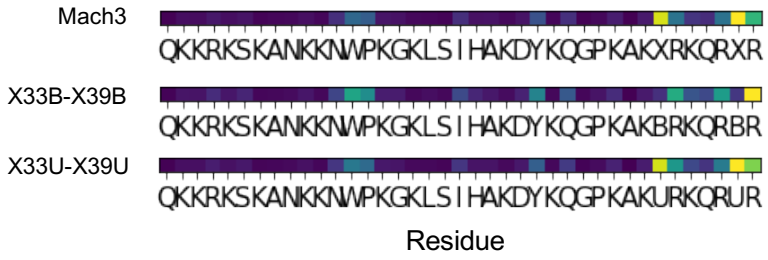


Figure 2.20 Attribution analysis of mutations. Alanine and β-Alanine mutations of the most active residue(s) shows a fall in the predicted MFI and a corresponding change in the positive activation heatmap. Ahx remains the most active residue for single Ala mutations, that changes to Arg when both C-

terminus Ahx are mutated to Ala and β-Ala. However, when Ahx is mutated to undecanoic acid (U), both U are the most positively activated residues.

2.2.6 Biophysical evaluation of Mach Miniproteins

Several PMO-Mach constructs have greater potency than previously characterized PMO-CPPs, while remaining nontoxic. This type of macromolecular delivery is a historic challenge, often suffering from either membrane toxicity or endosomal entrapment. Therefore we investigated the constructs' mechanism of uptake using a panel of endocytosis inhibitors and the HeLa 654 assay. Chemical endocytosis inhibitors were used to probe the mechanism of delivery of PMO in a pulse-chase format. HeLa 654 cells were preincubated with various chemical inhibitors or incubated at 4 °C for 30 minutes before treatment with PMO-Mach constructs for three hours. Treatment media was then replaced with fresh media for 22 hours. Cells were then lifted as previously described and EGFP synthesis was measured by flow cytometry (Fig. 2.21). Chlorpromazine (CPZ) has a dose-dependent negative impact on PMO delivery. CPZ is known to inhibit clathrin-mediated endocytosis, indicating that these PMO constructs may be utilizing this particular pathway, although multiple modes of uptake may be used.

We then investigated the Mach miniproteins' propensities for secondary structure. Peptides in PBS buffer at 20 μM, with or without 10 mM sodium dodecyl sulfate (SDS) were measured using circular dichroism (Fig. 2.22). While all peptides showed random coil character in aqueous solution, Mach1, 2, and 7 showed a shift towards alpha helical character in a lipid environment. Curiously, these sequences resulted in higher toxicity compared to Mach3 and Mach4, which did not show any alpha helical character in either condition. These three peptides are also longer (>50 residues) compared to Mach3 and Mach4 (<40 residues). Further tests are required to draw any conclusions from these trends, although peptides with strong alpha helical character have been shown to be more membrane lytic than sequences without rigid structure.[47]

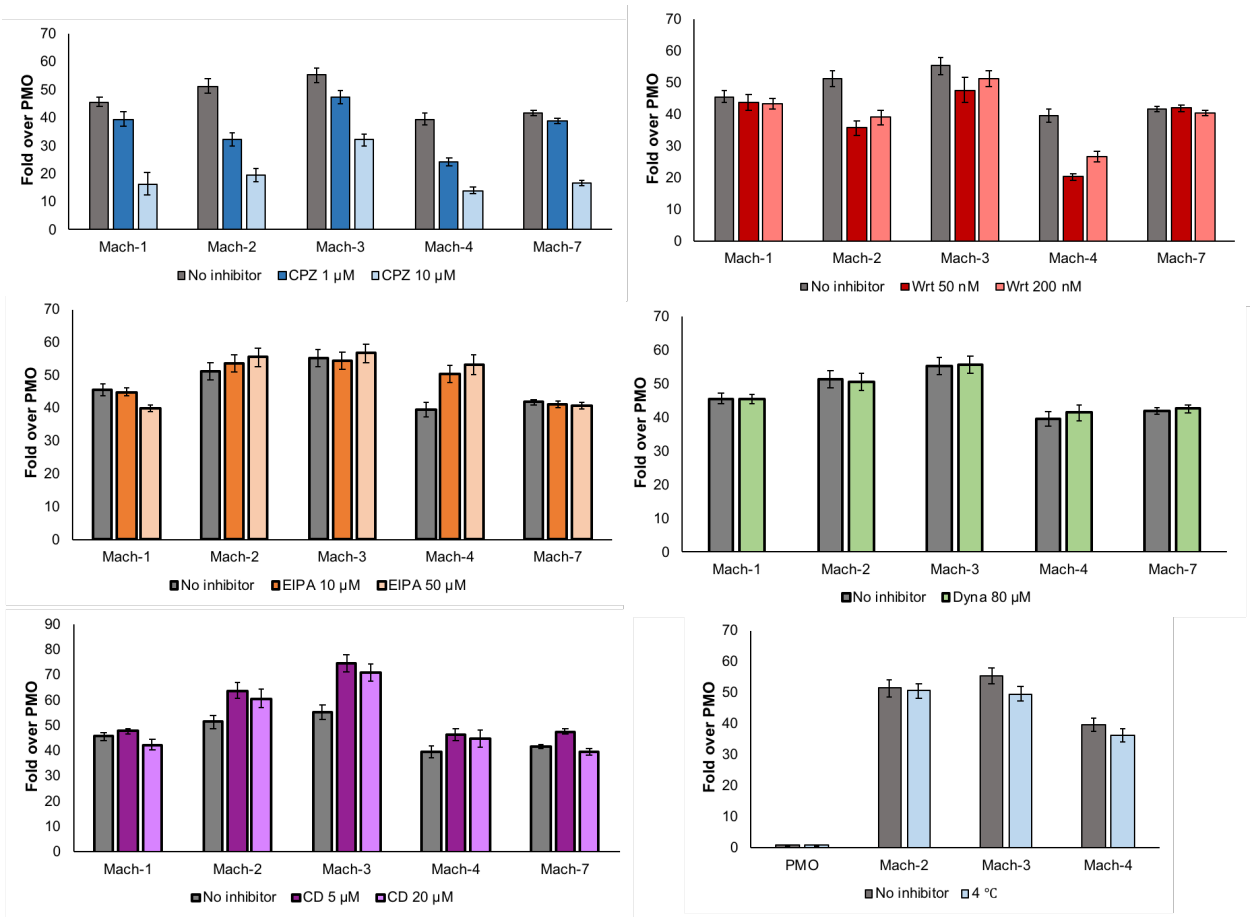


Figure 2.21 PMO-Mach peptides enter cells by energy-dependent endocytosis. PMO-Mach peptides enter cells by energy-dependent endocytosis. PMO activity of Mach constructs when treated with various endocytosis inhibitors. Chlorpromazine (CPZ) has a dose-dependent effect on PMO activity for each of the Mach constructs, indicating that constructs may enter via clathrin-mediated endocytosis. Each bar represents group mean ± SD, n = 3.

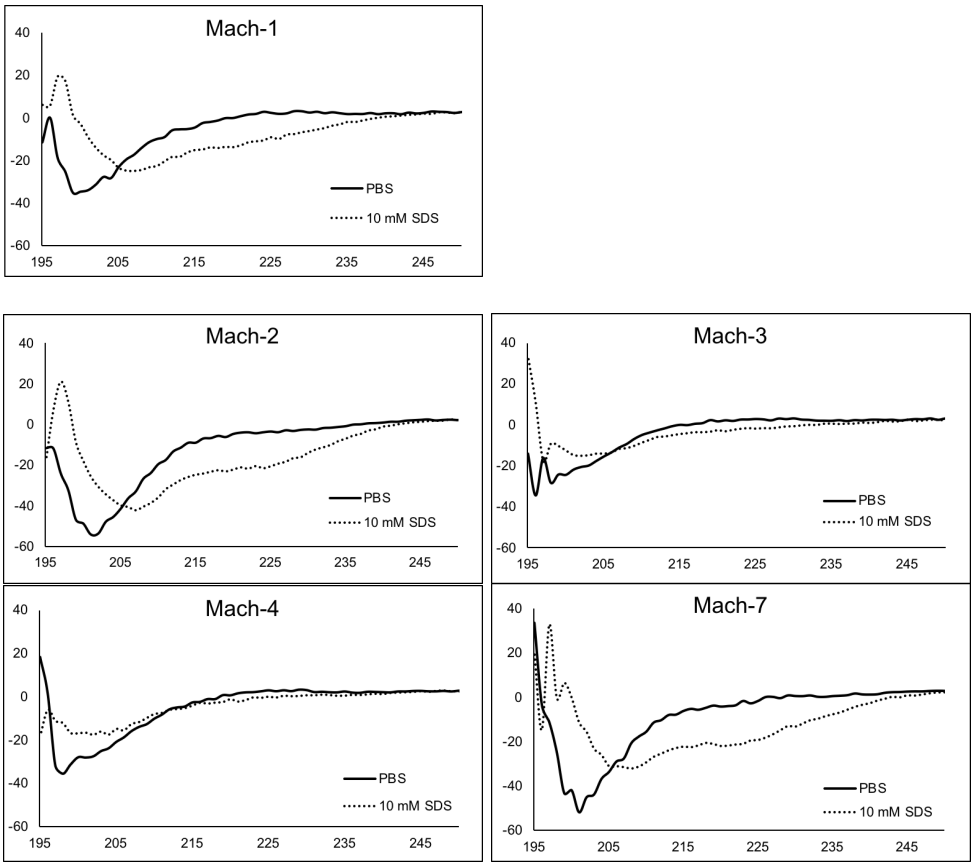


Figure 2.22 azide-Mach peptides show limited secondary structure by circular dichroism Circular dichroism of azide-Mach peptides. 20 μM Mach peptides were either incubated in PBS or 10 mM SDS before analysis using circular dichroism. In buffer, these peptides do not exhibit secondary structure. In a lipid environment, Mach1, Mach2, and Mach7 exhibit partial alpha helicity.

2.2.7 Mach miniproteins deliver other biomacromolecules

Mach miniproteins are versatile in that they can deliver other large biomolecules to the cytosol. Peptide nucleic acid (PNA), is a class of synthetic antisense oligonucleotides that has the same mechanism of action as PMO but has a highly flexible backbone structure.[48] We tested for delivery of a PNA variant of PMO 654 that is compatible with the EGFP assay. Each of the four Mach miniproteins tested was able to significantly enhance PNA delivery (2.23a).

In addition to antisense oligonucleotides, Mach peptides can also deliver charged proteins, such as Diphtheria toxin A (DTA). DTA is a 21 kDa anionic protein segment containing the catalytic domain of the toxin but lacking the portions that endow cell entry.[49] Delivery of this enzyme can be monitored using a cell proliferation assay as it inhibits protein synthesis in the cytosol. We found that Mach-DTA constructs were delivered into the cell cytosol significantly more efficiently than protein alone, and that covalent linkage was required for delivery (Fig.

2.23b). Furthermore, we confirmed that toxicity is due to the cytosolic delivery of active DTA by comparing the wild-type constructs to those containing DTA(E148S), a mutant with 300-fold lower activity that the wild-type.[50] As expected, the mutant DTA conjugates lead to significantly reduced toxicity.

Conjugation to Mach miniproteins also improves the delivery of EGFP, a fluorescent protein commonly used as a reporter. Confocal micrographs of HeLa cells displayed diffuse green fluorescence in the cytosol and intense fluorescence in the nucleus after incubation with Mach-EGFP (Fig. 2.23c). This observation is in contrast with the EGFP alone condition, in which no diffuse fluorescence was observed in either location, indicating reduced uptake.

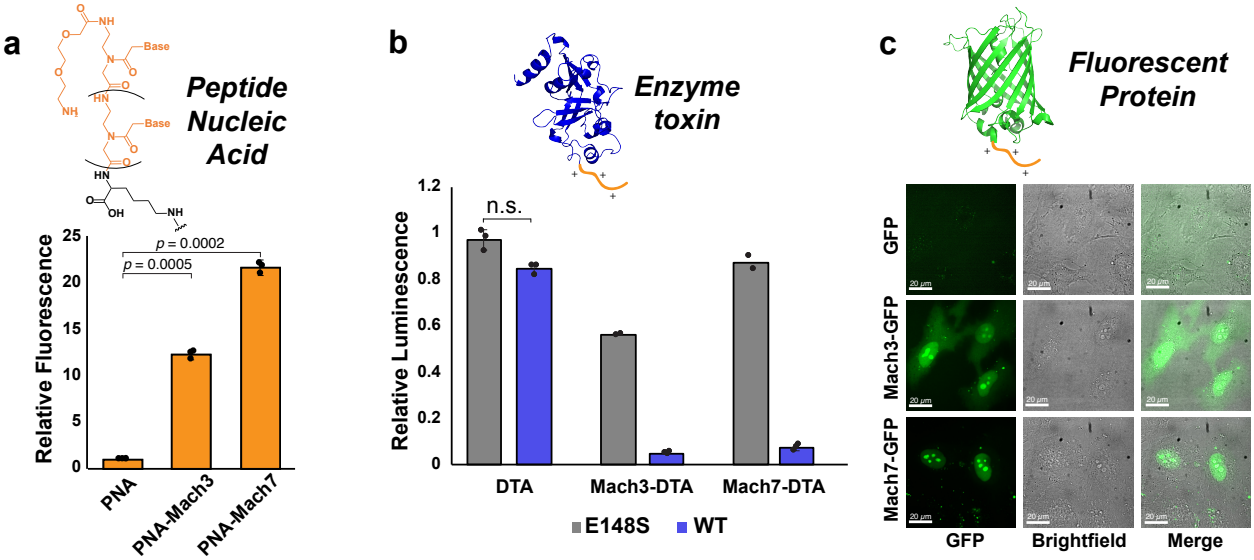
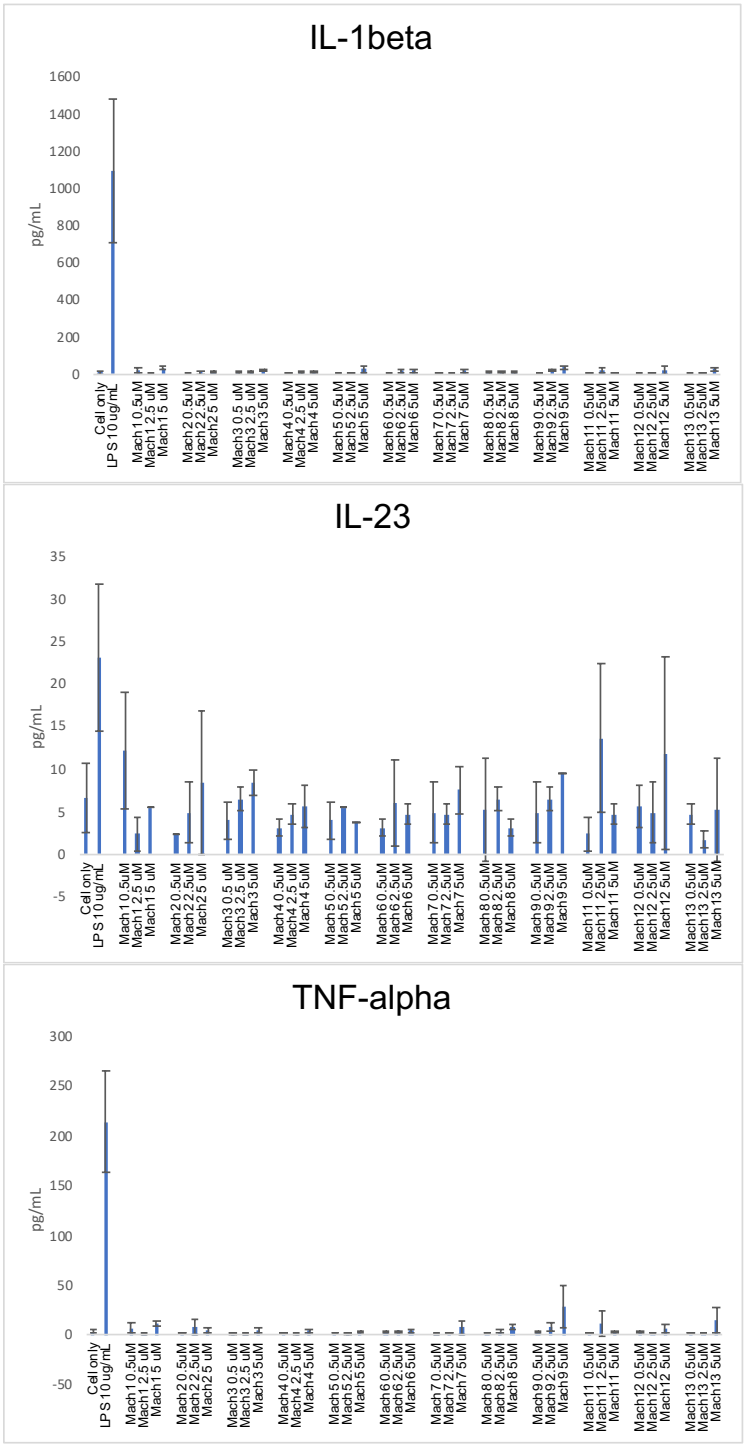


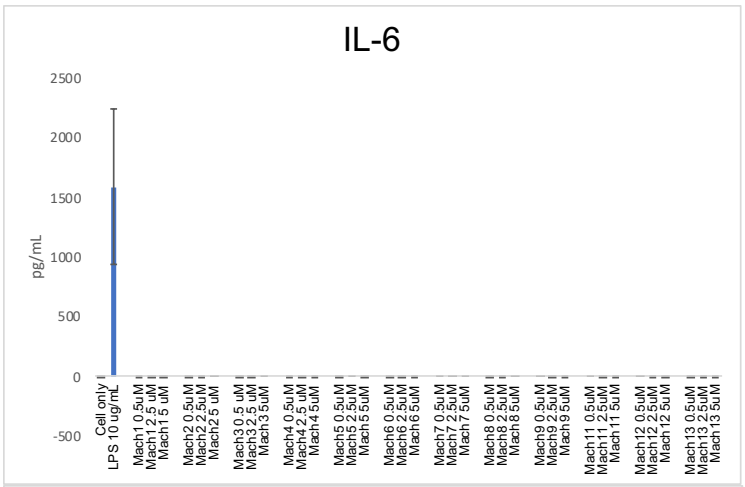
Figure 2.23 Mach peptides can deliver other biomacromolecules (a) The relative fluorescence of Mach conjugated to PNA 654 are compared to PNA alone, as determined by EGFP assay. (b) Comparison of the toxicity of wild-type and inactive mutant DTA and DTA(E148S) alone or conjugated to Mach3 or Mach7. Delivery of the active toxin to the cytosol results in toxicity as measured by luminescence. (c) Confocal micrographs displaying green fluorescence produced by EGFP, Mach3-EGFP, or Mach7-EGFP in HeLa cells after 3h incubation at 10 µM. (a) Each bar represents group mean ± SD, n = 3 distinct samples. P-values calculated from student's two-tailed t-test. (b) Each bar represents group mean ± SD, n=3, except for Mach3-DTA(E148S) and Mach7-DTA(E148S) which show mean, n=2 distinct samples. Full concentration curves are reported in Supplementary Figure 25. (c) Scale bar shows 10 µm. This experiment was conducted twice independently with similar results.

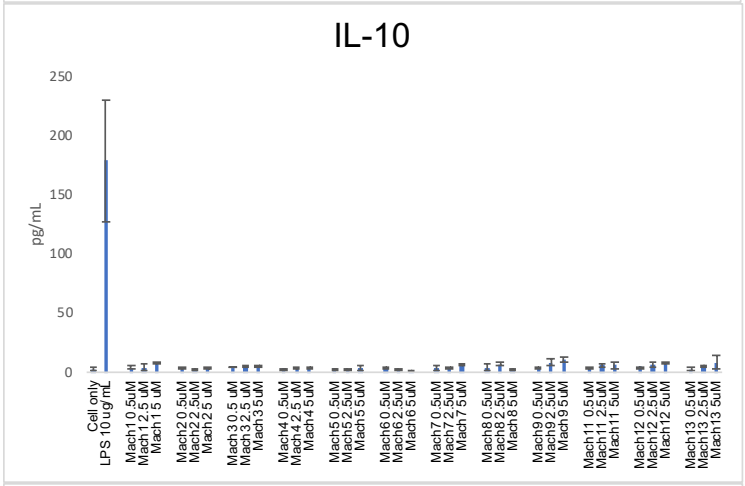
2.2.8 PMO-Mach restores protein synthesis in mice

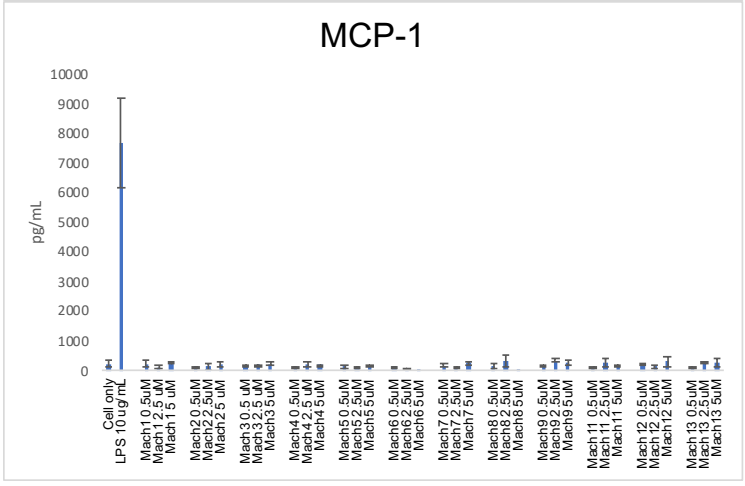
After verifying Mach miniproteins' propensity for in vitro macromolecule delivery, we looked towards in vivo antisense applications. In vitro tests with human macrophages suggested that the constructs are not inflammatory and therefore may be safe to evaluate in animals (Fig. 2.24).

Existing predictive models also suggest that Mach sequences would not be T cell epitopes (Fig. 2.25).




IL-6
pg/mL
2500
2000
1500
1000
500
0
-500
Cell only
LPS 10 ug/mL
Mach1 0.5uM
Mach1 2.5 uM
Mach1 5 uM
Mach2 0.5uM
Mach2 2.5uM
Mach2 5 uM
Mach3 0.5 uM
Mach3 2.5 uM
Mach3 5uM
Mach4 0.5uM
Mach4 2.5 uM
Mach4 5uM
Mach5 0.5uM
Mach5 2.5uM
Mach5 5uM
Mach6 0.5uM
Mach6 2.5uM
Mach6 5uM
Mach7 0.5uM
Mach7 2.5uM
Mach7 5uM
Mach8 0.5uM
Mach8 2.5uM
Mach8 5uM
Mach9 0.5uM
Mach9 2.5uM
Mach9 5uM
Mach11 0.5uM
Mach11 2.5uM
Mach11 5uM
Mach12 0.5uM
Mach12 2.5uM
Mach12 5uM
Mach13 0.5uM
Mach13 2.5uM
Mach13 5uM


IL-10
pg/mL
250
200
150
100
50
0
Cell only
LPS 10 ug/mL
Mach1 0.5uM
Mach1 2.5 uM
Mach1 5 uM
Mach2 0.5uM
Mach2 2.5uM
Mach2 5 uM
Mach3 0.5 uM
Mach3 2.5 uM
Mach3 5uM
Mach4 0.5uM
Mach4 2.5 uM
Mach4 5uM
Mach5 0.5uM
Mach5 2.5uM
Mach5 5uM
Mach6 0.5uM
Mach6 2.5uM
Mach6 5uM
Mach7 0.5uM
Mach7 2.5uM
Mach7 5uM
Mach8 0.5uM
Mach8 2.5uM
Mach8 5uM
Mach9 0.5uM
Mach9 2.5uM
Mach9 5uM
Mach11 0.5uM
Mach11 2.5uM
Mach11 5uM
Mach12 0.5uM
Mach12 2.5uM
Mach12 5uM
Mach13 0.5uM
Mach13 2.5uM
Mach13 5uM


MCP-1
pg/mL
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Cell only
LPS 10 ug/mL
Mach1 0.5uM
Mach1 2.5 uM
Mach1 5 uM
Mach2 0.5uM
Mach2 2.5uM
Mach2 5 uM
Mach3 0.5 uM
Mach3 2.5 uM
Mach3 5uM
Mach4 0.5uM
Mach4 2.5 uM
Mach4 5uM
Mach5 0.5uM
Mach5 2.5uM
Mach5 5uM
Mach6 0.5uM
Mach6 2.5uM
Mach6 5uM
Mach7 0.5uM
Mach7 2.5uM
Mach7 5uM
Mach8 0.5uM
Mach8 2.5uM
Mach8 5uM
Mach9 0.5uM
Mach9 2.5uM
Mach9 5uM
Mach11 0.5uM
Mach11 2.5uM
Mach11 5uM
Mach12 0.5uM
Mach12 2.5uM
Mach12 5uM
Mach13 0.5uM
Mach13 2.5uM
Mach13 5uM

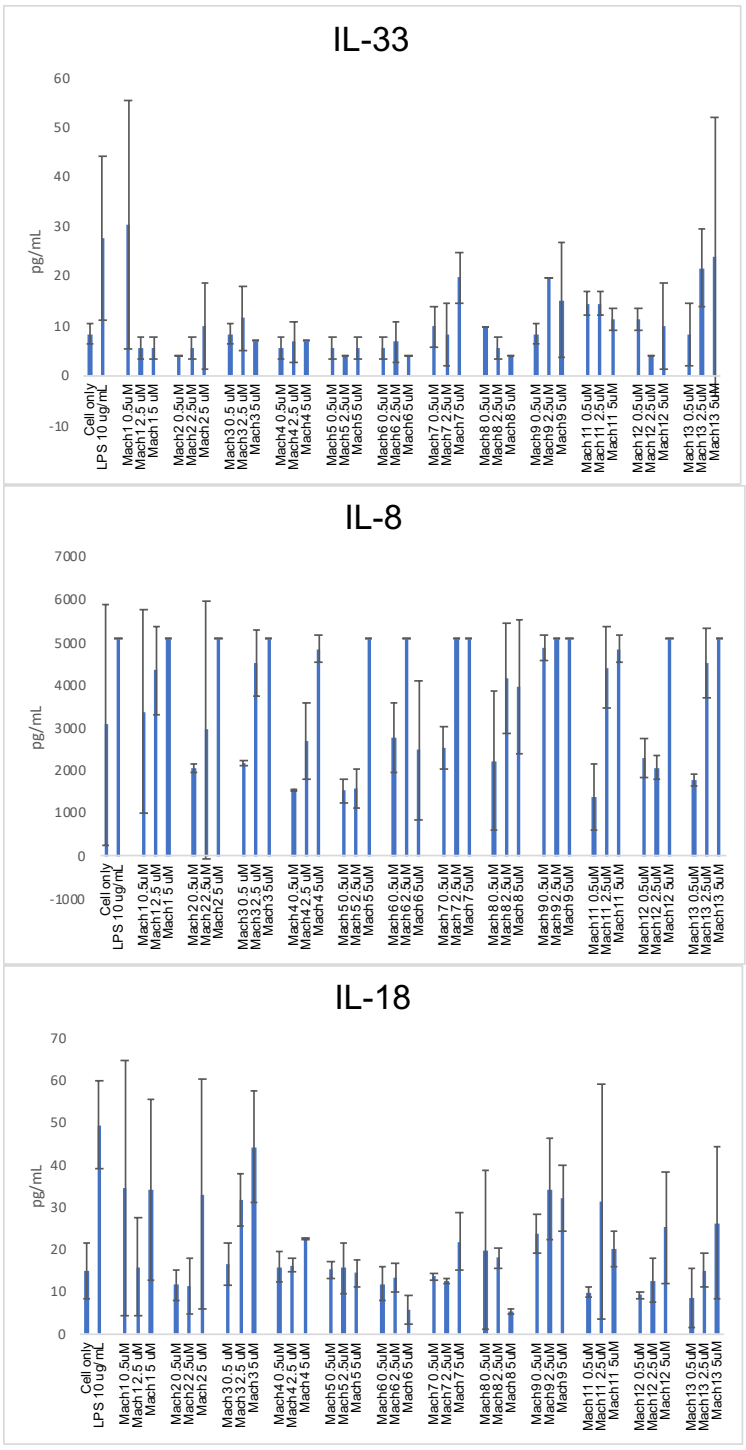


Figure 2.24 PMO-Mach constructs are nonimmunogenic in vitro. PMO-Mach constructs are nonimmunogenic in vitro. Inflammation panel results of cytokines that were detected in human monocyte-derived macrophages. IL-1B, TNF-a, IL-6, IL-10, and MCP-1 are all released after treatment with lipopolysaccharide (LPS), but exhibit no significant increase after treatment with PMO-Mach constructs. Each bar represents group mean ± SD, n = 2.

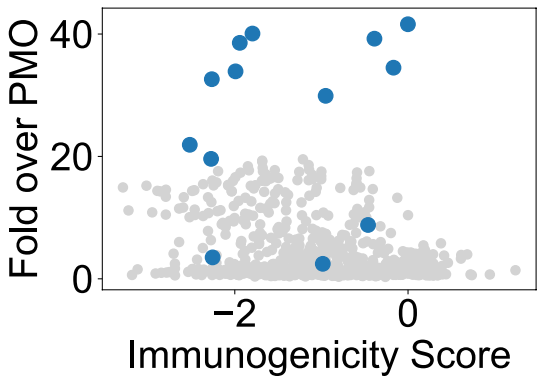


Figure 2.25 In silico immunogenicity score for Mach and training sequences. In silico immunogenicity score for Mach and training sequences. Predicted immunogenicity for Mach sequences is within the range of the predicted immunogenicity for the sequences used in training of predictor. Mach sequences have a substantially higher experimental normalized MFI within the same range of immunogenicity, in comparison to the sequences used in the training of the predictor. The immunogenicity scores are the likelihood of being a T-cell epitope. The values are calculated using an online predictor.[51]

Lastly, we demonstrated that PMO-Mach constructs safely correct protein synthesis in animals. Transgenic mice containing the same EGFP IVS2-654 gene as used in cell assays were given a single intravenous injection of varying doses of PMO-Mach3 or PMO-Mach4 and evaluated after 7 days. Both constructs exhibited a dose-dependent increase in EGFP expression in quadriceps, diaphragm, and heart (Fig. 2.26a-c). PMO delivery to the heart is a critical but challenging objective. Here we observe similar levels of protein synthesis in both skeletal and cardiac tissue. In addition, there were no significant changes in the level of renal function biomarkers 7-days post-treatment (Fig. 2.26d-f). These findings indicate that Mach miniproteins may be safe delivery materials for PMO to muscle tissue.

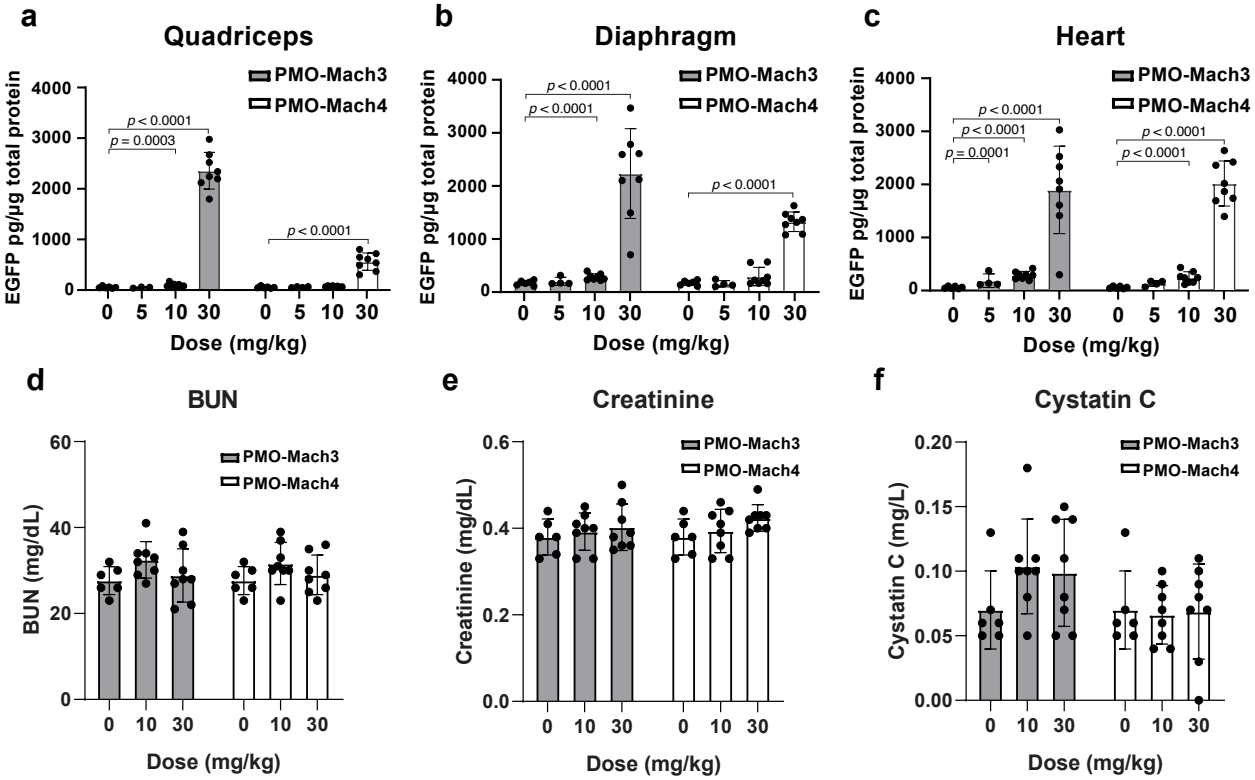


Figure 2.26 PMO-Mach constructs restore protein synthesis in mice and do not induce kidney toxicity. PMO-Mach constructs do not induce kidney toxicity in mice. In EGFP 654 mice, levels of (a) blood urea nitrogen (BUN), (b) creatinine, and (c) cystatin C remained unchanged. Saline (n = 6), Mach3 and Mach4 at 5 mg/kg (n = 4), all other n = 8. Each bar represents group mean ± SD. Mann-Whitney U test showed no significant difference between groups.

2.3 Discussion

We demonstrate a method to efficiently sample the vast chemical search space of functional peptides using machine learning and standardized experimentation. Our model was applied to the design of abiotic miniproteins that can deliver an antisense PMO to the nucleus with greater efficiency than any previously known polypeptide-based variant. Importantly, the new constructs are effective in animals and are non-toxic up to a dose of 30 mg/kg. These miniproteins are versatile intracellular carriers, delivering other classes of biomolecules to the nucleus and cytosol, including antisense PNA, fluorescent protein, and enzymes. The core strengths of our model lie in: (1) standardized quantitative activity data, (2) the model's ability to extrapolate beyond the training set and (3) a visual attribution tool to interpret the decision-making process of the model.

A critical factor in building a robust machine learning model was the training dataset; the 600-member library was synthesized by combining peptide modules, and tested in a standardized assay that provides quantitative activity information. Synthesis and testing of the modular PMO-CPP

library produced a broad spectrum of sequence and activity data with which we trained the model. By representing peptide sequences as topological fingerprints rather than categorical choices or descriptors such as molecular weight, charge, and hydrophobicity, the model has access to inherent structural information and can be used on monomers not encountered in the training data. The standardized activity values allowed us to use a quantitative regressor, rather than an active/inactive classifier, and thus design sequences with a broad spectrum of activity predictions. While we have previously tested CPPs designed by other machine learning methods, we found that they were not able to deliver PMO.[7] The CNN model using fingerprints was able to extrapolate predicted activity beyond that of the training dataset, while models using other frameworks and representations were not. While the other models and methods to generate seed sequences may be able to produce sequences with high experimental activity, the ability to predict that high activity is critical to differentiate predicted sequences. Since our goal is to discover unique peptides with higher activity than any previously known, a model able to predict values outside the range of the training data is required, thus necessitating the use of CNN with fingerprint representations.

The interpretability of the model is an additional advantage. By overlaying the output of the predictor with the sequence matrix of a given peptide, we can visualize the activated residues and substructures important for the decision-making process. Several observations from the interpretations match our current understanding of CPP motifs, such as the benefit of cationic residues. The model also identified Ahx as an important residue, one which has only been investigated in the context of endosomal escape in Arg-rich sequences.[52] This tool allowed for post-hoc analysis to validate empirical hypotheses and enhance the activity of Mach3 by mutating aminohexanoic acid to aminoundecanoic acid, an amino acid not present in the training dataset.

In addition to PMO, Mach peptides deliver other antisense oligonucleotides as well as functional proteins into the cell cytosol. Delivery of EGFP reveals diffuse green fluorescence in the cytosol and clear accumulation of EGFP to the nucleus. We believe that Mach peptides may contain nuclear localization sequences (NLS), which have been described previously and are typically lysine-rich.[53] However, nuclear localization is not solely due to the cationic charge, as shown by a previous study.[54] The model likely selected for such NLS sequences because the activity used in the training was acquired from a nuclear delivery-based assay.

PMO-Mach conjugates effected a dose-dependent increase in protein synthesis in all three examined mouse muscle tissues including heart after a single intravenous injection. The mouse

model used contains the same transgene as the in vitro assay, and the Mach sequences recapitulated the in vitro results in vivo, indicating that the model implemented here could be applied to data obtained from animal experiments. If a sequence-activity training set were generated from data obtained in animals, then this model may be applicable further downstream in the drug design pipeline. A greater challenge remains toward in vivo delivery to target tissues. In Duchenne muscular dystrophy, PMO must access the nucleus of muscle cells to have therapeutic effect. Targeting to cardiac tissue is a primary concern given that the leading cause of death from this disease is heart failure. Our animal model confirmed localization of the PMO-Mach constructs to the heart, suggesting a potential solution to the tissue targeting challenge.

In conclusion, this strategy illustrates how deep learning can be applied to de novo design of functional abiotic miniproteins. The Mach miniproteins are the most effective PMO delivery constructs developed to date and are effective in animals. Our machine learning framework could potentially be repurposed to discover sequence-optimized peptides with other desired activities, solely requiring a standardized high-quality input dataset. We envision that this strategy will enable the rapid future design of de novo functional peptides with impact on chemical, biological and material sciences.

2.4 Materials & Methods

2.4.1 Reagents & Solvents

H-Rink Amide-ChemMatrix resin was obtained from PCAS BioMatrix Inc. (St-Jean-sur-Richelieu, Quebec, Canada). 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium-3-oxid-hexafluorophosphate (HATU), 4-pentynoic acid, 5-azidopentanoic acid, Fmoc-β-Ala-OH, Fmoc-6-aminohexanoic acid, and Fmoc-L-Lys(N_3) were purchased from Chem-Impex International (Wood Dale, IL). PyAOP was purchased from P3 BioSystems (Louisville, KY). Fmoc-protected amino acids (Fmoc-Ala-OHxH_2O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(O*t*Bu)-OH; Fmoc-Cys(Trt)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(O*t*Bu)-OH; Fmoc-Gly-OH; Fmoc-His(Trt)-OH; Fmoc-Ile-OH; Fmoc-Leu-OH; Fmoc-Lys(Boc)-OH; Fmoc-Met-OH; Fmoc-Phe-OH; Fmoc-Pro-OH; Fmoc-Ser(But)-OH; Fmoc-Thr(*t*Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(*t*Bu)-OH; Fmoc-Val-OH), were purchased from the Novabiochem-line from Sigma Millipore. Peptide synthesis-grade *N,N*-dimethylformamide (DMF), CH_2Cl_2, diethyl ether, and HPLC-grade acetonitrile were obtained from VWR International (Radnor, PA). All other

reagents were purchased from Sigma-Aldrich (St. Louis, MO). Milli-Q water was used exclusively.

2.4.2 Liquid-chromatography mass-spectrometry

LCMS analyses were performed on either an Agilent 6520 Accurate-Mass Q-TOF LCMS (abbreviated as 6520) or an Agilent 6550 iFunnel Q-TOF LCMS system (abbreviated as 6550) coupled to an Agilent 1260 Infinity HPLC system. Mobile phases were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LCMS methods were used for characterization:

Method A: 1-61% B over 9 min, Zorbax C3 column (6520)

LC: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, 12-15 min 95% B; flow rate: 0.8 mL/min.
MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 300–3000 *m*/*z*. MS is on from 4 to 11 min.

Method B: 1-61% B over 10 min, Phenomenex Jupiter C4 column (6550)

LC: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min 1-61% B, 12-16 min 61-90% B; 16-20 min 90% B; flow rate: 0.1 mL/min.
MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 100–1700 *m*/*z*. MS is on from 4 to 12 min.

Method C: 1-61% B over 10 min, Agilent EclipsePlus C18 column (6550)

LC: Agilent EclipsePlus C18 RRHD column: 2.1 × 50 mm, 1.8 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min, 1-61% B, 12-13 min, 61% B, 13-16 min, 1% B; flow rate: 0.1 mL/min.
MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 300–3000 m/z. MS is on from 4 to 12 min. This method was used exclusively for characterization of the modular library.

All data were processed using Agilent MassHunter software package. Y-axis in all chromatograms shown represents total ion current (TIC) unless noted.

2.4.3 General method for peptide preparation

All peptides and miniproteins were synthesized by automated solid-phase peptide synthesis as previously described.[45,55] The preparation of the 600-member library was previously described in the thesis of Dr. Justin Wolfe.[30]

Fast-flow Peptide Synthesis: Peptides were synthesized on a 0.1 mmol scale using an automated fast-flow peptide synthesizer. A 100 mg portion of ChemMatrix Rink Amide HYR resin was loaded into a reactor maintained at 90 ºC. All reagents were flowed at 40 mL/min with HPLC pumps through a stainless-steel loop maintained at 90 ºC before introduction into the reactor. For each coupling, 10 mL of a solution containing 0.4 M amino acid and 0.38 M HATU in DMF were mixed with 600 μL diisopropylethylamine and delivered to the reactor. Fmoc removal was accomplished using 10.4 mL of 20% (v/v) piperidine. Between each step, DMF (15 mL) was used to wash out the reactor. For peptides in the modular library, special coupling conditions were used for arginine, in which 10 mL of a solution containing 0.4 M Fmoc-L-Arg(Pbf)-OH and 0.38 M PyAOP in DMF were mixed with 600 μL diisopropylethylamine and delivered to the reactor. For Mach peptides, additional special coupling conditions were used according to the optimized peptide synthesis protocol.[56] To couple unnatural amino acids or to cap the peptide (e.g. with 4-pentynoic acid), the resin was incubated for 30 min at room temperature with amino acid (1 mmol) dissolved in 2.5 mL 0.4 M HATU in DMF with 500 μL diisopropylethylamine. After completion of the synthesis, the resin was washed 3 times with dichloromethane and dried under vacuum.

Peptide Cleavage and Deprotection: Each peptide was subjected to simultaneous global side-chain deprotection and cleavage from resin by treatment with 5 mL of 94% trifluoroacetic acid (TFA), 2.5% 1,2-ethanedithiol (EDT), 2.5% water, and 1% triisopropylsilane (TIPS) (v/v) for 7 min at 60 °C or at room temperature for 2 to 4 hours. For arginine-rich sequences, the resin was treated with a cleavage cocktail consisting of 82.5% TFA, 5% phenol, 5% thioanisole, 5% water, and 2.5% EDT (v/v) for 14 hours at room temperature. For peptides containing azide, EDT was substituted for thioanisole. The cleavage cocktail was first concentrated by bubbling N_2 through the mixture, and cleaved peptide was precipitated and triturated with 40 mL of cold ether (chilled in dry ice). The crude product was pelleted by centrifugation for three minutes at 4,000 rpm and the ether was decanted. This wash step was repeated two more times. After the third wash, the pellet was dissolved in 50% water and 50% acetonitrile containing 0.1% TFA, filtered through a fritted syringe to remove the resin and lyophilized.

Peptide Purification: The peptides were dissolved in water and acetonitrile containing 0.1% TFA, filtered through a 0.22 μm nylon filter and purified by mass-directed semi-preparative reversed-phase HPLC. Solvent A was water with 0.1% TFA additive and Solvent B was acetonitrile with 0.1% TFA additive. A linear gradient that changed at a rate of 0.5% B/min was used. Most of the peptides were purified on an Agilent Zorbax SB C3 column: 9.4 x 250 mm, 5 μm. Extremely hydrophilic peptides, such as the arginine-rich sequences were purified on an Agilent Zorbax SB C18 column: 9.4 x 250 mm, 5 μm. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

Macrocyclization: Mach12 and Mach13 contained cysteine linked macrocycles. Purified unprotected peptide (1 mM) was dissolved in DMF with decafluorobiphenyl (2 mM) and DIEA (50 mM) and incubated at room temperature for 3 h. The solution was then diluted 100-fold in 1% acetonitrile, 2% TFA in water and purified directly by reverse-phase HPLC.

2.4.4 PMO and PNA conjugation to peptides

PMO IVS-654 (50 mg, 8 μmol), acquired from Sarepta Therapeutics, was dissolved in 150 μL DMSO. To the solution was added a solution containing 2 equivalents of dibenzocyclooctyne acid (5.3 mg, 16 μmol) activated with HBTU (37.5μL of 0.4 M HBTU in DMF, 15 μmol) and DIEA (2.8 μL, 16 μmol) in 40 μL DMF (Final reaction volume = 0.23 mL). The reaction proceeded for 25 min before being quenched with 1 mL of water and 2 mL of ammonium hydroxide. The ammonium hydroxide hydrolyzed any ester formed during the course of the reaction. After 1 hour, the solution was diluted to 40 mL in water/acetonitrile and purified using reverse-phase HPLC (Agilent Zorbax SB C3 column: 21.2 x 100 mm, 5 μm) and a linear gradient from 2 to 60% B (solvent A: water; solvent B: acetonitrile) over 58 min (1% B / min). Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

Mach miniproteins were conjugated to PMO via strain-promoted azide-alkyne cycloaddition. PMO-DBCO (5 mM in water) was stoichiometrically combined with azide-peptide (5 mM in water) and incubated at room temperature until reaction completed (between 2 and 12 hours), monitored by LCMS. The reaction was purified using reversed-phase HPLC (Agilent Zorbax SB C3 column: 21.2 x 100 mm, 5 μm) and a linear gradient from 2 to 60% B (solvent A: 100 mM

ammonium acetate in water pH 7.2; solvent B: acetonitrile) over 58 min (1% B / min). pure fractions were pooled as determined by LCMS and lyophilized.

50 nmol of PNA 654 (O-GCTATTACCTTAACCCAG-Lys(DBCO)) was purchased from PNABio. PNA-DBCO (1 mM in water) was stoichiometrically combined with azide-peptide (1 mM in water) and incubated at 4 °C for 12 hours. The product was then used in cell assays without purification. Conversion was checked by LCMS.

2.4.5 Endocytosis inhibitors & Circular dichroism

Peptides were dissolved in PBS buffer to obtain stock solutions of 1 mM. The circular dichroism (CD) spectra was obtained from 195 to 250 nm using an AVIV 420 circular dichroism spectrometer with a 1 mm path length quartz cuvette.

2.4.6 Cell assays

Activity Assays: HeLa 654 cells obtained from the University of North Carolina Tissue Culture Core facility were maintained in MEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin-streptomycin at 37 °C and 5% CO_2. 18 h prior to treatment, the cells were plated at a density of 5,000 cells per well in a 96-well plate in MEM supplemented with 10% FBS and 1% penicillin-streptomycin.

For testing of the library, on the day of the experiment, the 384 well plate containing the crude reaction mixtures in DMSO was diluted to 100 μM by the addition of 16.8 μL of PBS to the 3.2 μL reaction mixture. Then, each construct was diluted to 5 μM in MEM supplemented with 10% FBS and 1% penicillin-streptomycin. For individual peptide testing, PMO-peptides were dissolved in PBS without Ca^{2+} or Mg^{2+} at a concentration of 1 mM (determined by UV) before being diluted in MEM. Cells were incubated at the designated concentrations for 22 h at 37 °C and 5% CO_2. Next, the treatment media was removed, and the cells were washed once before being incubated with 0.25% Trypsin-EDTA for 15 min at 37 °C and 5% CO_2. Lifted cells were transferred to a V-bottom 96-well plate and washed once with PBS, before being resuspended in PBS containing 2% FBS and 2 μg/mL propidium iodide (PI). Flow cytometry analysis was carried out on a BD LSRII flow cytometer at the Koch Institute. Gates were applied to the data to ensure that cells that were positive for propidium iodide or had forward/side scatter readings that were sufficiently different

from the main cell population were excluded. Each sample was capped at 5,000 gated events (SM).

Analysis was conducted using Graphpad Prism 7 and FlowJo. For each sample, the mean fluorescence intensity (MFI) and the number of gated cells was measured. To report activity, triplicate MFI values were averaged and normalized to the PMO alone condition.

Toxicity assays: Cytotoxicity assays were performed in both HeLa 654 cells and human RPTEC (Human Renal Proximal Tubule Epithelial cells, TH-1, ECH001, Kerafast). RPTEC were maintained in high glucose DMEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin-streptomycin at 37 °C and 5% CO_2. Treatment of RPTEC was performed as with the HeLa 654 cells. After treatment, supernatant was transferred to a new 96-well plate. To each well of the 96-well plate containing supernatant, described above, was added CytoTox 96 Reagent (Promega). The plate was shielded from light and incubated at room temperature for 30 minutes. Equal volume of Stop Solution was added to each well, mixed, and the absorbance of each well was measured at 490 nm. The blank measurement was subtracted from each measurement, and % LDH release was calculated as % cytotoxicity = 100 × Experimental LDH Release (OD490) / Maximum LDH Release (OD490).

2.4.7 Preparation of protein conjugates

His_6-SUMO-G_5-DTA(C186S), His_6-SUMO-G_5-DTA(C186S, E148S) and G_5-EGFP-His_6 were overexpressed in E. coli BL21 (DE3) cells. Approximately 10 g of cell pellet was lysed by sonication in 50 mL of 20 mM Tris, 150 mM NaCl, pH 7.5 buffer containing 30 mg lysozyme, 2 mg DNAase I, and 1 tablet of cOmplete™ Protease Inhibitor Cocktail. The suspension was centrifuged at 16,000 rpm for 30 min to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap FF Ni-NTA column (GE Healthcare, UK) and washed with 30 mL of 100 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. Protein was eluted from the column with buffer containing 300 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. Imidazole was removed from protein via centrifugation in Millipore centrifugal filter unit (10K).

For the DTA constructs, the His_6-SUMO tag was then cleaved from the protein with SUMO protease (previously recombinantly expressed) by incubating a 1:1000 protease:protein ratio in 20 mM Tris, 150 mM NaCl, pH 7.5 overnight at 4 °C. Desired protein was separated from His_6-SUMO tag by flowing the mixture through a 5 mL HisTrap FF Ni-NTA column. Finally, purified

protein was isolated by size exclusion chromatography using HiLoad 26/600 Superdex 200 prep grade size exclusion chromatography column (GE Healthcare, UK) in 20 mM Tris, 150 mM NaCl, pH 7.5 buffer.

For the EGFP construct, purified protein was isolated by anion exchange chromatography using HiTrap Q HP anion exchange chromatography column (GE Healthcare, UK) in (0-40 %B over 20 CV) where A: 20 mM Tris, pH 8.5 buffer and B: 1 M NaCl, 20 mM Tris, pH 8.5 buffer.

Proteins were analyzed using an SDS-Page gel. In addition, proteins were analyzed by ESI-QTOF LCMS to confirm molecular weight and purity. The protein charge-state envelope was deconvoluted using Agilent MassHunter Bioconfirm using maximum entropy.

Synthesis and testing of Mach-DTA: Mach-LPSTGG peptides were synthesized and purified by standard protocol as described. G_5-DTA (50 μM) was incubated with either Mach3-LPSTGG (250 μM) or Mach7-LPSTGG (750 μM) and SrtA* (2.5 μM) for 90 minutes at 4 °C in SrtA buffer (10 mM $CaCl_2$, 50 mM Tris, 150 mM NaCl, pH 7.5). The reaction was monitored by LCMS and gel electrophoresis. After 90 minutes, Mach-DTA conjugate was isolated using HiLoad 26/600 Superdex 200 prep grade size exclusion chromatography column (GE Healthcare, UK) in 20 mM Tris, 150 mM NaCl, pH 7.5 buffer. Fractions containing the pure product as determined by LCMS and gel electrophoresis were concentrated using a centrifugal filter unit (10K, Millipore).

To test for DTA delivery to the cytosol, HeLa cells were plated at 5,000 cells/well in a 96-well plate the day before the experiment. Wild-type and mutant constructs of G_5-DTA, Mach3-DTA, and Mach7-DTA, as well as Mach3-LPSTGG and Mach7-LPSTGG were prepared at varying concentrations in complete media and transferred to the plate. Cell proliferation was measured after 48 h using the CellTiter-Glo assay.

Synthesis and testing of Mach-EGFP: G_5-EGFP (60 μM) was incubated with either Mach3-LPSTGG (1000 μM) or Mach7-LPSTGG (1000 μM) and SrtA* (5 μM) in SrtA buffer (10 mM $CaCl_2$, 50 mM Tris, 150 mM NaCl, pH 7.5) for 90 minutes at room temperature under exclusion of light. The reaction was monitored by LCMS and gel electrophoresis. After 90 minutes, Mach-eGFP conjugate was isolated using cation exchange chromatography using HiTrap SP HP cation exchange chromatography column (GE Healthcare, UK) in (0-100 %B over 20 CV) where A: 50 mM NaCl, 20 mM Tris, pH 7.5 buffer and B: 1 M NaCl, 20 mM Tris, pH 12 buffer. Fractions containing the pure product as determined by LCMS and gel electrophoresis were immediately desalted and concentrated using a centrifugal filter unit (10K, Millipore).

To visualize delivery of EGFP into cells, HeLa were plated at 5,000 cells/well in a coverslip glass-bottomed 96-well plate the day before the experiment. Mach3-EGFP, Mach7-EGFP, or EGFP at varying concentrations were added to each well and incubated at 37 °C and 5% CO_2 for 1 h. Treatment media was replaced with fresh media 1 h before being imaged in the W.M. Keck microscopy facility on an RPI Spinning Disk Confocal microscope on brightfield and GFP setting (488 nm, 150 mW OPSL excitation laser, 525/50 nm emission).

2.4.8 In vitro inflammation panel

THP-1 cells (ATCC TIB-202) were grown in RPMI 1640 media supplemented with 10% (v/v) FBS, 1% (v/v) penicillin-streptomycin, L-glutamine, non-essential amino acids, sodium pyruvate at 37 °C and 5% CO_2. THP-1 cells (450k/mL) were treated with 25 nM phorbol 12-myristate 13-acetate (PMA) at 37 °C and 5% CO_2 for 24 h to trigger differentiation into macrophages. Then, media was replaced with fresh RPMI media and the cells were incubated for another 24 h. Cells were then collected, spun down, and brought up in complete RPMI media to a cell density of 500k/mL. 100k cells were plated in each well of a 96-well plate, leaving the first two columns empty. Duplicate wells were treated with varying concentrations of the PMO-peptide conjugates at 37 °C and 5% CO_2 for 2 h. Media-only and no treatment wells were used as negative controls, and 10 μg/mL bacterial lipopolysaccharide (LPS) treatment was used as a positive control. Following treatment, each well was washed three times, given fresh media, and incubated for 12 h. Supernatant was transferred to a V-bottom plate. Inflammatory cytokines in the supernatant were assayed using LEGENDplex Human Inflammation panel (BioLegend). Analysis was carried out on a BD LSRII flow cytometer and data was analyzed using BioLegend's accompanying software.

2.4.9 In vivo studies

EGFP-654 transgenic mice obtained from Dr. Ryszard Kole's lab[57] ubiquitously express EGFP-654 transgene throughout body under chicken β-actin promoter. Identical to the HeLa 654 cell line, a mutated nucleotide 654 at intron 2 of human β-globin gene interrupts EGFP-654 coding sequence and prevents proper translation of EGFP protein. The antisense activity of PMO blocks aberrant splicing and resulted in EGFP expression, the same as in the HeLa 654 assay. In this study, 6- to 8- week-old male EGFP-654 mice bred at Charles River Laboratory were shipped to

the vivarium at Sarepta Therapeutics (Cambridge, MA). These mice were group housed with ad libitum access to food and water. All animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Sarepta Therapeutics.

After 3 days of acclimation, mice were randomized into groups to receive a single *i.v.* tail vein injection of either saline or PMO-peptide (PMO-Mach3 or PMO-Mach4) at the indicated doses; 5, 10 and 30 mg/kg. 7-days after the injection, the mice were euthanized for serum and tissue sample collection. Quadriceps, diaphragm, heart were rapidly dissected, snap-frozen in liquid nitrogen and stored at -80 °C until analysis.

Serum from all groups were collected 7 days post-injection and tested for kidney injury markers using a Vet Axcel Clinical Chemistry System (Alfa Wassermann Diagnostic Technologies, LLC.) Specifically, serum BUN, creatinine, and cystatin C levels were measured using ACE® Creatinine Reagent (Alfa Wassermann, Cat# SA1012), ACE® Blood Urea Nitrogen Reagent (Alfa Wassermann, Cat# SA2024) and Diazyme Cystatin C immunoassay (Diazyme Laboratories, Cat# DX133C-K), respectively, per manufacturer's recommendation. 20-25 mg of mouse tissue was homogenized in RIPA buffer (Thermo Fisher, Cat# 89900) with protease inhibitor cocktail (Roche, 04693124001) using a Fast Prep 24-5G instrument (MP Biomedical). Homogenates were centrifuged at 12,000 g for 10 min at 4 °C. The resultant supernatant lysates were quantified by Pierce BCA Protein Assay Kit (Thermo Fisher, Cat# 23225) and saved for EGFP expression measurement. Specifically, 80 µg of lysates were aliquoted in each well in a black-wall clear-bottom 96-well microplate (Corning). EGFP fluorescent intensity of each sample was measured in duplicates using a SpectraMAx i3x microplate reader (Molecular devices) by default setting. The average EGFP fluorescent intensity of each sample was then plotted against a standard curve constructed by recombinant EGFP protein (Origen, Cat#TP790050) to quantify EGFP protein level per µg protein lysate.

2.4.10 General development of machine learning model

Inverse Design Model:

Generator – Recurrent Neural Network. The generator is a data-driven tool to generate new peptide sequences that follow the ontology of cell penetrating peptides to seed the optimization from likely starting points, and is based on recurrent neural network (RNN) - Nested LSTM architecture.[39] It was trained using one-hot encoding representations of the amino acids to predict the next amino acid in the sequence, from the preceding sequence. The inputs were size 5 to 50

amino acids, left-padding with zeros and representing termination with a unique token. The training dataset comprised of 1,150 sequences and a total of 19,800 sequence-next character pairs, including the non-modular sequences used in the creation of the library and sequences from CPPSite2.0.[40] The training was performed using 80% of this dataset, and validated using the remaining 20%. A validation accuracy of 76% was obtained in the training. For the model, multiple combinations of LSTM and Nested LSTM layers were tried with different cell sizes.[39] The final model was chosen after the optimization of hyperparameters. All hyperparameters were optimized using SigOpt.[58]

Predictor – Convolutional Neural Network. The predictor, based on convolutional neural network (CNN), estimates the normalized fluorescence intensity from PMO delivery by a given peptide sequence, as measured in the HeLa 654 assay. The model was trained on a row matrix of residue fingerprints. The row matrix of 2048-bit vectors (vector of 0s and 1s) represents the arrangement of the residues along the backbone of the peptide chain. This representation is analogous to 1D image with 2048 color channels. Fingerprints have radius 3 and were generated using RDKit.[59] By combining the CPP library from this work as well as the collection of CPPs from previous work, we compiled 640 PMO-peptides sequences for training.[7]

We used fingerprints and one-hot encodings to train non-CNN models such as those based on support vector regression, Gaussian process regression, kernel ridge regression, k-nearest neighbors regression and XGBoost regression.

Optimizer. The optimization was done using genetic algorithm (GA), where single residue mutations involved insertion, deletion and swapping, and multi-residue mutation was done using hybridization. For hybridization, the sequence length and position to be hybridized, and the hybridized sequence (from the list of all CPPs) were all chosen randomly. In the case of hybridization mutation, the selection and replacement of motifs was done at random without conservation of the sequence length. For the case of mutations with cysteine macrocycles, explicit conditions were built in to keep the number and position of cysteine residues separate in the case of a single through-space covalent bond or bicycle. A constrained hybridization condition conserving the sequence length was also set-up for specific optimization tasks. In the case of cysteine macrocycles, different fingerprints were used to denote the residues. The GA was used the following objective function, starting from LSTM-generated sequences and taking 1000 evolution steps:

$$GA\ Score = \frac{1}{2}\ Intensity - \frac{1}{2}\left(\frac{1}{2}\ R_{count} + \frac{1}{5}\ Length - \frac{1}{10}\ Net\ Charge + Similarity\right)$$

Set up of Generator-Predictor-Optimizer Loop: The generator was primed with a 5-long random sequence from the training dataset and sampled until a termination character is produced. The randomly sampled sequences were then set-up for optimization. The directed evolution of the generated sequences was carried using the predictor-optimizer feedback loop. Each sequence was mutated by the optimizer. Post mutations, the normalized fluorescence values for the new sequence was predicted by the predictor and the optimization parameters (similarity, % arginine, length, net charge) were calculated. The objective function (equation with optimization parameters) was evaluated for both the old and mutated sequences. If the value for the mutated sequence was higher for the mutated than the older, then the old sequence was replaced by the mutated sequence. 1000 such optimization rounds were conducted for each sequence. The output was hundreds of sequences with varying predicted activity.

We benchmarked the fingerprint (FP) representation and convolutional neural network (CNN) model against other model architectures and one-hot representation-based models in their ability to predict activities of both library and Mach sequences (Figs. S2, S3; Tables S2, S3). 8 scikit-learn model architectures (ridge, lasso, stochastic gradient descent, gaussian process, random forest, support vector and gradient boosting regression) and extreme gradient boosting regression were evaluated.[60,61]

Relevant hyperparameters for every model were optimized with Bayesian search using scikit-optimize (Table S1).[62] The hyperparameter optimization was done by 3-fold cross-validation, with the objective function as minimization of average root mean-squared error (RMSE) on the randomly held out validation dataset. The metrics have been reported on the test dataset. The train-valid-test dataset split was 70:10:20.

Table 2.8 Hyperparameters and optimized hyperparameters for regression and classification model architectures*

Model Architecture	Hyperparameters	Optimized parameters			
		Regression - Fingerprint	Regression - One Hot	Classification - Fingerprint	Classification - One Hot
Ridge	fit_intercept, Categorical, [True, False];	alpha=546.999, fit_intercept=True,	alpha=41.986, fit_intercept=True,	alpha=118.608, fit_intercept=False,	alpha=96.513, fit_intercept=False,

	normalize, Categorical, [True, False]; alpha, Real, [1e-3, 1e3]; solver, Categorical, ['svd', 'cholesky', 'lsqr', 'sparse_cg', 'sag', 'saga']	normalize=False, solver=sparse_cg	normalize=False, solver=sag	normalize=True, solver=svd	normalize=True, solver=cholesky
Lasso	fit_intercept, Categorical, [True, False]; normalize, Categorical, [True, False]; alpha, Real, [1e-3, 1e3]; precompute, Categorical, [True, False]; selection, Categorical, [‘selection’, ‘random’]	alpha=0.014, fit_intercept=False, normalize=True,precompute=False,selection=random	alpha=0.007, fit_intercept=True, normalize=False, precompute=True selection=random	-	-
SGD	loss, Categorical, ['squared_loss', 'epsilon_insensitive', 'huber', 'squared_epsilon_insensitive'; penalty, Categorical; ['l1', 'l2', 'elasticnet']; alpha, Real, [1e-3, 1e3]; fit_intercept, Categorical, [True, False]; l1_ratio, Real, [1e-3, 1]; learning_rate, Categorical, ['invscaling', 'constant', 'optimal', 'adaptive']; epsilon, Real, [1e-3, 1e3]; eta0, Real, [1e-2, 10]; power_t, Real, [1e-2, 10]; average, Categorical, [True, False]	alpha=0.002, average=True, epsilon=0.004, eta0=1.738, fit_intercept=True, l1_ratio=0.003, learning_rate=invscaling, loss=huber, penalty=l2, power_t=0.544	alpha=0.002, average=True, epsilon=0.004, eta0=1.738, fit_intercept=True, l1_ratio=0.003, learning_rate=invscaling, loss=huber, penalty=l2, power_t=0.544	alpha=0.001, average=True, epsilon=0.077, eta0=0.026, fit_intercept=False, l1_ratio=0.005, learning_rate=optimal, loss=log, penalty=l2, power_t=4.976	alpha=0.023, average=False, epsilon=0.009, eta0=0.01, fit_intercept=True, l1_ratio=0.1, learning_rate=adaptive, loss=hinge, penalty=l2, power_t=0.01
Gaussian Process	kernel, Categorical, [Matern, RBF, DotProduct]; alpha, Real, [1e-11, 1e-6]; n_restarts_optimizer, Integer, [0, 10]	alpha=1.57e-7, kernel=Matern, n_restarts_optimizers=8	alpha=7.320e-9, kernel=Matern, n_restarts_optimizer=10	kernel=Matern, n_restarts_optimizers=5	kernel=Matern, n_restarts_optimizers=5
Random Forest	criterion, Categorical, ['mse', 'mae']-Regression; criterion, Categorical, ['mse', 'entropy]-Classification n_estimators, Integer, [10, 1000]; max_depth, Integer, [1, 10]	criterion=’mse’, max_depth=7, n_estimators=626	criterion=’mse’, max_depth=6, n_estimators=127	criterion=entropy, max_depth=8, n_estimators=908	criterion=entropy, max_depth=10, n_estimators=630
XGBoost	gamma, Real, [1e-6, 10]; eta,Real, [1e-3, 1]; max_depth, Integer, [1, 10]; tree_method, Categorical, ['auto', 'exact', 'approx', 'hist']; alpha, Real, [1e-3, 1e3]; lambda, Real, [1e-3, 1e3]; sketch_eps,Real, [1e-3, 1]	alpha=0.035, eta=0.864, gamma=2.285, reg_lambda=3.945, sketch_eps=0.003, tree_method='exact'	eta=0.864, gamma=2.285, reg_alpha=0.035, reg_lambda=3.945, sketch_eps=0.003,tree_method=’exact'	alpha=2.407, eta=0.035, lambda=0.004, max_depth=5, tree_method=exact	alpha=0.835, eta=0.155, lambda=15.360, max_depth=4, tree_method=auto
Support Vector	kernel, Categorical, ['linear', 'poly', 'rbf', 'sigmoid']; degree, Integer, [1, 6]; gamma, Real, [1e-6, 10]; C, Real, [1e-2, 10], epsilon,Real, [1e-3, 10]; shrinking, Categorical, [True, False]	C=2.671, epsilon=0.009, gamma=0.002, kernel='linear'	C=0.024, epsilon=0.220, gamma=4.966e-5, kernel='linear'	C=0.030, degree=5, gamma=0.011, kernel=linear, shrinking=False	C=0.288, degree=5, gamma=0.076, kernel='poly', probability=True, shrinking=False

Gradient Boosting	loss, Categorical, ['ls', 'lad', 'huber', 'quantile']; learning_rate, Real, [1e-2, 1]; n_estimators, Integer, [10, 1000]; criterion, Categorical, ['friedman_mse', 'mse', 'mae']; max_depth, Integer, [1, 10]	criterion='mse', learning_rate=0.018, loss='huber', max_depth=6, n_estimators=250	criterion='mse', learning_rate=0.018, loss='huber', max_depth=6, n_estimators=250	criterion=mse, learning_rate=0.018, loss=deviance, max_depth=6, n_estimators=250	criterion=mse, learning_rate=0.018, loss=deviance, max_depth=6, n_estimators=250
Nearest Neighbors	weights, Categorical, [‘uniform’, ‘distance’]; leaf_size, Integer, [10, 100]; n_neighbors, Integer, [2, 20]; algorithm, Categorical, ['auto', 'ball_tree', 'kd_tree', 'brute']; p, Integer, [1, 5]	algorithm=ball_tree, leaf_size=16, n_neighbors=7, p=5, weights=uniform	algorithm=brute, leaf_size=56, n_neighbors=4, p=3, weights=uniform	algorithm=auto, leaf_size=57, n_neighbors=4, p=4, weights=uniform	algorithm=auto, leaf_size=57, n_neighbors=4, p=4, weights=uniform

* The hyperparameters follow a notation – parameter, datatype, values. Datatypes are categorical, integer and real. For categorical datatype, the list of hyperparameters is noted, and for integer and real datatypes, minimum and maximum values are noted.

2.4.11 Analysis and benchmarking of CNN predictor model

We evaluated the CNN-fingerprint (CNN-FP) models against individual models and model ensembles trained with FP and one-hot encoding representations.

RMSE and other metrics between the predicted and experimental activity values were used to compare individual models. For the validation dataset, random forest-FP (RF-FP) has lowest RMSE, but RF models cannot extrapolate outside the range of the training data, so they are limited toward the task of designing more active peptides. For the activity values of the Mach sequences, other models such as CNN one-hot had better RMSE, R^2, Pearson and Spearman's rank correlation coefficients. Based on the RMSE values, the best performing model architecture was one-hot based CNN. In practice, however, we observed that this model was not able to extrapolate activities beyond the training data, whereas the CNN-FP model could. We determined that RMSE and other metrics for the CNN-FP model are significantly affected by outlying predicted activity values. Upon removing the outlier (Mach12 with 140 predicted activity by CNN-FP model), we observed that the CNN-FP model outperforms all other models in terms of RMSE and R^2.

Unlike CNN-FP, none of the simpler models predict the activity of Mach peptides to be above the maximum of training dataset, as apparent in the parity plot and the RMSE and R^2 metrics for the Mach dataset without the outlier (Table S2). This experiment shows that simpler models are limited by the range of the training data, and are unable to extrapolate in the co-domain space. While other models may be able to produce sequences with high (>20-fold) experimental activity, the ability to extrapolate predicted activities is critical for the informed selection of predicted sequences to validate. Extrapolation is a necessary model feature for our goal of designing sequences with activities higher than those in the training set.

To mitigate the role of outliers that impact performance, we evaluated the use of model ensembling. Ensembled CNN one-hot model performed the best amongst all models on the validation dataset, while ensembled CNN-FP outperformed model-feature combinations on RMSE for Mach dataset, with and without the outlier (Table S3). Although the choice of sequences for experimental validation was not based on predictions from ensemble models, we note that ensemble models can robustly extrapolate predictions outside the training data for future studies.

From our analysis, we observe that simpler models can complement the CNN predictions in decision-making, such as ranking of predicted peptides, as noted from the high Pearson's and

Spearman's correlation coefficients. The CNN model is still necessary to be able to predict peptides with higher activity than the training set.

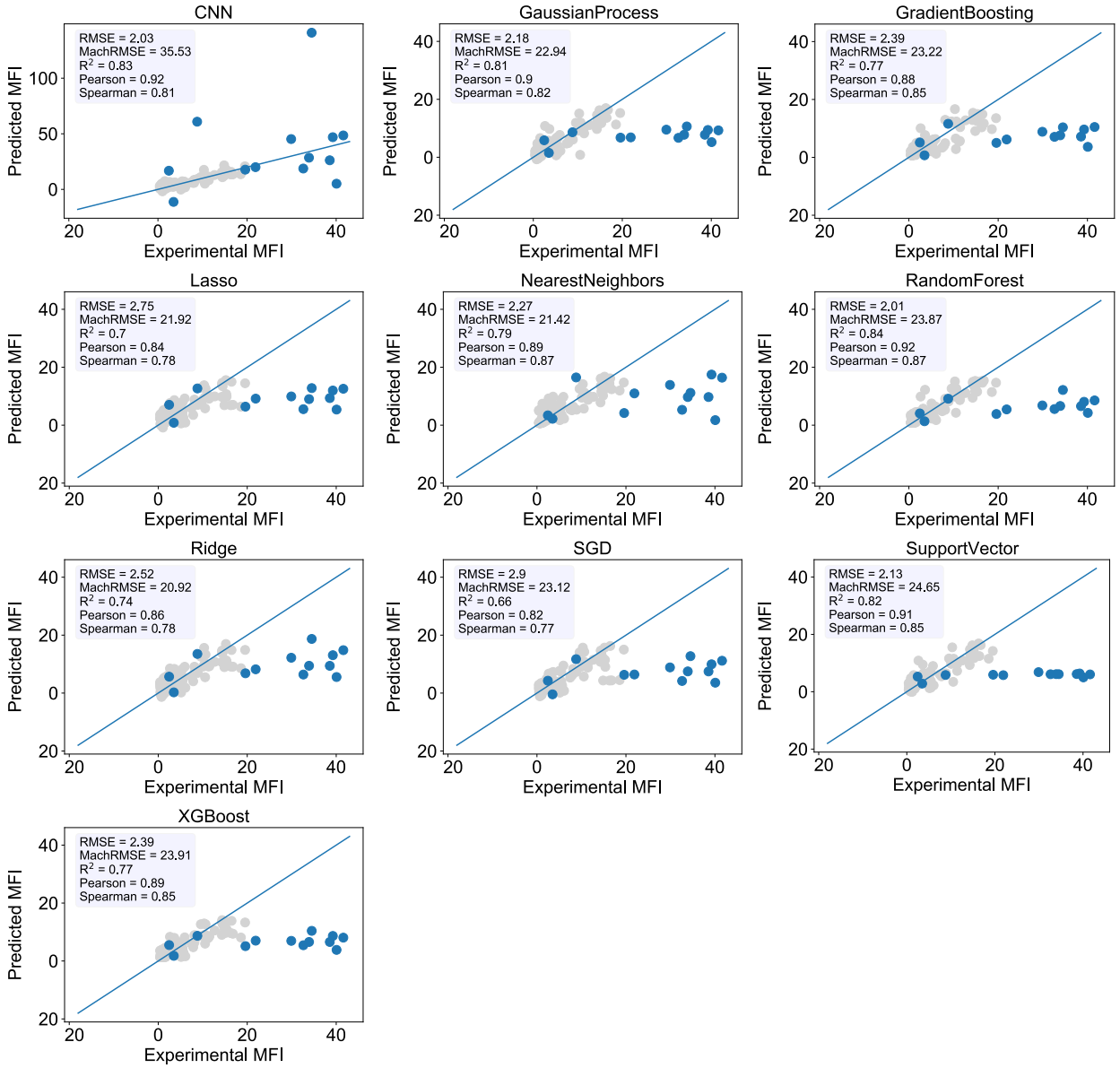


Figure 2.27 Parity plots for CNN and other models trained using 2048-bit fingerprints Parity plots for CNN and other models trained using 2048-bit fingerprints are shown. Random forest regressor performs best on the validation set, however is unable to extrapolate for Mach peptides. The second best, and the optimal model is support vector regression. The held-out data for validation of the model is shown in grey, and predictions for Mach peptides are shown in blue. Key evaluation metrics have been noted in the data inset. Only the CNN model shows a range of predicted values above the training data, as do the Mach peptides.

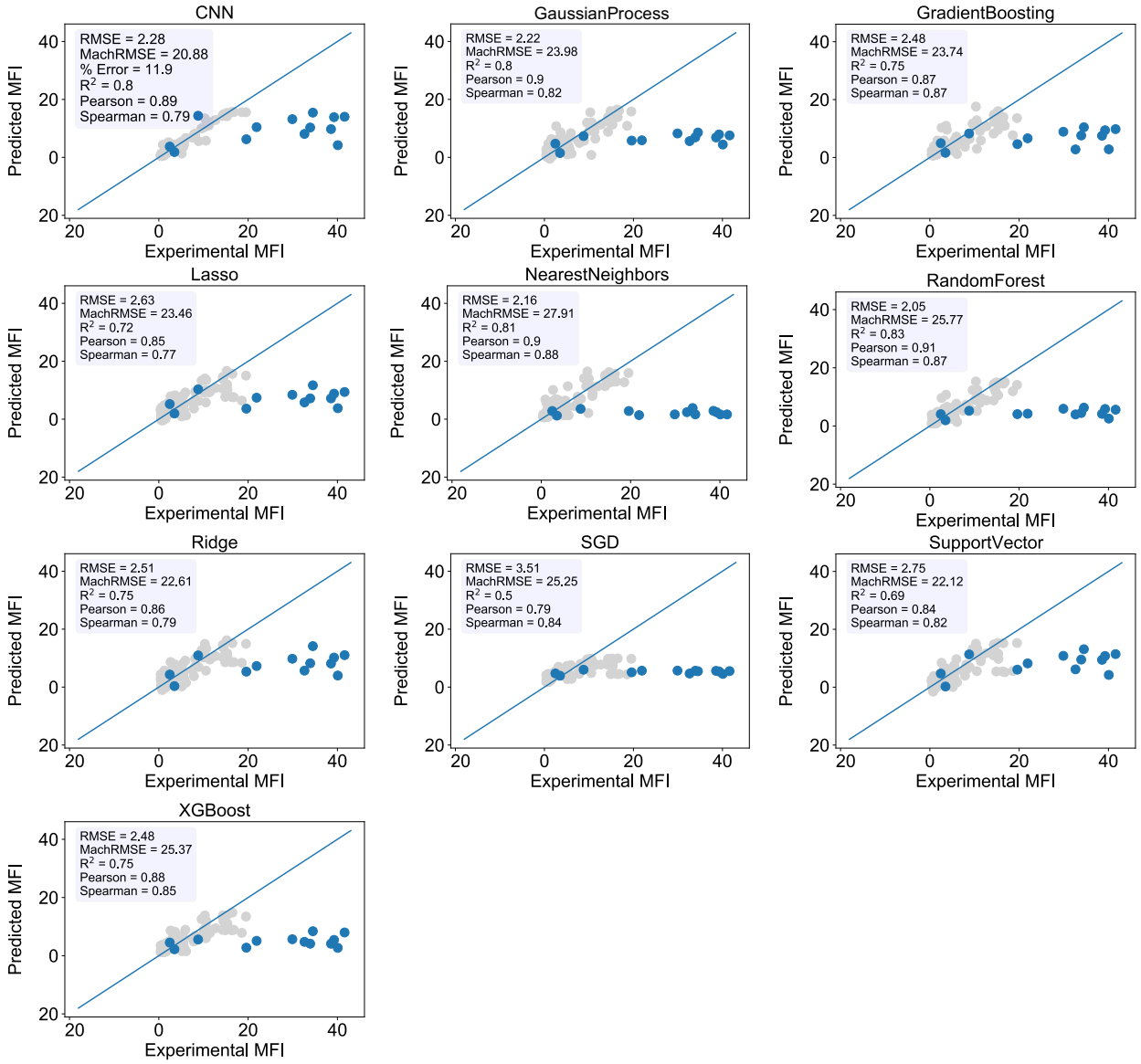


Figure 2.28 Parity plots for CNN and other models trained using one-hot encodings Parity plots for CNN and other models trained using one-hot encodings are shown. The held-out data for validation of the model is shown in grey, and predictions for Mach peptides are shown in blue. Key evaluation metrics have been noted in the data inset. Only the CNN model shows a range of predicted values above the training data, as do the Mach peptides.

Table 2.9 Metrics for CNN model compared to other model architectures using one hot encodings*

Model - Feature	Validation						Mach				Mach, without outlier			
	uRMSE	RMSE	% Error	R2	r	ρ	RMSE	R2	r	ρ	RMSE	R2	r	ρ
CNN-FP	0.41	2.03	10.55	0.83	0.92	0.81	35.53	-5.90	0.29	0.37	20.16	-1.11	0.22	0.35
Ridge-FP	0.51	2.51	13.04	0.75	0.87	0.78	21.23	-1.46	0.46	0.41	21.56	-1.41	0.45	0.39
Lasso-FP	0.55	2.74	14.24	0.70	0.84	0.78	22.23	-1.70	0.46	0.41	22.34	-1.59	0.44	0.38
SGD-FP	0.61	3.03	15.76	0.63	0.80	0.81	23.28	-1.96	0.42	0.35	23.33	-1.83	0.40	0.32
GP-FP	0.44	2.17	11.32	0.81	0.90	0.82	22.93	-1.88	0.55	0.40	22.86	-1.71	0.53	0.38
RF-FP	0.41	2.02	10.51	0.84	0.92	0.87	23.83	-2.10	0.49	0.46	23.93	-1.97	0.50	0.45
XGBoost-FP	0.47	2.34	12.19	0.78	0.88	0.82	23.68	-2.07	0.49	0.47	23.81	-1.94	0.49	0.45
SVR-FP	0.43	2.13	11.06	0.82	0.91	0.85	24.65	-2.32	0.57	0.43	24.32	-2.07	0.56	0.44
GB-FP	0.48	2.37	12.33	0.77	0.88	0.86	23.19	-1.94	0.44	0.33	23.11	-1.77	0.41	0.34
kNN-FP	0.46	2.27	11.81	0.79	0.89	0.87	21.42	-1.51	0.35	0.29	21.25	-1.34	0.34	0.28
CNN-Onehot	0.46	2.28	11.90	0.80	0.89	0.79	20.88	-1.38	0.47	0.38	21.02	-1.29	0.45	0.37
Ridge-Onehot	0.51	2.51	13.08	0.75	0.87	0.79	22.70	-1.82	0.48	0.42	22.86	-1.71	0.46	0.41
Lasso-Onehot	0.53	2.62	13.64	0.72	0.85	0.78	23.28	-1.96	0.39	0.36	23.36	-1.83	0.36	0.33
SGD-Onehot	0.73	3.61	18.78	0.48	0.74	0.80	25.68	-2.61	0.28	0.03	25.33	-2.33	0.27	0.05
GP-Onehot	0.45	2.21	11.51	0.80	0.90	0.82	24.02	-2.16	0.56	0.39	23.85	-1.95	0.54	0.38
RF-Onehot	0.42	2.06	10.74	0.83	0.91	0.86	25.89	-2.67	0.31	0.24	25.62	-2.41	0.28	0.22
XGBoost-Onehot	0.50	2.49	12.96	0.75	0.87	0.78	25.83	-2.65	0.31	0.23	25.80	-2.46	0.27	0.21
SVR-Onehot	0.55	2.74	14.26	0.70	0.84	0.82	22.12	-1.68	0.51	0.40	22.18	-1.55	0.49	0.38
GB-Onehot	0.50	2.49	12.97	0.75	0.87	0.87	23.54	-2.03	0.45	0.41	23.55	-1.88	0.43	0.39
kNN-Onehot	0.44	2.16	11.24	0.81	0.90	0.88	27.91	-3.26	-0.10	-0.07	27.46	-2.92	-0.07	-0.05

*The best values for each metric have been highlighted in red. The RF model slightly outperforms the original CNN in the validation dataset metrics; however, it is known to be limited in predicting within the range of training data only. As regards testing against the Mach dataset, CNN-Onehot model outperforms the CNN-FP model. However, upon removing the outlier sequence (CNN-FP predicted activity: 140), CNN-FP turns out to be the most optimal model. **r** and **ρ** refers to Pearson's and Spearman's correlation, respectively.

Validation loss is defined as unitless root-mean-squared (uRMSE), since the training and validation data are normalized through scaling by the standard deviation of the training data (σ). Re-scaled RMSE is defined as root-mean-squared error in fold-over-PMO units (uRMSE × σ). %Error is defined through the equation below, where range is the difference between the maximum and minimum value in the training data (19.52 and 0.31 respectively) in fold-over-PMO units.

$$\%\, Error = \frac{uRMSE \times \boldsymbol{\sigma}}{range} \times 100\%$$

Table 2.10 Metrics for CNN model compared to other model architectures using fingerprint representations*

Model - Feature	Validation						Mach				Mach, without outlier			
	uRMSE	RMSE	% Error	R2	r	ρ	RMSE	R2	r	ρ	RMSE	R2	r	ρ
CNN-FP	0.41	2.03	10.58	0.83	0.91	0.80	19.89	-1.42	0.33	0.34	17.49	-0.59	0.30	0.32
Ridge-FP	0.51	2.51	13.04	0.75	0.87	0.78	21.23	-1.46	0.46	0.41	21.56	-1.41	0.45	0.39
Lasso-FP	0.55	2.74	14.25	0.70	0.84	0.78	22.23	-1.70	0.46	0.41	22.34	-1.59	0.44	0.38
SGD-FP	0.61	3.01	15.64	0.64	0.80	0.79	23.36	-1.98	0.47	0.40	23.34	-1.83	0.44	0.38
GP-FP	0.44	2.17	11.32	0.81	0.90	0.82	22.94	-1.88	0.55	0.40	22.86	-1.71	0.53	0.38
RF-FP	0.41	2.01	10.48	0.84	0.92	0.87	23.97	-2.14	0.47	0.48	24.10	-2.02	0.49	0.47
XGBoost-FP	0.47	2.34	12.19	0.78	0.88	0.82	23.68	-2.07	0.49	0.47	23.81	-1.94	0.49	0.45
SVR-FP*	0.43	2.13	11.06	0.82	0.91	0.85	24.65	-2.32	0.57	0.43	24.32	-2.07	0.56	0.44
GB-FP	0.48	2.38	12.37	0.77	0.88	0.86	23.12	-1.92	0.42	0.34	23.05	-1.76	0.39	0.34
kNN-FP*	0.46	2.27	11.81	0.79	0.89	0.87	21.42	-1.51	0.35	0.29	21.25	-1.34	0.34	0.28
CNN-Onehot	0.34	1.66	8.64	0.89	0.94	0.85	20.29	-1.25	0.48	0.46	20.48	-1.18	0.46	0.45
Ridge-Onehot	0.61	3.03	15.77	0.63	0.82	0.83	25.13	-2.45	0.39	0.20	24.84	-2.20	0.36	0.20
Lasso-Onehot	0.57	2.82	14.66	0.68	0.83	0.81	23.21	-1.95	0.31	0.31	23.10	-1.77	0.27	0.28
SGD-Onehot	0.73	3.60	18.75	0.48	0.74	0.80	25.52	-2.56	0.29	0.01	25.17	-2.29	0.27	0.02
GP-Onehot	0.45	2.21	11.51	0.80	0.90	0.82	23.87	-2.12	0.55	0.39	23.72	-1.92	0.54	0.38
RF-Onehot	0.41	2.01	10.48	0.84	0.92	0.87	25.75	-2.63	0.33	0.29	25.50	-2.38	0.29	0.26
XGBoost-Onehot	0.50	2.49	12.96	0.75	0.87	0.78	25.83	-2.65	0.31	0.23	25.80	-2.46	0.27	0.21
SVR-Onehot*	0.55	2.74	14.26	0.70	0.84	0.82	22.12	-1.68	0.51	0.40	22.18	-1.55	0.49	0.38
GB-Onehot	0.50	2.48	12.91	0.75	0.87	0.87	23.76	-2.09	0.40	0.42	23.77	-1.93	0.37	0.40
kNN-Onehot*	0.44	2.16	11.24	0.81	0.90	0.88	27.91	-3.26	-0.10	-0.07	27.46	-2.92	-0.07	-0.05

5 models with distinct random initialization seeds were trained for all possible model-feature combinations (models with '' do not have a random state initialization, in the sklearn implementation). In comparison to metrics obtained from individual models, we note that both CNN models, based on FP and one-hot encoding respectively, stand out for the ensemble models. CNN one-hot is the optimal model on the validation dataset. CNN-FP outperforms CNN-One-hot and other models, however, on RMSE for Mach dataset and Mach dataset without the outlier. **r** and **ρ** refers to Pearson's and Spearman's correlation, respectively.

We benchmarked our regression models with corresponding classification models for both fingerprint and one-hot encoding representations (Tables S4, S5). The classification model architectures are similar to those reported in the literature for CPP prediction.[63–67] Similar to our earlier work, the classes were obtained by setting fold over PMO activity threshold of 3.0, above which the sequences were classified as active, otherwise inactive.[68] The hyperparameter optimization and train-valid-test split was same as the benchmarking of regression models (SI Section 2.1, Table S1). For the CNN models, the architecture was kept largely the same as regression models, with the only modifications being the activation function of last layer as sigmoid and loss function as binary crossentropy, which are conventional modifications for classification model architecture. In addition to conventional metrics for evaluating classification models, we used Spearman's coefficient to estimate the rank correlation between the predicted probabilities from the classification models and experimental MFI values of the sequences.

The best metrics for the performance of the classification models against the held-out validation and Mach datasets varied across different model architectures. The CNN, support vector, random forest and stochastic gradient descent models were the optimal models across both representations. This benchmarking experiment further confirms that CNN model architecture is optimal at both classification and regression tasks.

Table 2.11 The CNN model benchmarked against classification models using fingerprint representation*

	Validation Dataset Metrics						Mach Dataset Metrics					
	Accu-racy	Preci-sion	F1	Recall	ROC-AUC	Spear-man	Accu-racy	Preci-sion	F1	Recall	ROC-AUC	Spear-man
CNN-FP	0.85	0.81	0.85	0.92	0.92	0.81	0.46	0.92	0.66	0.82	0.84	1.00
Gaussian Process	0.84	0.69	0.77	0.87	0.85	0.82	0.54	0.50	0.67	1.00	0.57	0.01
Gradient Boosting	0.78	0.67	0.70	0.73	0.77	0.81	0.77	0.83	0.87	0.91	0.45	0.21
Nearest Neighbors	0.88	0.76	0.82	0.90	0.88	0.79	0.69	0.75	0.82	0.90	0.45	0.28
Random Forest	0.87	0.82	0.82	0.83	0.86	0.85	0.77	0.83	0.87	0.91	0.45	0.02
Ridge	0.90	0.86	0.87	0.88	0.89	0.75	0.85	0.92	0.92	0.92	0.46	0.39
SGD	0.89	0.86	0.86	0.86	0.88	0.72	0.77	0.83	0.87	0.91	0.45	0.00
Support Vector	0.91	0.88	0.88	0.88	0.90	0.75	0.77	0.83	0.87	0.91	0.45	0.00
XGBoost	0.84	0.73	0.77	0.82	0.83	0.84	0.46	0.50	0.63	0.86	0.43	0.22

***** Spearman correlation coefficient is calculated using predicted probabilities from the classification model and the experimental MFI values. The best values for each metric have been highlighted in red.

Table 2.12 The CNN model benchmarked against classification models using one hot encodings*

	Validation Dataset Metrics						Mach Dataset Metrics					
	Accuracy	**Precision**	**F1**	**Recall**	**ROC-AUC**	**Spear-man**	**Accuracy**	**Precision**	**F1**	**Recall**	**ROC-AUC**	**Spear-man**
CNN-One-Hot	0.88	0.83	0.77	0.87	0.83	0.62	0.45	0.91	0.58	0.82	0.87	0.62
Gaussian Process	0.87	0.73	0.81	0.90	0.88	0.83	0.77	0.75	0.86	1.00	0.63	0.01
Gradient Boosting	0.86	0.82	0.82	0.82	0.85	0.83	0.77	0.75	0.86	1.00	0.63	-0.27
Nearest Neighbors	0.88	0.80	0.84	0.89	0.88	0.82	0.08	0.00	0.00	0.00	N/A	0.27
Random Forest	0.90	0.82	0.86	0.91	0.90	0.84	0.77	0.75	0.86	1.00	0.63	-0.07
Ridge	0.87	0.80	0.82	0.85	0.86	0.69	0.77	0.83	0.87	0.91	0.45	0.00
SGD	0.90	0.88	0.87	0.86	0.89	0.76	0.77	0.83	0.87	0.91	0.45	0.00
Support Vector	0.84	0.78	0.78	0.79	0.83	0.77	0.08	0.00	0.00	0.00	N/A	0.12
XGBoost	0.88	0.80	0.83	0.87	0.87	0.83	0.46	0.42	0.59	1.00	0.56	0.09

* Spearman correlation coefficient is calculated using predicted probabilities from the classification model and the experimental MFI values. ROC-AUC for Mach dataset could not be calculated for nearest neighbors and support vector model architectures as all predicted probabilities were less than 0.50. The best values for each metric have been highlighted in red.

2.5 Acknowledgements

Data and materials availability: Computer code, training data, and trained models are available at https://github.com/learningmatter-mit/peptimizer. Tutorial Jupyter notebooks are also in the repository, and demo Google Colab notebooks can be found at github.com/pikulsomesh/tutorials. All other data is available in the main text or the supplementary materials.

Acknowledgments: We thank Alexander R. Loftis and Jacob Rodriguez for assistance with recombinant protein expression, Dr. Coralie Backlund for assistance with immunoassays, Wendy C. Salmon at the W. M. Keck Microscopy Facility at the Whitehead Institute for help with imaging, the Swanson Biotechnology Center Flow Cytometry Facility at the Koch Institute for the use of their flow cytometers, and Bryan Mastis and Sam Foley for help with in vivo studies. We also thank Zi-Ning Choo for igniting our interest in machine learning.

Funding: This research was funded by Sarepta Therapeutics, by the MIT-SenseTime Alliance on Artificial Intelligence, and by an award from the Abdul Latif Jameel Clinic for Machine Learning in Health (J-Clinic). C.K.S. (4000057398) acknowledges the National Science Foundation Graduate Research Fellowship (NSF Grant No. 1122374) for research support.

2.6 Author contributions:

C.K.S., S.M., J.M.W., B.L.P., and R.G.B. conceptualized the research; J.M.W. and C.M.F. synthesized and tested the modular library; S.M. and R.G.B. developed the machine learning model with input from C.K.S. and B.L.P.; C.K.S. synthesized Mach peptides and constructs, performed experiments and analyzed the results. K.B., C.L.W., and J.A.W. performed in vivo study with input from A.M.; C.K.S., S.M., A.L., B.L.P., and R.G.B. wrote the manuscript with input from all authors.

2.7 Appendix I: Library Sequences & Activity

Sequence and activity data used to train and validate the model. PMO is inherent in every input sequence as well as predicted sequence. Any C shown indicates a cysteine macrocycle linked by decafluorobiphenyl. X = aminohexanoic acid, B = beta-alanine, 2 = linker 2, and 3 = linker 3. Sequences are shown in descending order of activity.

Sequences	Activity
GRPRESGKKRKRKRLKP2PKKKRKV3RRWWRRWRR	20
GRPRESGKKRKRKRLKP2PKKKRKV3RXRRBRCRXRRBRC	19
GRPRESGKKRKRKRLKP2PKKKRKV3CRXRRBRRXRRBRC	19
AGYLLGKINLKALAALAKKIL2KRVK3CRXRRBRRXRRBRC	19
GRPRESGKKRKRKRLKP2PKKKRKV3RLRWR	18
GRPRESGKKRKRKRLKP2PKKKRKV3RXRRBRRXRRBR	18
AGYLLGKINLKALAALAKKIL2KRVK3RXRRBRRXRRBR	18
GRPRESGKKRKRKRLKP2PKKKRKV3RRRRRRCRRRRRRC	18
GRPRESGKKRKRKRLKP2PKKKRKV3TRSSRAGLQWPVGRVHRLLRK	18
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RRRRRRRRRRRR	18
GRPRESGKKRKRKRLKP2PKKKRKV3RGGRLSYSRRRFSTSTGR	18
GRPRESGKKRKRKRLKP2PKKKRKV3LLIILRRRIRKQAHAHSKRXRRBRR XRRBR	18
AGYLLGKINLKALAALAKKIL2KRVK3RRRRRRRRRRRR	18
GRPRESGKKRKRKRLKP2KRVK3RRRRRRRRR	17
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RXRRBRRXRRBR	17
GRPRESGKKRKRKRLKP2KRVK3RKKRRQRRR	17
GRPRESGKKRKRKRLKP2PKKKRKV3CRRRRRRCCRRRRRRC	17
AGYLLGKINLKALAALAKKIL2KRVK3CRRRRRRCCRRRRRRC	17
AGYLLGKINLKALAALAKKIL2KRVK3CRRRRRRRRRRRRC	17
GRPRESGKKRKRKRLKP2PKKKRKV3CRRRRRRRRRRRRC	17
GRPRESGKKRKRKRLKP2PKKKRKV3PPRPPRPPRPPR	17
GRPRESGKKRKRKRLKP2PKKKRKV3RRIRPRPPRLPRPRPRPLPFPRPG	16
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CRXRRBRRXRRBRC	16
GRPRESGKKRKRKRLKP2PKKKRKV3CRRRRRRCRRRRRRC	16
AGYLLGKINLKALAALAKKIL2KRVK3CRRRRRRCRRRRRR	16
AGYLLGKINLKALAALAKKIL2KRVK3RRRRRRCRRRRRRC	16
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RKKRRQRRR	16
AGYLLGKINLKALAALAKKIL2KRVK3RXRRBRCRXRRBRC	16
AGYLLGKINLKALAALAKKIL2KRVK3GRPRESGKKRKRKRLKP	16
GRPRESGKKRKRKRLKP2PKKKRKV3RRRRRRRRR	16
GRPRESGKKRKRKRLKP2PKKKRKV3PPRPPRPPR	16
GRPRESGKKRKRKRLKP2PKKKRKV3CRRRRRRCRRRRRR	16

RQIKIWFQNRRMKWKK2PKKKRKV3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	16
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RXRRBRCRXRRBRC	16
GRPRESGKKRKRKRLKP2PKKKRKV3RKKRRQRRR	15
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RRRRRRRRR	15
GRPRESGKKRKRKRLKP2PKKKRKV3RQIKIWFQNRRMKWKK	15
GRPRESGKKRKRKRLKP2KRVK3RGGRLSYSRRRFSTSTGR	15
GRPRESGKKRKRKRLKP2PKKKRKV3CQIKIWFCNKRAKIKK	15
GRPRESGKKRKRKRLKP2PKKKRKV3RQIKIWFQNRRMKWKK	15
GRPRESGKKRKRKRLKP2PKKKRKV3KGTYKKKLMRIPLKGT	15
GRPRESGKKRKRKRLKP2PKKKRKV3GAYDLRRRERQSRLRRRERQSR	15
GRPRESGKKRKRKRLKP2PKKKRKV3ALWKTLLKKVLKAPKKKRKV	15
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CRRRRRRCRRRRRR	15
AGYLLGKINLKALAALAKKIL2KRVK3CRRRRRRCRRRRRRC	15
GRPRESGKKRKRKRLKP2KRVK3RRWWRRWRR	15
RQIKIWFQNRRMKWKK2KRVK3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	15
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RRIRPRPPRLPRPRPRPLPFPRPG	15
AGYLLGKINLKALAALAKKIL2KRVK3GRKAARAPGRRKQ	15
GRPRESGKKRKRKRLKP2PKKKRKV3GRPRESGKKRKRKRLKP	15
RQIKIWFQNRRMKWKK2PKKKRKV3TRSSRAGLQWPVGRVHRLLRK	15
GRPRESGKKRKRKRLKP2PKKKRKV3LYKKGPAKKGRPPLRGWFH	14
GRPRESGKKRKRKRLKP2PKKKRKV3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	14
GRPRESGKKRKRKRLKP2PKKKRKV3KKYRGRKRHPR	14
GRPRESGKKRKRKRLKP2PKKKRKV3GRKAARAPGRRKQ	14
GRPRESGKKRKRKRLKP2KRVK3ALWKTLLKKVLKAPKKKRKV	14
GRPRESGKKRKRKRLKP2KRVK3LYKKGPAKKGRPPLRGWFH	14
GRPRESGKKRKRKRLKP2KRVK3RRIRPRPPRLPRPRPRPLPFPRPG	14
RQIKIWFQNRRMKWKK2PKKKRKV3CRRRRRRCCRRRRRRC	14
GRPRESGKKRKRKRLKP2KRVK3RLRWR	14
GRPRESGKKRKRKRLKP2PKKKRKV3RRRRRRRRRRRR	14
GRPRESGKKRKRKRLKP2KRVK3GAYDLRRRERQSRLRRRERQSR	14
AGYLLGKINLKALAALAKKIL2KRVK3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	14
GRPRESGKKRKRKRLKP2PKKKRKV3KRVKAGYLLGKINLKALAALAKKIL	13
GRPRESGKKRKRKRLKP2KRVK3KGTYKKKLMRIPLKGT	13
GRPRESGKKRKRKRLKP2PKKKRKV3RQVTIWSQNRRVKSKK	13
GRPRESGKKRKRKRLKP2KRVK3RQVTIWSQNRRVKSKK	13

GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RRWWRRWRR	13
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RRRRRRCRRRRRRC	13
AGYLLGKINLKALAALAKKIL2KRVK3KKYRGRKRHPR	13
GRPRESGKKRKRKRLKP2KRVK3PPRPPRPPRPPR	13
GRPRESGKKRKRKRLKP2PKKKRKV3CRLRWRC	13
GRPRESGKKRKRKRLKP2PKKKRKV3LKTLTETLKELTKTLTEL	13
GRPRESGKKRKRKRLKP2KRVK3TRSSRAGLQWPVGRVHRLLRK	13
RQIKIWFQNRRMKWKK2KRVK3CRRRRRRCCRRRRRRC	12
AGYLLGKINLKALAALAKKIL2KRVK3RQIKIWFQNRRMKWKK	12
GRPRESGKKRKRKRLKP2KRVK3PPRPPRPPR	12
RQIKIWFQNRRMKWKK2PKKKRKV3CRXRRBRRXRRBRC	12
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CRRRRRRRRRRRRC	12
GRPRESGKKRKRKRLKP2PKKKRKV3AGYLLGKINLKALAALAKKILKRVK	12
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3GRPRESGKKRKRKRLKP	12
GRPRESGKKRKRKRLKP2PKKKRKV3CSQIKIWFQNKRAKIKKC	11
GRPRESGKKRKRKRLKP2KRVK3RQIKIWFQNRRMKWKK	11
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	11
GRPRESGKKRKRKRLKP2KRVK3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	11
GRPRESGKKRKRKRLKP2KRVK3AGYLLGKINLKALAALAKKILKRVK	11
RQIKIWFQNRRMKWKK2PKKKRKV3RXRRBRCRXRRBRC	11
GRPRESGKKRKRKRLKP2PKKKRKV3PKKKRKVAGYLLGKINLKALAALAKKIL	11
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3ALWKTLLKKVLKAPKKKRKV	11
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CRRRRRRCRRRRRRC	11
GRPRESGKKRKRKRLKP2PKKKRKV3KLALKALKALKAALKLA	11
RQIKIWFQNRRMKWKK2PKKKRKV3RXRRBRRXRRBR	11
RQIKIWFQNRRMKWKK2KRVK3RXRRBRCRXRRBRC	11
GRPRESGKKRKRKRLKP2PKKKRKV3FKIYDKKVRTRVVKH	11
GRPRESGKKRKRKRLKP2PKKKRKV3IAWVKAFIRKLRKGPLG	11
RQIKIWFQNRRMKWKK2KRVK3CRRRRRRRRRRRRC	11
GRPRESGKKRKRKRLKP2KRVK3LKTLTETLKELTKTLTEL	11
GRPRESGKKRKRKRLKP2KRVK3PKKKRKVAGYLLGKINLKALAALAKKIL	10
GRPRESGKKRKRKRLKP2PKKKRKV3SQIKIWFQCKRAKIKC	10
RQIKIWFQNRRMKWKK2PKKKRKV3CRRRRRRRRRRRRC	10
RRRRRRRRRRR	10
RQIKIWFQNRRMKWKK2PKKKRKV3CRRRRRRCRRRRRRC	10

RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CRXRRBRRXRRBRC	10
GRPRESGKKRKRKRLKP2PKKKRKV3LLIILRRRIRKQAHAHSK	10
GRPRESGKKRKRKRLKP2KRVK3KRVKAGYLLGKINLKALAALAKKIL	10
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3KLALKALKALKAALKLA	10
GRPRESGKKRKRKRLKP2KRVK3KLALKALKALKAALKLA	10
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CRRRRRRCCRRRRRRC	10
RQIKIWFQNRRMKWKK2KRVK3RRRRRRRRRRRR	10
GRPRESGKKRKRKRLKP2PKKKRKV3GSPWGLQHHPPRT	10
GRPRESGKKRKRKRLKP2PKKKRKV3VRLPPPVRLPPPVRLPPP	10
RQIKIWFQNRRMKWKK2PKKKRKV3GRKAARAPGRRKQ	9
RQIKIWFQNRRMKWKK2KRVK3RXRRBRRXRRBR	9
GRPRESGKKRKRKRLKP2PKKKRKV3RSVTIWFQSRRVKEKK	9
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3KKYRGRKRHPR	9
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3PKKKRKVAGYLLGKINLKALAALAKKIL	9
RQIKIWFQNRRMKWKK2KRVK3CRRRRRRCRRRRRRC	9
GRPRESGKKRKRKRLKP2PKKKRKV3VSALK	9
GRPRESGKKRKRKRLKP2KRVK3FKIYDKKVRTRVVKH	9
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3KRVKAGYLLGKINLKALAALAKKIL	9
RQIKIWFQNRRMKWKK2KRVK3RRWWRRWRR	9
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RXRRBRCRXRRBRC	9
GRPRESGKKRKRKRLKP2PKKKRKV3PLSSIFSRIGDP	8
RQIKIWFQNRRMKWKK2PKKKRKV3RRRRRRCRRRRRRC	8
RQIKIWFQNRRMKWKK2PKKKRKV3KGTYKKKLMRIPLKGT	8
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RGGRLSYSRRRFSTSTGR	8
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3AGYLLGKINLKALAALAKKILKRVK	8
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RQVTIWSQNRRVKSKK	8
RQIKIWFQNRRMKWKK2KRVK3RRIRPRPPRLPRPRPRPLPFPRPG	8
RQIKIWFQNRRMKWKK2PKKKRKV3LLIILRRRIRKQAHAHSK	8
GRPRESGKKRKRKRLKP2KRVK3IAWVKAFIRKLRKGPLG	8
GRPRESGKKRKRKRLKP2PKKKRKV3IWIAQELRRIGDEFNAYYARR	8
AGYLLGKINLKALAALAKKIL2KRVK3CQIKIWFCNKRAKIKK	8
GLAFLGFLGAAGSTMGAWSQPKKKRKV	8
RQIKIWFQNRRMKWKK2KRVK3TRSSRAGLQWPVGRVHRLLRK	7
RQIKIWFQNRRMKWKK2KRVK3RRRRRRCRRRRRRC	7
RQIKIWFQNRRMKWKK2PKKKRKV3RRIRPRPPRLPRPRPRPLPFPRPG	7
RQIKIWFQNRRMKWKK2PKKKRKV3RRWWRRWRR	7
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3GRKAARAPGRRKQ	7

AGYLLGKINLKALAALAKKIL2KRVK3CSQIKIWFQNKRAKIKKC	7
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RLRWR	7
RQIKIWFQNRRMKWKK2PKKKRKV3LYKKGPAKKGRPPLRGWFH	7
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3GAYDLRRRERQSRLRRRERQSR	7
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	7
RQIKIWFQNRRMKWKK2PKKKRKV3RLRWR	7
GRPRESGKKRKRKRLKP2KRVK3RSVTIWFQSRRVKEKK	7
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3LYKKGPAKKGRPPLRGWFH	7
RQIKIWFQNRRMKWKK2KRVK3LYKKGPAKKGRPPLRGWFH	7
RQIKIWFQNRRMKWKK2PKKKRKV3RQIKIWFQNRRMKWKK	7
RQIKIWFQNRRMKWKK2PKKKRKV3KKYRGRKRHPR	6
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	6
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3PPRPPRPPRPPR	6
RQIKIWFQNRRMKWKK2KRVK3CRLRWRC	6
RQIKIWFQNRRMKWKK2KRVK3GRPRESGKKRKRKRLKP	6
AGYLLGKINLKALAALAKKIL2KRVK3SQIKIWFQCKRAKIKC	6
RQIKIWFQNRRMKWKK2KRVK3CRRRRRRCRRRRRR	6
RRIRPRPPRLPRPRPRPLPFPRPG	6
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RLRWR	6
RQIKIWFQNRRMKWKK2KRVK3RKKRRQRRR	6
RQIKIWFQNRRMKWKK2PKKKRKV3PPRPPRPPRPPR	6
RKKRRQRRR	6
GRPRESGKKRKRKRLKP2KRVK3VSALK	6
RQIKIWFQNRRMKWKK2KRVK3RQIKIWFQNRRMKWKK	6
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RXRRBRRXRRBR	6
RQIKIWFQNRRMKWKK2PKKKRKV3RGGRLSYSRRRFSTSTGR	6
RRRRRRRRRR	6
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3KGTYKKKLMRIPLKGT	6
RQIKIWFQNRRMKWKK2PKKKRKV3GRPRESGKKRKRKRLKP	6
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RRIRPRPPRLPRPRPRPLPFPRPG	6
AGYLLGKINLKALAALAKKIL2KRVK3LLIILRRRIRKQAHAHSK	5
RQIKIWFQNRRMKWKK2PKKKRKV3RRRRRRRRRRRR	5
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3GSPWGLQHHPPRT	5
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3PPRPPRPPR	5
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RRRRRRRRRRRR	5
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	5
RQIKIWFQNRRMKWKK2KRVK3RRRRRRRRR	5

GRPRESGKKRKRKRLKP	5
ALWKTLLKKVLKAPKKKRKV	5
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3KGTYKKKLMRIPLKGT	5
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3PLSSIFSRIGDP	5
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	5
RQIKIWFQNRRMKWKK2KRVK3KGTYKKKLMRIPLKGT	5
RQIKIWFQNRRMKWKK2PKKKRKV3LKTLTETLKELTKTLTEL	5
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3IAWVKAFIRKLRKGPLG	5
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3TRSSRAGLQWPVGRVHRLLRK	5
RQIKIWFQNRRMKWKK2PKKKRKV3CQIKIWFCNKRAKIKK	5
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3TRSSRAGLQWPVGRVHRLLRK	5
RRIPNRRPRR	5
RQIKIWFQNRRMKWKK2PKKKRKV3RQVTIWSQNRRVKSKK	5
RQIKIWFQNRRMKWKK2KRVK3KKYRGRKRHPR	5
GRPRESGKKRKRKRLKP2KRVK3GSPWGLQHHPPRT	5
TRRQRTRRARRNR	5
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3RSVTIWFQSRRVKEKK	5
RQIKIWFQNRRMKWKK2PKKKRKV3RQIKIWFQNRRMKWKK	5
RQIKIWFQNRRMKWKK2KRVK3RLRWR	5
HARIKPTFRRLKWKYKGKFW	5
RQIKIWFQNRRMKWKK2KRVK3RGGRLSYSRRRFSTSTGR	5
RQIKIWFQNRRMKWKK2PKKKRKV3PPRPPRPPR	5
GIGAVLKVLTTGLPALISWIKRKRQQ	5
RQIKIWFQNRRMKWKK2KRVK3PKKKRKVAGYLLGKINLKALAALAKKIL	4
RQIKIWFQNRRMKWKK2KRVK3GRKAARAPGRRKQ	4
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3GSPWGLQHHPPRT	4
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CRRRRRRCRRRRRR	4
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CQIKIWFCNKRAKIKK	4
LRRERQSRLRRERQSR	4
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3VSALK	4
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RKKRRQRRR	4
RQIKIWFQNRRMKWKK2PKKKRKV3RKKRRQRRR	4
RRRRRRRRR	4
RQIKIWFQNRRMKWKK	4
KRARNTEAARRSRARKLQRMKQ	4
RQIKIWFQNRRMKWKK2PKKKRKV3CRLRWRC	4

RHIKIWFQN RRM KWKK	4
LLIILRRRIRKQAHAHSK2PKKKRKV3LLIILRRRIRKQAHAHSKRXRRBRR XRRBR	4
RQIKIWFQNRRMKWKK2PKKKRKV3KRVKAGYLLGKINLKALAALAKKI L	4
RRRRRRRR	4
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3KLALKALKALKAALKLA	4
KMTRAQRRAAARRNRWTAR	4
RGGRLSYSRRRFSTSTGR	4
RQIKIWFQNRRMKWKK2KRVK3RQVTIWSQNRRVKSKK	4
RQIKIWFQNRRMKWKK2PKKKRKV3IWIAQELRRIGDEFNAYYARR	4
AGYLLGKINLKALAALAKKIL2KRVK3CRLRWRC	4
RQIKIWFQNRRMKWKK2PKKKRKV3RRRRRRRRR	4
KQINNWFINQRKRHWK	4
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RRRRRRRRR	4
RQIKIWFQNRRMKWKK2PKKKRKV3CSQIKIWFQNKRAKIKKC	4
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	4
KLWMRWYSPTTRRYG	4
RQIKIWFQNRRMKWKK2PKKKRKV3PKKKRKVAGYLLGKINLKALAALA KKIL	4
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CRLRWRC	4
RQIKIWFQNRRMKWKK2PKKKRKV3GAYDLRRRERQSRLRRRERQSR	4
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3PLSSIFSRIGDP	4
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CRLRWRC	4
RRWWRRWRR	4
RQIKIWFQNRRMKWKK2KRVK3CSQIKIWFQNKRAKIKKC	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CRRRRRRRRRRRC	3
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3FKIYDKKVRTRVVKH	3
RQIKIWFQNRRMKWKK2PKKKRKV3AGYLLGKINLKALAALAKKILKRV K	3
SQIKIWFQN KRAKIKK	3
RQIKIWFQNRRMKWKK2KRVK3GAYDLRRRERQSRLRRRERQSR	3
GAYDLRRRERQSRLRRRERQSR	3
TRRNKRNRIQEQLNRK	3
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3SQIKIWFQCKRAKIKC	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3GRPRESGKKRKRKRLKP	3
RQIKIWFQNRRMKWKK2PKKKRKV3VRLPPPVRLPPPVRLPPP	3
GKRKKKGKLGKKRDP	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RRWWRRWRR	3
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3VRLPPPVRLPPPVRLPPP	3

RQIKIWFQNRRMKWKK2KRVK3AGYLLGKINLKALAALAKKILKRVK	3
RQIKIWFQNRRMKWKK2KRVK3LLIILRRRIRKQAHAHSK	3
RQIKIWFQNRRMKWKK2PKKKRKV3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	3
RQVTIWFQNRRVKEKK	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3GRKAARAPGRRKQ	3
RQIKIWFQNRRMKWKK2PKKKRKV3KLALKALKALKAALKLA	3
GRPRESGKKRKRKRLKP2KRVK3VRLPPPVRLPPPVRLPPP	3
RQIKIWFQNRRMKWKK2PKKKRKV3SQIKIWFQCKRAKIKC	3
RQIKIWFQNRRMKWKK2KRVK3CQIKIWFCNKRAKIKK	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CRRRRRRCCRRRRRRC	3
RQIKIWFQNRRMKWKK2PKKKRKV3GSPWGLQHHPPRT	3
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3LLIILRRRIRKQAHAHSK	3
RLRWR	3
RQIKIWFQNRRMKWKK2KRVK3RQIKIWFQNRRMKWKK	3
RQIKIWFQNRRMKWKK2PKKKRKV3IAWVKAFIRKLRKGPLG	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3GSPWGLQHHPPRT	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RRRRRRCRRRRRRC	3
GRPRESGKKRKRKRLKP2KRVK3IWIAQELRRIGDEFNAYYARR	3
PPRPPRPPRPPRPPR	3
AGYLLGKINLKALAALAKKIL2KRVK3AGYLLGKINLKALAALAKKILKRVK	3
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3LKTLTETLKELTKTLTEL	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3PKKKRKVAGYLLGKINLKALAALAKKIL	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3ALWKTLLKKVLKAPKKKRKV	3
RQIKIWFQNRRMKWKK2PKKKRKV3CRRRRRRCRRRRRR	3
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3TRSSRAGLQWPVGRVHRLLRK	3
CAYHRLRRC	3
RQIKIWFQNRRMKWKK2KRVK3PPRPPRPPRPPR	3
LLIILRRRIRKQAHAHSK2KRVK3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	3
RQIKIWFQNRRMKWKK2KRVK3PPRPPRPPR	3
SRRARRSPRHLGSG	3
RQIKIWFQNRRMKWKK2PKKKRKV3PLSSIFSRIGDP	2
AGYLLGKINLKALAALAKKIL2PKKKRKV3KRVKAGYLLGKINLKALAALAKKIL	2
RQIKIWFQNRRMKWKK2KRVK3ALWKTLLKKVLKAPKKKRKV	2
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3CSQIKIWFQNKRAKIKKC	2

RQIKIWFQNRRMKWKK2PKKKRKV3VSALK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3AGYLLGKINLKALAALAKKIL KRVK	2
RQIKIWFQNRRMKWKK2KRVK3KRVKAGYLLGKINLKALAALAKKIL	2
RQIKIWFQNRRMKWKK2KRVK3KLALKALKALKAALKLA	2
PPRPPRPPRPPR	2
NAKTRRHERRRKLAIER	2
VKRGLKLRHVRPRVTRMDV	2
GRPRESGKKRKRKRLKP2SDGTLAVPFKA3IWIAQELRRIGDEFNAYYARR	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CQIKIWFCNKRAKIKK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3KKYRGRKRHPR	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3KLALKALKALKAALKL A	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3IWIAQELRRIGDEFNAY YARR	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3VSALK	2
LYKKGPAKKGRPPLRGWFH	2
RQIKIWFQNRRMKWKK2PKKKRKV3RSVTIWFQSRRVKEKK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RGGRLSYSRRRFSTSTGR	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3LYKKGPAKKGRPPLRGWFH	2
RQIKIWFQNRRMKWKK2PKKKRKV3ALWKTLLKKVLKAPKKKRKV	2
AGYLLGKINLKALAALAKKIL2KRVK3PPRPPRPPRPPR	2
AGYLLGKINLKALAALAKKIL2KRVK3IWIAQELRRIGDEFNAYYARR	2
RQIKIWFQNRRMKWKK2PKKKRKV3FKIYDKKVRTRVVKH	2
AGYLLGKINLKALAALAKKIL2KRVK3LLIILRRRIRKQAHAHSKRXRRBR RXRRBR	2
AGYLLGKINLKALAALAKKIL2KRVK3LYKKGPAKKGRPPLRGWFH	2
TAKTRYKARRAELIAERR	2
KGTYKKKLMRIPLKGT	2
AGYLLGKINLKALAALAKKIL2KRVK3RLRWR	2
RQIKIWFQNRRMKWKK2KRVK3LKTLTETLKELTKTLTEL	2
RQIKIWFQNRRMKWKK2KRVK3IAWVKAFIRKLRKGPLG	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CRRRRRRCRRRRRRC	2
AGYLLGKINLKALAALAKKIL2KRVK3RGGRLSYSRRRFSTSTGR	2
RQIKIWFQNRRMKWKK2KRVK3KLIKGRTPIKFGKADCDRPPKHSQNGM GK	2
AGYLLGKINLKALAALAKKIL2KRVK3TRSSRAGLQWPVGRVHRLLRK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RQVTIWSQNRRVKSKK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3LLIILRRRIRKQAHAHSK	2
AGYLLGKINLKALAALAKKIL2PKKKRKV3RRRRRRRRRRRR	2

RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3SQIKIWFQCKRAKIKC	2
PPRPPRPPR	2
RASKRDGSWVKKLHRILE	2
TRSSRAGLQWPVGRVHRLLRK	2
AGYLLGKINLKALAALAKKIL2KRVK3CQIKIWFCNKRAKIKK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RLRWR	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CRXRRBRRXRRBRC	2
AGYLLGKINLKALAALAKKIL2KRVK3PPRPPRPPR	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3GAYDLRRRERQSRLRRRERQSR	2
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	2
AGYLLGKINLKALAALAKKIL2KRVK3KRVKAGYLLGKINLKALAALAKKIL	2
FKIYDKKVRTRVVKH	2
VRLPPPVRLPPPVRLPPP	2
GPFHFYQFLFPPV	2
PLILLRLLRGQF	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3AGYLLGKINLKALAALAKKILKRVK	2
AGYLLGKINLKALAALAKKIL2KRVK3IAWVKAFIRKLRKGPLG	2
AGYLLGKINLKALAALAKKIL2KRVK3PKKKRKVAGYLLGKINLKALAALAKKIL	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3LYKKGPAKKGRPPLRGWFH	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RXRRBRCRXRRBRC	2
GRPRESGKKRKRKRLKP2KRVK3CRLRWRC	2
RQIKIWFQNRRMKWKK2KRVK3SQIKIWFQCKRAKIKC	2
LLIILRRRIRKQAHAHSK2KRVK3AGYLLGKINLKALAALAKKILKRVK	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3VRLPPPVRLPPPVRLPPP	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3PPRPPRPPRPPR	2
LLIILRRRIRKQAHAHSK2KRVK3KKYRGRKRHPR	2
AGYLLGKINLKALAALAKKIL2KRVK3KGTYKKKLMRIPLKGT	2
AGYLLGKINLKALAALAKKIL2PKKKRKV3AGYLLGKINLKALAALAKKILKRVK	2
AGYLLGKINLKALAALAKKIL2KRVK3KLALKALKALKAALKLA	2
AGYLLGKINLKALAALAKKIL2PKKKRKV3LLIILRRRIRKQAHAHSKRXRRBRRXRRBR	2

AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3FKIYDKKVRTRVVKH	2
AGYLLGKINLKALAALAKKIL2KRVK3GSPWGLQHHPPRT	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3KRVKAGYLLGKINLKALAALAKKIL	2
AGYLLGKINLKALAALAKKIL2PKKKRKV3PKKKRKVAGYLLGKINLKALAALAKKIL	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3PKKKRKVAGYLLGKINLKALAALAKKIL	2
RQIKIWFQNRRMKWKK2KRVK3GSPWGLQHHPPRT	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3IAWVKAFIRKLRKGPLG	2
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RXRRBRRXRRBR	2
AGYLLGKINLKALAALAKKIL2KRVK3GAYDLRRRERQSRLRRRERQSR	2
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3PPRPPRPPR	2
AGYLLGKINLKALAALAKKIL2KRVK3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	2
YTAIAWVKAFIRKLRK	2
LLIILRRRIRKQAHAHSK2KRVK3RXRRBRCRXRRBRC	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CSQIKIWFQNKRAKIKKC	1
AGYLLGKINLKALAALAKKIL2KRVK3SQIKIWFQCKRAKIKC	1
AGYLLGKINLKALAALAKKIL2KRVK3CSQIKIWFQNKRAKIKKC	1
RQIKIWFQNRRMKWKK2KRVK3IWIAQELRRIGDEFNAYYARR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3ALWKTLLKKVLKAPKKKRKV	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3CRLRWRC	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3SQIKIWFQCKRAKIKC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3LKTLTETLKELTKTLTEL	1
KETWWETWWTEWSQ PKKRKV	1
LIRLWSHLIHIWFQNRRLKWKKK	1
VDKGSYLPRPTPPRPIYNRN	1
MDAQTRRRERRAEKQAQWKAAN	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RGGRLSYSRRRFSTSTGR	1
AGYLLGKINLKALAALAKKIL2KRVK3LLIILRRRIRKQAHAHSK	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3RSVTIWFQSRRVKEKK	1
AGYLLGKINLKALAALAKKIL2KRVK3LKTLTETLKELTKTLTEL	1
LLIILRRRIRKQAHAHSK2KRVK3VSALK	1
AGYLLGKINLKALAALAKKIL2KRVK3RQVTIWSQNRRVKSKK	1
GSPWGLQHHPPRT	1
KLALKALKALKAALKLA	1

IPALK	1
VPALR	1
RQIKIWFQNRRMKWKK2KRVK3RSVTIWFQSRRVKEKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3LLIILRRRIRKQAHAHSK	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RRRRRRRRRRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RXRRBRRXRRBR	1
AGYLLGKINLKALAALAKKIL2KRVK3RQIKIWFQNRRMKWKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3VRLPPPVRLPPPVRLPPP	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3KLALKALKALKAALKLA	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RRWWRRWRR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RRIRPRPPRLPRPRPRPL PFPRPG	1
AGYLLGKINLKALAALAKKIL2KRVK3RRWWRRWRR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CQIKIWFCNKRAKIKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CRLRWRC	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3GAYDLRRRERQSRLRR RERQSR	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3FKIYDKKVRTRVVKH	1
LLIILRRRIRKQAHAHSK2KRVK3GRKAARAPGRRKQ	1
LLIILRRRIRKQAHAHSK	1
IAWVKAFIRKLRKGPLG	1
AAVLLPVLLAAPVQRKRQKLP	1
TSPLNIHNGQKL	1
VPTLK	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3IAWVKAFIRKLRKGPLG	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3LKTLTETLKELTKTLTE L	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3GRKAARAPGRRKQ	1
LLIILRRRIRKQAHAHSK2KRVK3RXRRBRRXRRBR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3SQIKIWFQCKRAKIKC	1
LLIILRRRIRKQAHAHSK2KRVK3LYKKGPAKKGRPPLRGWFH	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3LKTLTETLKELTKTLTEL	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3KKYRGRKRHPR	1
AGYLLGKINLKALAALAKKIL2KRVK3RRIRPRPPRLPRPRPRPLPFPRPG	1
LLIILRRRIRKQAHAHSK2KRVK3RXRRBRCRXRRBRC	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3IWIAQELRRIGDEFNAYYARR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RLRWR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3VSALK	1
LLIILRRRIRKQAHAHSK2KRVK3PKKKRKVAGYLLGKINLKALAALAKKI L	1

LLIILRRRIRKQAHAHSK2PKKKRKV3AGYLLGKINLKALAALAKKILKRVK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CSQIKIWFQNKRAKIKKC	1
LLIILRRRIRKQAHAHSK2KRVK3CRRRRRRCCRRRRRRC	1
AGYLLGKINLKALAALAKKIL2KRVK3RRRRRRRRR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3KKYRGRKRHPR	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3VSALK	1
AGYLLGKINLKALAALAKKIL2KRVK3VSALK	1
VSALK	1
AGYLLGKINLKALAALAKKIL2KRVK3ALWKTLLKKVLKAPKKKRKV	1
LLIILRRRIRKQAHAHSK2KRVK3CRXRRBRRXRRBRC	1
LLIILRRRIRKQAHAHSK2KRVK3PPRPPRPPR	1
LLIILRRRIRKQAHAHSK2KRVK3TRSSRAGLQWPVGRVHRLLRK	1
LLIILRRRIRKQAHAHSK2KRVK3GSPWGLQHHPPRT	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RQVTIWSQNRRVKSKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3LYKKGPAKKGRPPLRGWFH	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3PLSSIFSRIGDP	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RRRRRRRRRRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CRXRRBRRXRRBRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RRIRPRPPRLPRPRPRPLPFPRPG	1
LLIILRRRIRKQAHAHSK2KRVK3KGTYKKKLMRIPLKGT	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CQIKIWFCNKRAKIKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RRIRPRPPRLPRPRPRPLPFPRPG	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RXRRBRRXRRBR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3TRSSRAGLQWPVGRVHRLLRK	1
RQIKIWFQNRRMKWKK2KRVK3GRKAARAPGRRKQ	1
LLIILRRRIRKQAHAHSK2KRVK3CRRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2KRVK3KLALKALKALKAALKLA	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	1
LLIILRRRIRKQAHAHSK2KRVK3RXRRBRRXRRBR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3PPRPPRPPRPPR	1
AGYLLGKINLKALAALAKKIL2KRVK3RKKRRQRRR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3PPRPPRPPRPPR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3PPRPPRPPR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3PKKKRKVAGYLLGKINLKALAALAKKIL	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2KRVK3RSVTIWFQSRRVKEKK	1

LLIILRRRIRKQAHAHSK2KRVK3CRXRRBRRXRRBRC	1
LLIILRRRIRKQAHAHSK2KRVK3KRVKAGYLLGKINLKALAALAKKIL	1
LLIILRRRIRKQAHAHSK2KRVK3LLIILRRRIRKQAHAHSK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3TRSSRAGLQWPVGRVHRLLRK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RXRRBRCRXRRBRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3KLALKALKALKAALKLA	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3KRVKAGYLLGKINLKALAALAKKIL	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CSQIKIWFQNKRAKIKKC	1
LLIILRRRIRKQAHAHSK2KRVK3PPRPPRPPRPPR	1
LLIILRRRIRKQAHAHSK2KRVK3RRIRPRPPRLPRPRPRPLPFPRPG	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3GRPRESGKKRKRKRLKP	1
LLIILRRRIRKQAHAHSK2KRVK3CRRRRRRCCRRRRRRC	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3PPRPPRPPRPPR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3GRKAARAPGRRKQ	1
LLIILRRRIRKQAHAHSK2KRVK3SQIKIWFQCKRAKIKC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3FKIYDKKVRTRVVKH	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RRRRRRRRRRRR	1
LLIILRRRIRKQAHAHSK2KRVK3CRLRWRC	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RRRRRRRRR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3GRKAARAPGRRKQ	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3PKKKRKVAGYLLGKINLKALAALAKKIL	1
LLIILRRRIRKQAHAHSK2PKKKRKV3VSALK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3SQIKIWFQCKRAKIKC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RSVTIWFQSRRVKEKK	1
GRPRESGKKRKRKRLKP2KRVK3PLSSIFSRIGDP	1
LLIILRRRIRKQAHAHSK2KRVK3CRRRRRRRRRRRRC	1
AGYLLGKINLKALAALAKKIL2KRVK3RQIKIWFQNRRMKWKK	1
RQIKIWFQNRRMKWKK2KRVK3KKYRGRKRHPR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3IAWVKAFIRKLRKGPLG	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3IWIAQELRRIGDEFNAYYARR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RKKRRQRRR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RGGRLSYSRRRFSTSTGR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RGGRLSYSRRRFSTSTGR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	1
LLIILRRRIRKQAHAHSK2KRVK3RRRRRRCRRRRRRC	1

LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3GRPRESGKKRKRKRLKP	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3KGTYKKKLMRIPLKGT	1
LLIILRRRIRKQAHAHSK2KRVK3RGGRLSYSRRRFSTSTGR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RXRRBRCRXRRBRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3GSPWGLQHHPPRT	1
LLIILRRRIRKQAHAHSK2PKKKRKV3LLIILRRRIRKQAHAHSK	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3KGTYKKKLMRIPLKGT	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3LLIILRRRIRKQAHAHSK RXRRBRRXRRBR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3GRKAARAPGRRKQ	1
RQIKIWFQNRRMKWKK2SDGTLAVPFKA3PLSSIFSRIGDP	1
LLIILRRRIRKQAHAHSK2PKKKRKV3KRVKAGYLLGKINLKALAALAKKI L	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3PPRPPRPPR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RKKRRQRRR	1
RQIKIWFQNRRMKWKK2KRVK3CRRRRRRRRRRRRC	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RRRRRRRRR	1
LLIILRRRIRKQAHAHSK2KRVK3RRRRRRRRR	1
LLIILRRRIRKQAHAHSK2KRVK3LKTLTETLKELTKTLTEL	1
RQIKIWFQNRRMKWKK2KRVK3PLSSIFSRIGDP	1
RQIKIWFQNRRMKWKK2KRVK3GRPRESGKKRKRKRLKP	1
LLIILRRRIRKQAHAHSK2KRVK3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RLRWR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3VRLPPPVRLPPPVRLPPP	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RSVTIWFQSRRVKEKK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RSVTIWFQSRRVKEKK	1
RQIKIWFQNRRMKWKK2KRVK3RRRRRRRRRRRR	1
DPKGDPKGVTVTVTVTVTGKGDPKPD	1
AGYLLGKINLKALAALAKKIL	1
VELPPPVELPPPVELPPP	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2KRVK3RLRWR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RSVTIWFQSRRVKEKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3GAYDLRRRERQSRLRRRERQ SR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3KRVKAGYLLGKINLKALAAL AKKIL	1
LLIILRRRIRKQAHAHSK2PKKKRKV3KKYRGRKRHPR	1
LLIILRRRIRKQAHAHSK2KRVK3GRPRESGKKRKRKRLKP	1

LLIILRRRIRKQAHAHSK2PKKKRKV3VRLPPPVRLPPPVRLPPP	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3AGYLLGKINLKALAALAKKILKRVK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3LLIILRRRIRKQAHAHSK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RRRRRRRRR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RRRRRRRRR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RQVTIWSQNRRVKSKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3GAYDLRRRERQSRLRRRERQSR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RQVTIWSQNRRVKSKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3KGTYKKKLMRIPLKGT	1
LLIILRRRIRKQAHAHSK2KRVK3PLSSIFSRIGDP	1
LLIILRRRIRKQAHAHSK2KRVK3RKKRRQRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CRRRRRRCCRRRRRRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3FKIYDKKVRTRVVKH	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RQIKIWFQNRRMKWKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RRIRPRPPRLPRPRPRPLPFPRPG	1
LLIILRRRIRKQAHAHSK2KRVK3IWIAQELRRIGDEFNAYYARR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3TRSSRAGLQWPVGRVHRLLRK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CRXRRBRRXRRBRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3PPRPPRPPR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CRRRRRRCCRRRRRRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3LYKKGPAKKGRPPLRGWFH	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3SQIKIWFQCKRAKIKC	1
RQIKIWFQNRRMKWKK2KRVK3CRRRRRRCRRRRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3KKYRGRKRHPR	1
LLIILRRRIRKQAHAHSK2KRVK3RQIKIWFQNRRMKWKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RXRRBRRXRRBR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3IWIAQELRRIGDEFNAYYARR	1
LLIILRRRIRKQAHAHSK2KRVK3CRRRRRRCRRRRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CRLRWRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CRXRRBRRXRRBRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3GRPRESGKKRKRKRLKP	1
LLIILRRRIRKQAHAHSK2KRVK3RRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3LYKKGPAKKGRPPLRGWFH	1
LLIILRRRIRKQAHAHSK2KRVK3GAYDLRRRERQSRLRRRERQSR	1

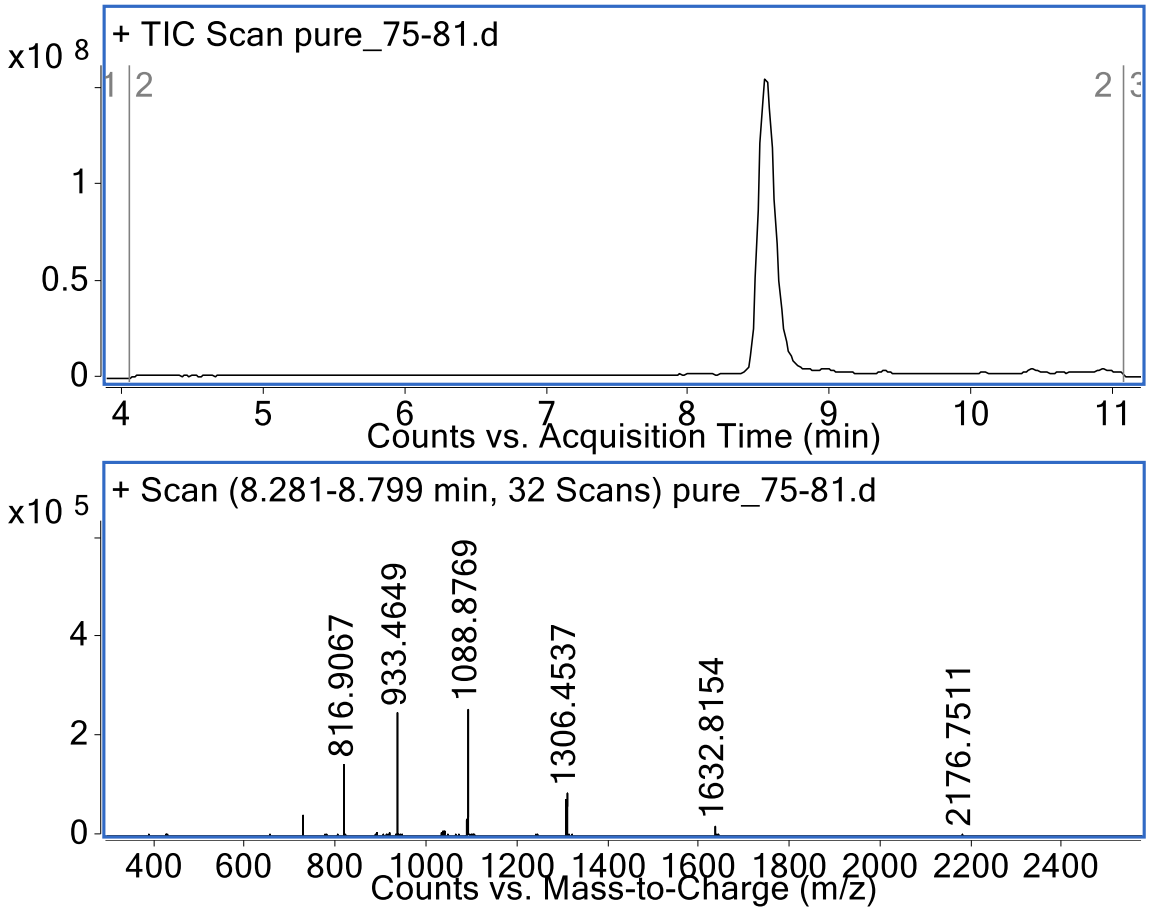
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3GSPWGLQHHPPRT	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CRRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CRRRRRRRRRRRRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3KGTYKKKLMRIPLKGT	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CRRRRRRCCRRRRRRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RQIKIWFQNRRMKWKK	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CRRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RKKRRQRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3IWIAQELRRIGDEFNAYYARR	1
AGYLLGKINLKALAALAKKIL2KRVK3RSVTIWFQSRRVKEKK	1
LLIILRRRIRKQAHAHSK2KRVK3IAWVKAFIRKLRKGPLG	1
LLIILRRRIRKQAHAHSK2KRVK3CSQIKIWFQNKRAKIKKC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RGGRLSYSRRRFSTSTGR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3GSPWGLQHHPPRT	1
LLIILRRRIRKQAHAHSK2KRVK3CQIKIWFCNKRAKIKK	1
LLIILRRRIRKQAHAHSK2PKKKRKV3LKTLTETLKELTKTLTEL	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3PPRPPRPPRPPR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RQVTIWSQNRRVKSKK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RLRWR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3KLALKALKALKAALKLA	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3ALWKTLLKKVLKAPKKKRKV	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3VSALK	1
RQIKIWFQNRRMKWKK2KRVK3VSALK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3GAYDLRRRERQSRLRRRERQSR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CQIKIWFCNKRAKIKK	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RKKRRQRRR	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RRWWRRWRR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3PPRPPRPPR	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3LLIILRRRIRKQAHAHSK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3ALWKTLLKKVLKAPKKKRKV	1
LLIILRRRIRKQAHAHSK2PKKKRKV3PLSSIFSRIGDP	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3LKTLTETLKELTKTLTEL	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3IAWVKAFIRKLRKGPLG	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CQIKIWFCNKRAKIKK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CRRRRRRRRRRRRC	1
GLWRALWRLLRSLWRLLWRA	1
PLSSIFSRIGDP	1

LLIILRRRIRKQAHAHSK2KRVK3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3RRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3KLIKGRTPIKFGKADCDRPPKHSQNGMGK	1
LLIILRRRIRKQAHAHSK2KRVK3RRWWRRWRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CRRRRRRCRRRRRR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3IAWVKAFIRKLRKGPLG	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CRRRRRRRRRRRRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CRRRRRRRRRRRRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CRLRWRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CSQIKIWFQNKRAKIKKC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CRRRRRRCRRRRRRC	1
LLIILRRRIRKQAHAHSK2PKKKRKV3FKIYDKKVRTRVVKH	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3CRRRRRRCCRRRRRRC	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CSQIKIWFQNKRAKIKKC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RXRRBRCRXRRBRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3GRPRESGKKRKRKRLKP	1
LLIILRRRIRKQAHAHSK2KRVK3ALWKTLLKKVLKAPKKKRKV	1
RQIKIWFQNRRMKWKK2KRVK3VRLPPPVRLPPPVRLPPP	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3VRLPPPVRLPPPVRLPPP	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RQIKIWFQNRRMKWKK	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RRWWRRWRR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3RRWWRRWRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3CRRRRRRCRRRRRRC	1
AGYLLGKINLKALAALAKKIL2PKKKRKV3RRRRRRCRRRRRRC	1
AGYLLGKINLKALAALAKKIL2SDGTLAVPFKA3CRRRRRRCRRRRRR	1
LLIILRRRIRKQAHAHSK2PKKKRKV3CRRRRRRCRRRRRR	1
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3PLSSIFSRIGDP	1
LLIILRRRIRKQAHAHSK2KRVK3RRRRRRRRRRRR	0
AGYLLGKINLKALAALAKKIL2KRVK3PLSSIFSRIGDP	0
LLIILRRRIRKQAHAHSK2KRVK3CRRRRRRCRRRRRRC	0
LLIILRRRIRKQAHAHSK2PKKKRKV3ALWKTLLKKVLKAPKKKRKV	0
AGYLLGKINLKALAALAKKIL2PKKKRKV3CRRRRRRCRRRRRR	0
RQIKIWFQNRRMKWKK2KRVK3FKIYDKKVRTRVVKH	0
RQIKIWFQNRRMKWKK2KRVK3CRXRRBRRXRRBRC	0
LLIILRRRIRKQAHAHSK2SDGTLAVPFKA3RRRRRRCRRRRRRC	0
AGYLLGKINLKALAALAKKIL2KRVK3VRLPPPVRLPPPVRLPPP	0
LLIILRRRIRKQAHAHSK2KRVK3VRLPPPVRLPPPVRLPPP	0

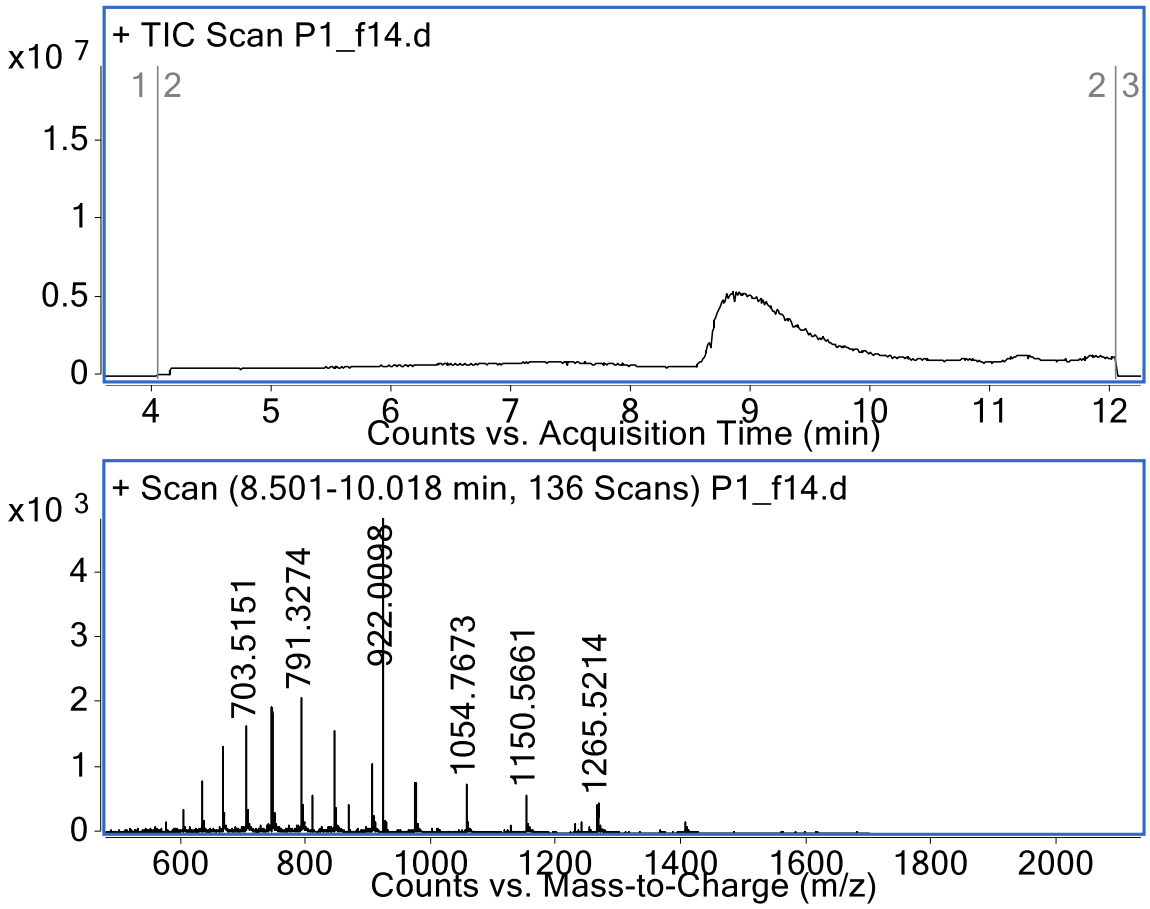
AGYLLGKINLKALAALAKKIL2KRVK3FKIYDKKVRTRVVKH	0
LLIILRRRIRKQAHAHSK2KRVK3RQVTIWSQNRRVKSKK	0
LLIILRRRIRKQAHAHSK2KRVK3FKIYDKKVRTRVVKH	0

2.8 Appendix II: LC-MS characterization

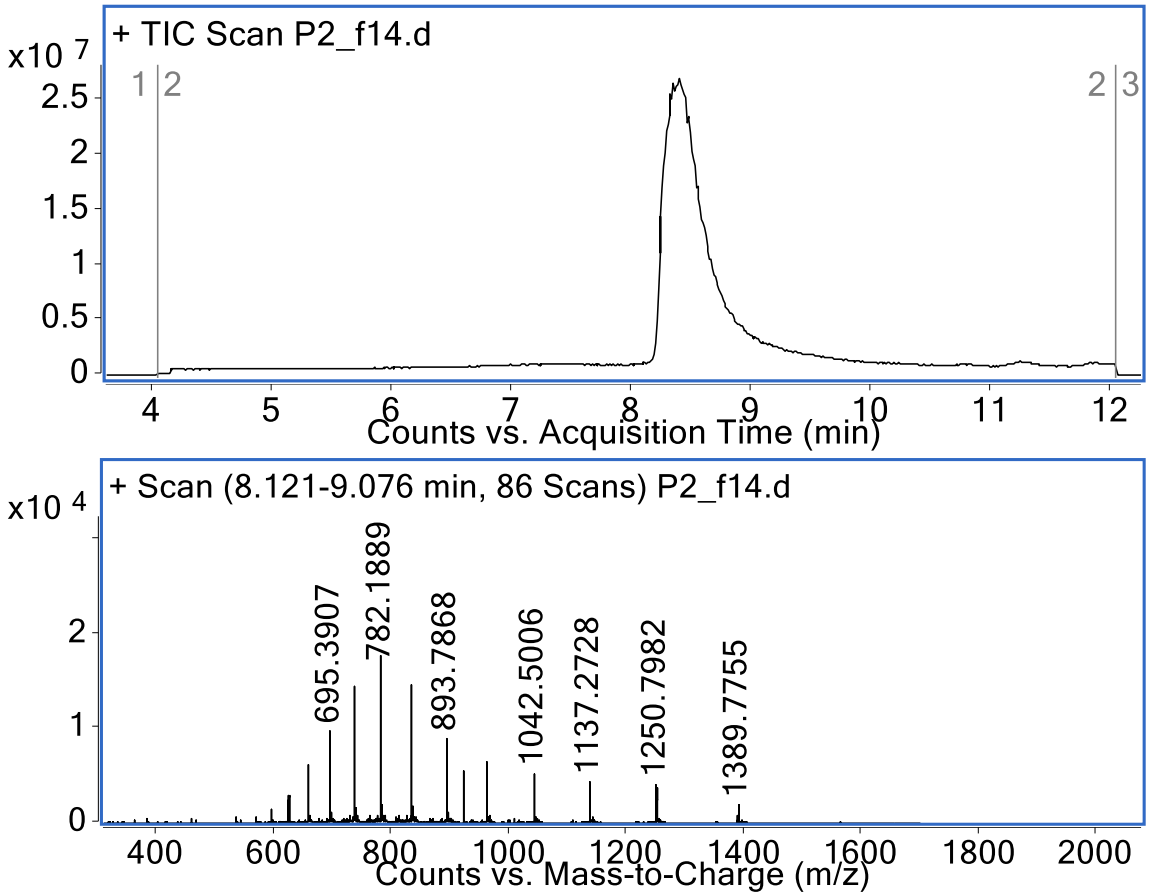
PMO-DBCO (Method A)
Mass Expected: 6527.9 Da
Mass Observed: 6527.9 Da
PMO sequence: GCT ATT ACC TTA ACC CAG



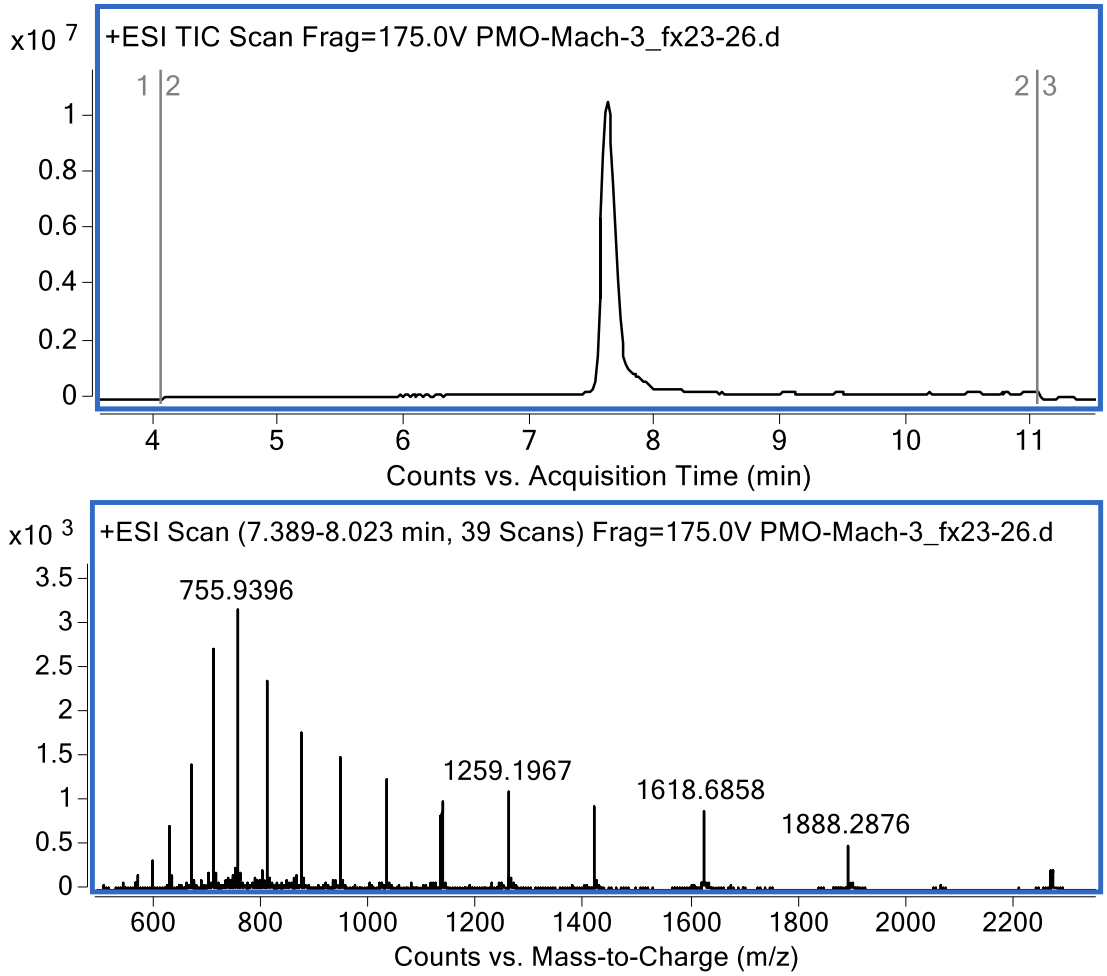
PMO-Mach1 (Method B)
Mass Expected: 12645.4 Da
Mass Observed: 12645.6 Da
Peptide sequence:
ALKBRSAAKAVRWPKKKIKQASKKVAKYALXXXRKKKAASKXWLQLHWPRW



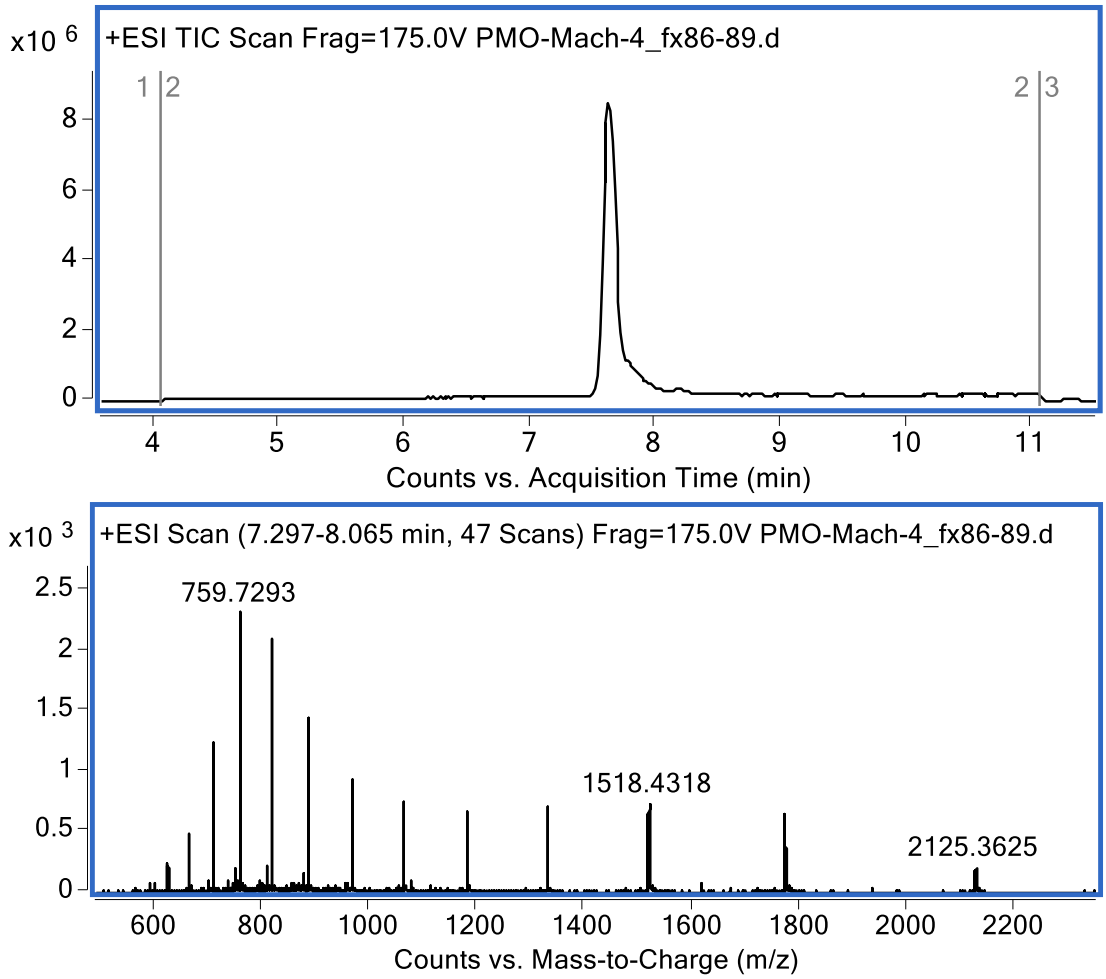
PMO-Mach2 (Method B)
Mass Expected: 12499.1 Da
Mass Observed: 12499.2 Da
Peptide sequence:
PPLRNAKKKNLKNNLKMDPKFTKKVKQGALKLNRRKKNRGPKGPXKHWTT



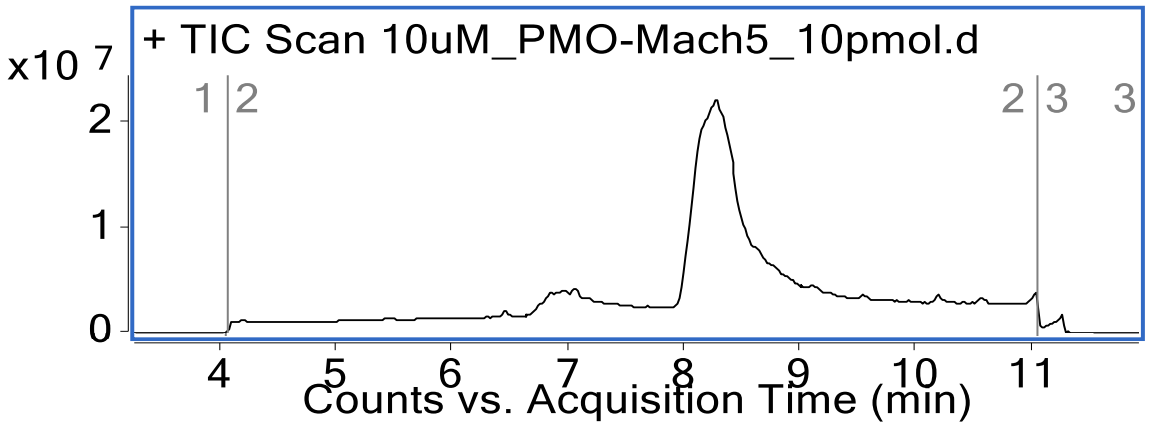
PMO-Mach3 (Method A)
Mass Expected: 11323.6 Da
Mass Observed: 11324.3 Da
Peptide sequence: QKKRKSKANKKNWPKGKLSIHAKDYKQGPKAKXRKQRXR

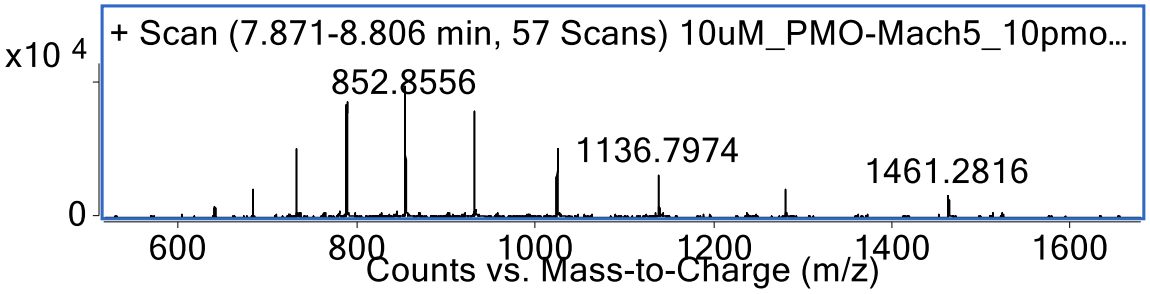


PMO-Mach4 (Method A)
Mass Expected: 10622.0 Da
Mass Observed: 10622.5 Da
Peptide sequence: KKGKKQNKKKHRWPKKKVPQPKKMFKQGABXRX

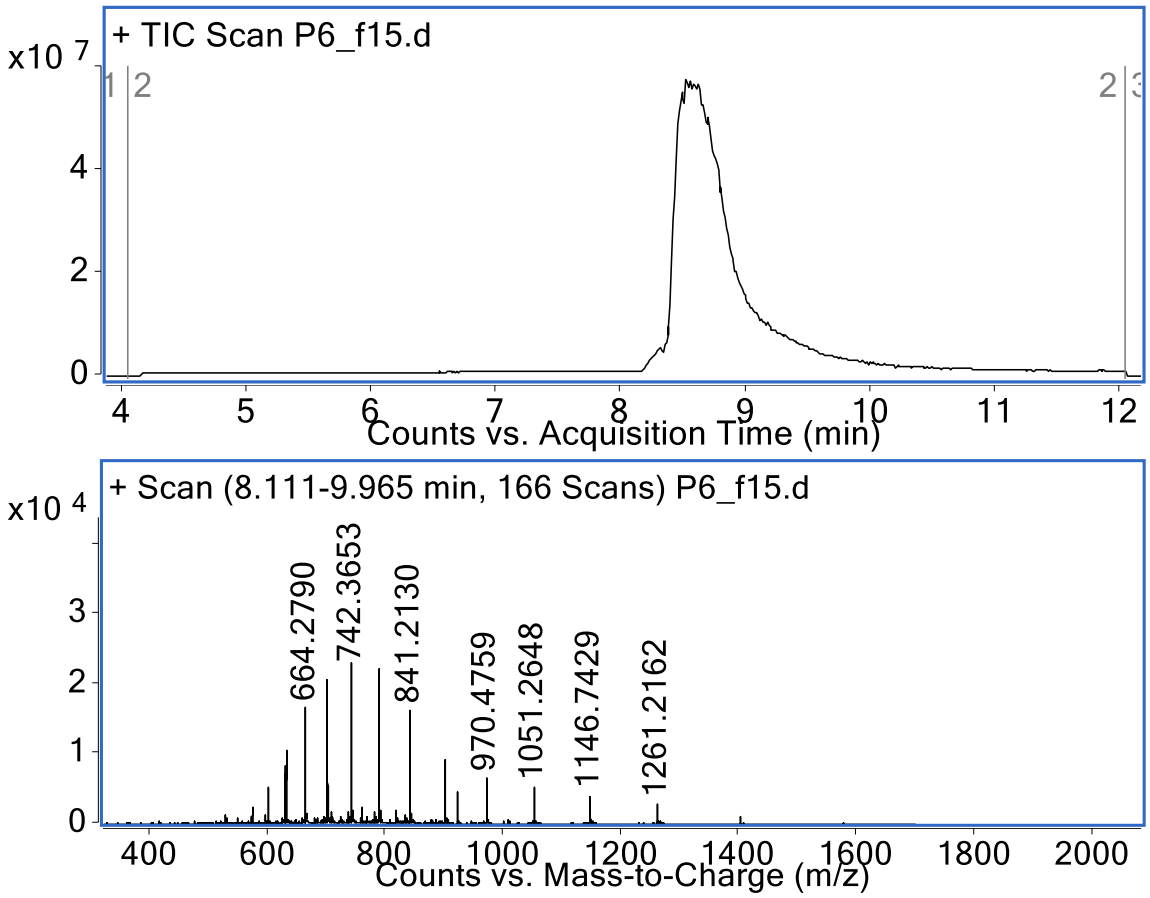


PMO-Mach5 (Method A)
Mass Expected: 10222.5 Da
Mass Observed: 10222.5 Da
Peptide sequence: AKKKIAKAKKHRGPNBGIHAPVSKIKDPLKXXX

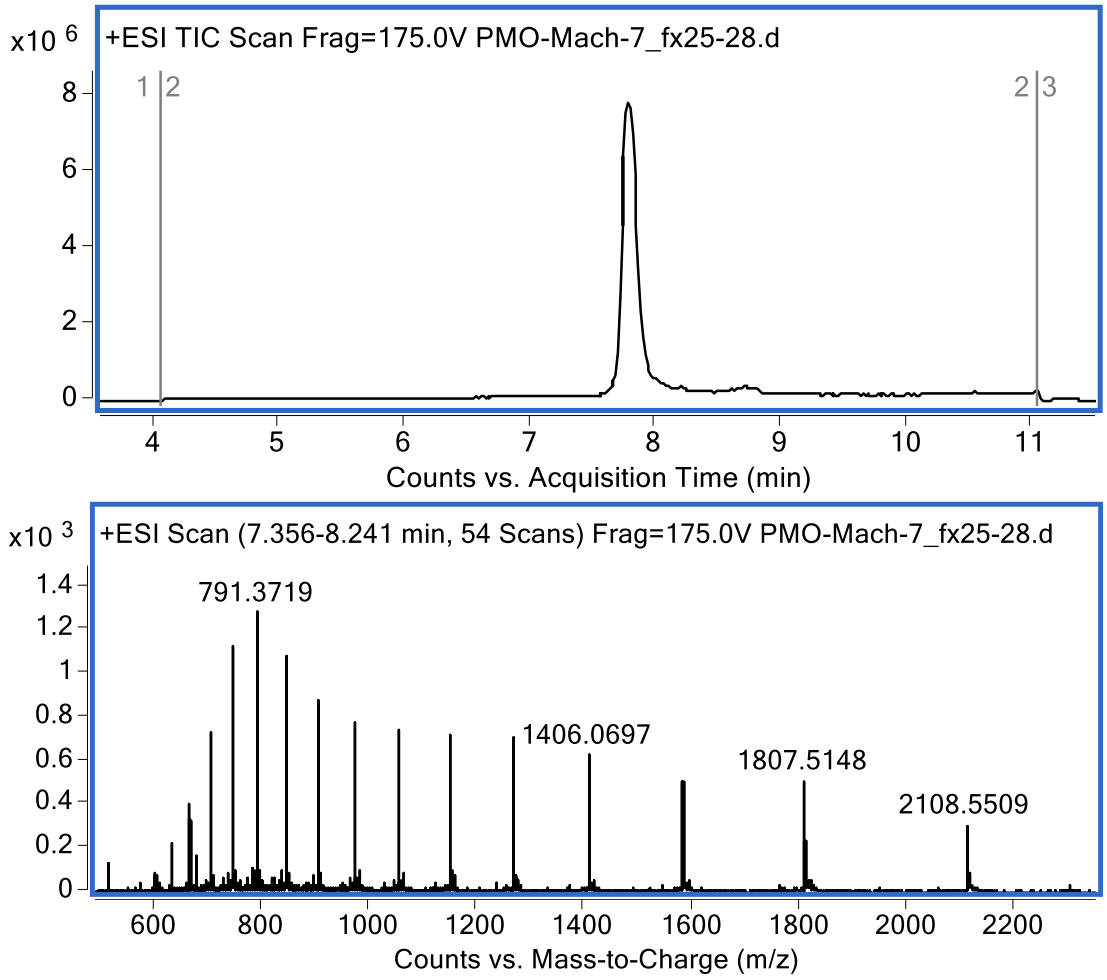




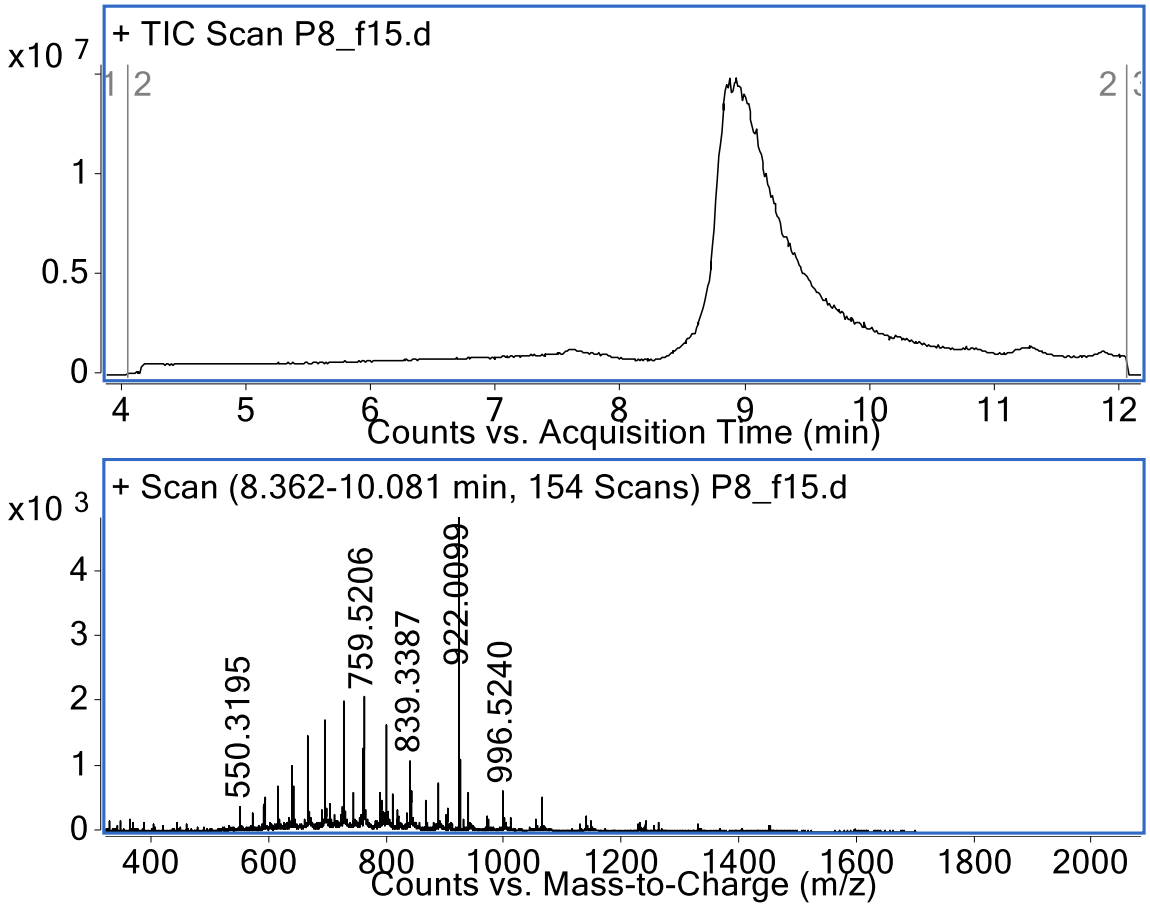
PMO-Mach6 (Method B)
Mass Expected: 12603.4 g/mol
Mass Observed: 12603.4 g/mol
Peptide sequence:
ALKBRSAAKAVRWPKKAIKQASKKVAKYALKXXRKKKAASKXWLQLHWPRW



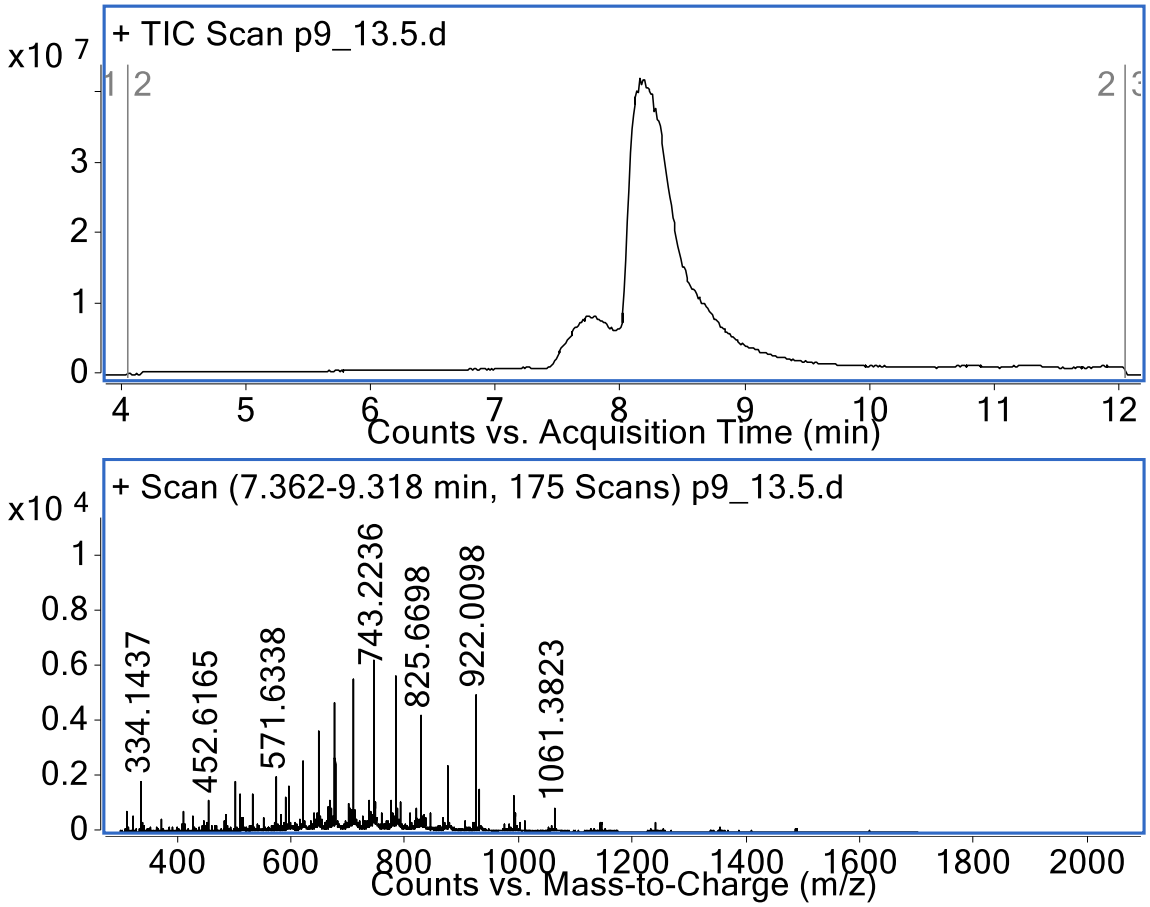
PMO-Mach7 (Method A)
Mass Expected: 12645.4 Da
Mass Observed: 12645.9 Da
Peptide sequence:
XKHPXAVQBAARAWKVPAAALWKKKRLKKSSKQKKKWLWKARSAXKYXRLI



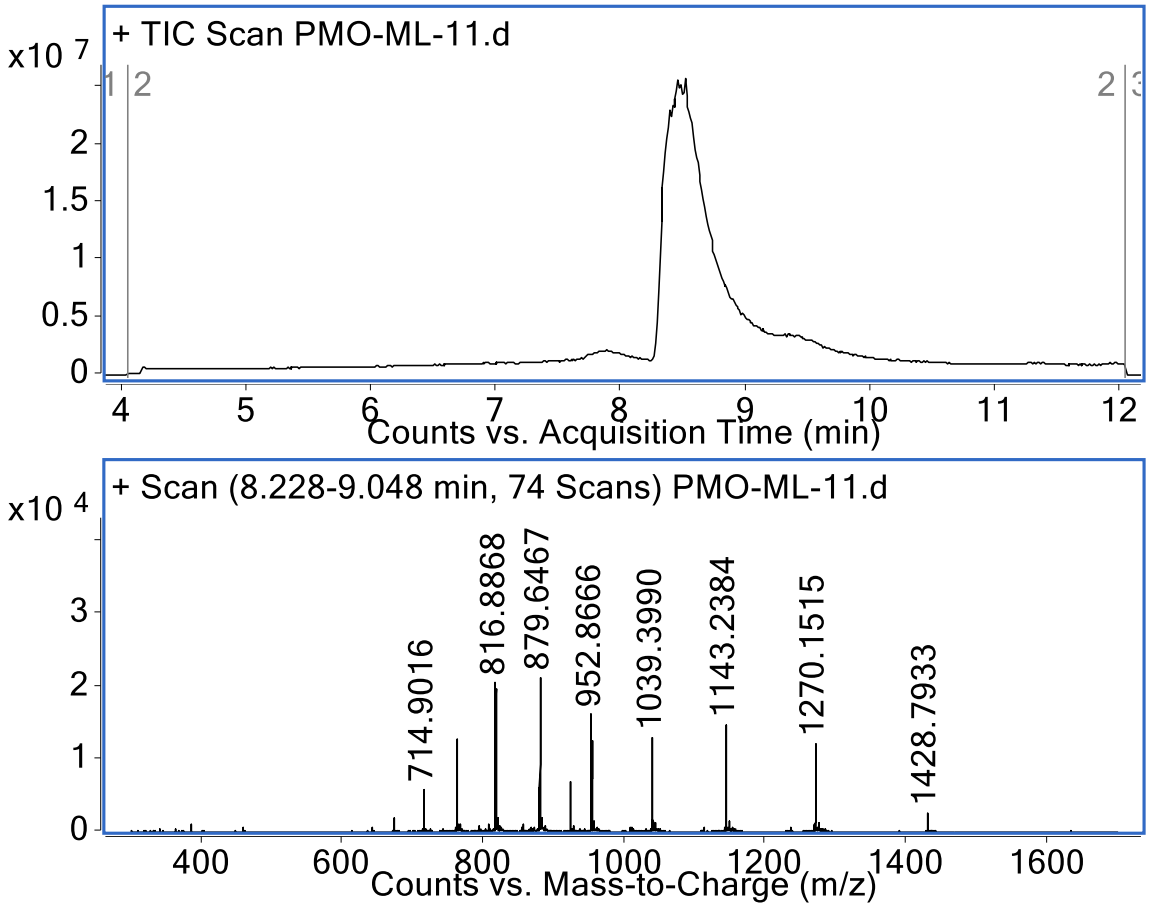
PMO-Mach8 (Method B)
Mass Expected: 15929.1 Da
Mass Observed: 15929.3 Da
Peptide sequence:
BKGKNLLAKIRRGPNGGNBQGSQGYLLYLLXRXRRQRXXYPWWRXKHXRWXXRXRG
HXRRRRQXLKPDRXRGGKGSVS



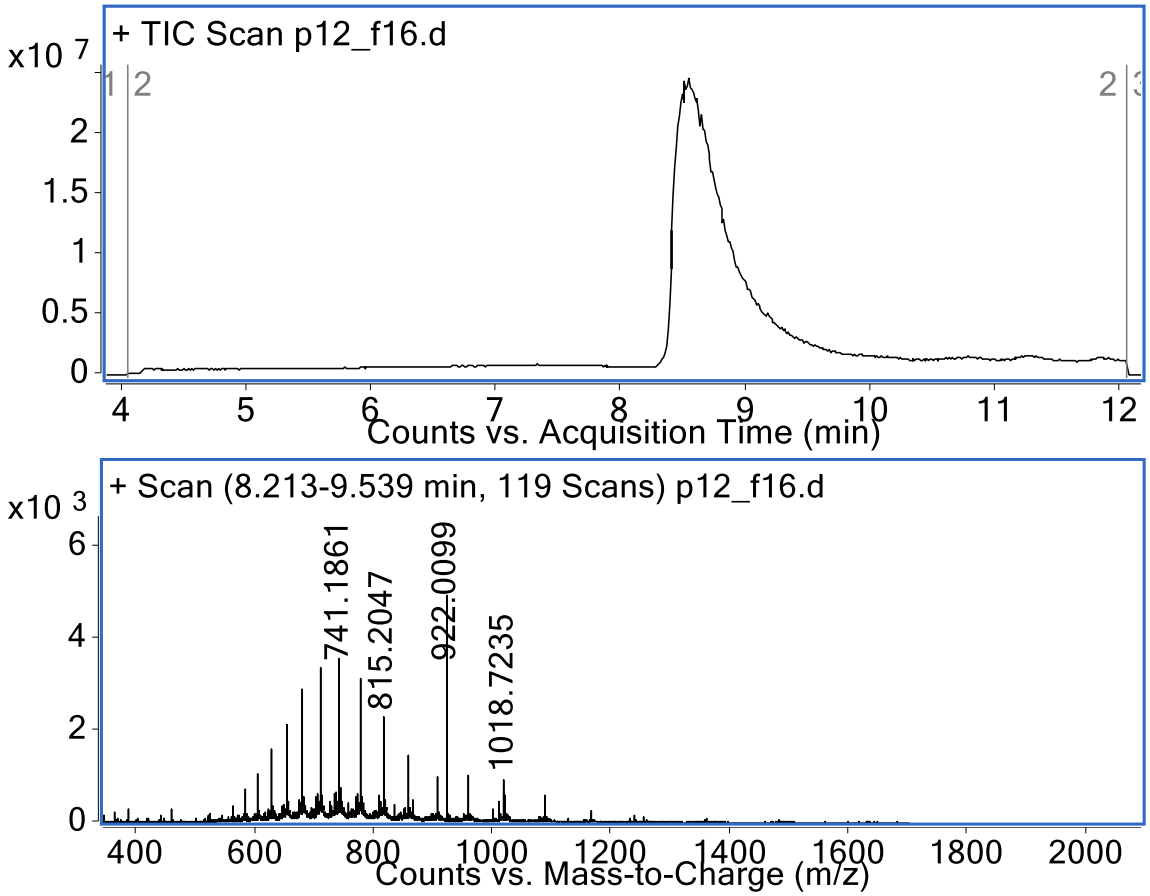
PMO-Mach9 (Method B)
Mass Expected: 14844.8 Da
Mass Observed: 14845.0 Da
Peptide sequence:
KKKKNLNBKSRRGPNGGALQPSQGYLQPLNXRXRRQRXXYPWWRXKHXRWRXRYH
XRRRRQXLKPG



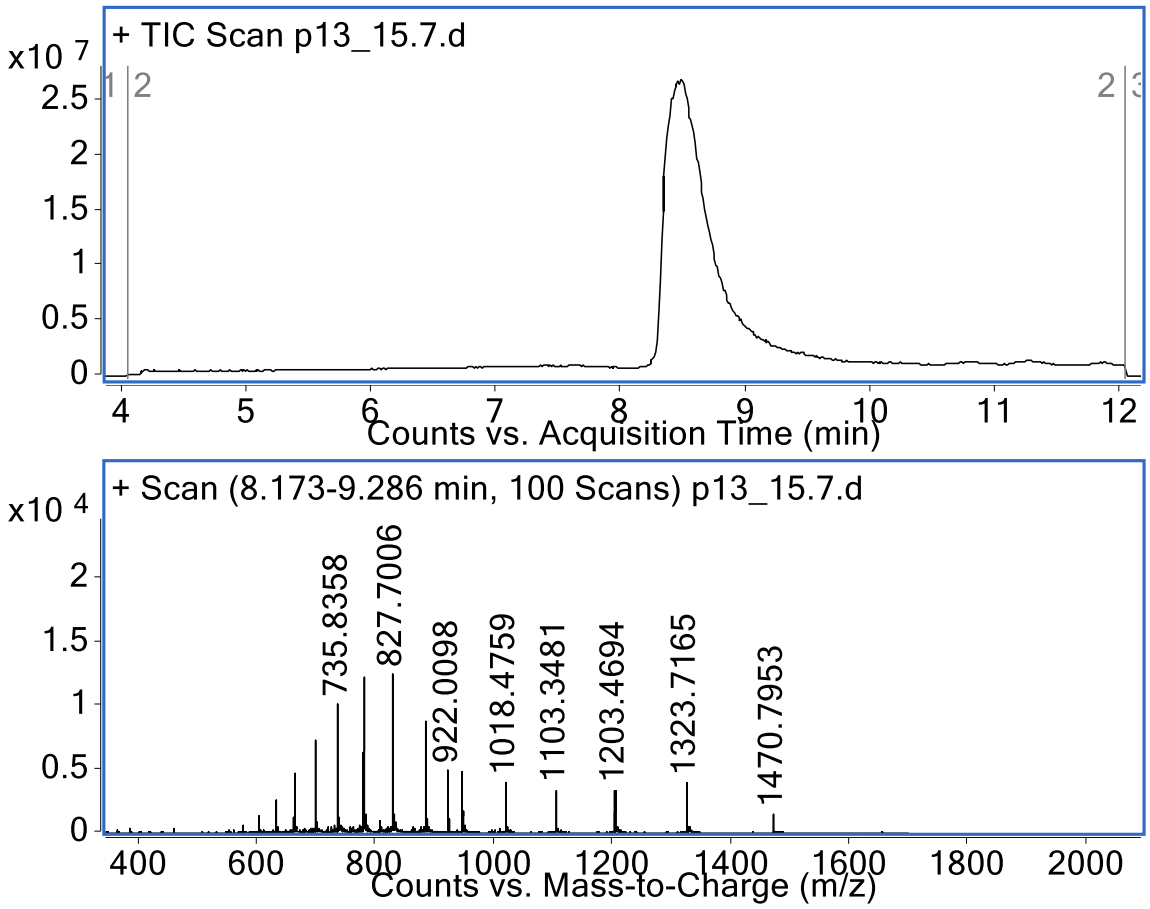
PMO-Mach11 (Method B)
Mass Expected: 11422.8 Da
Mass Observed: 11422.8 Da
Peptide sequence: TSNLKLHLAPPVKKKALKKPLYKAKKKKKVVSPTWXTDQEW



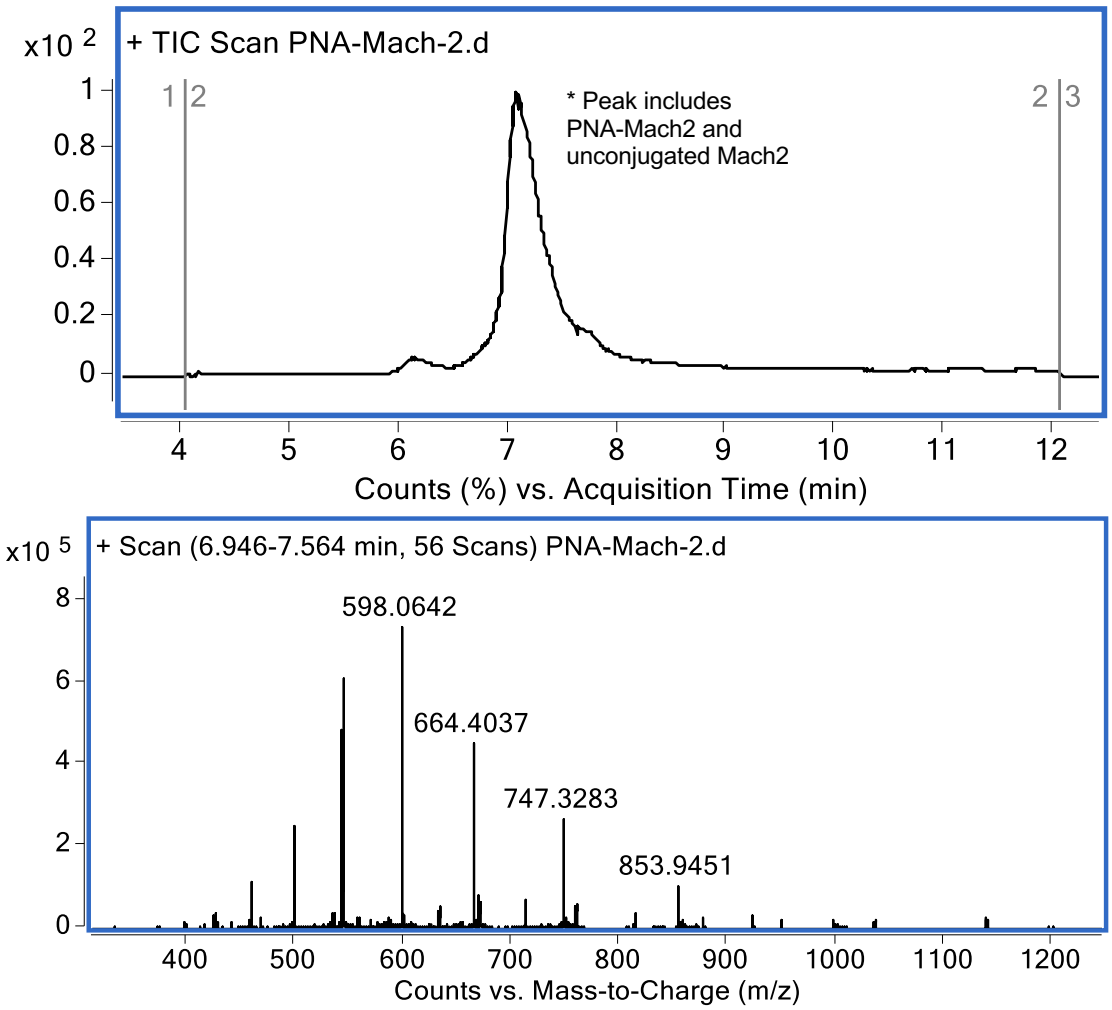
PMO-Mach12 (Method B)
Mass Expected: 16284.5 Da
Mass Observed: 16284.7 Da
Peptide sequence:
KGGKNLAKKIRRGPNGGALQPSQGYLLYLBXRXRRQRXXGPXWRXKHXRWXXXXXR
PTHXRRRRQXL**C**PGRXRP**C**RGSVS



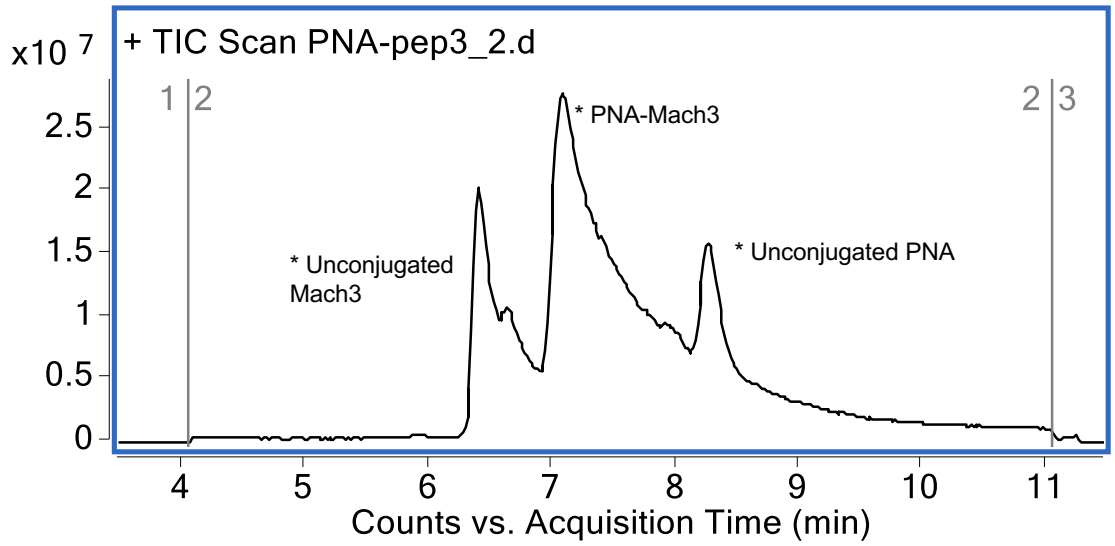
PMO-Mach13 (Method B)
Mass Expected: 13227.8 Da
Mass Observed: 13228.0 Da
Peptide sequence:
AKKKKLGBKALRWPNGK**C**PQPKEK**C**PKYLLGRXRRKRXRYPWWRXKHRRW

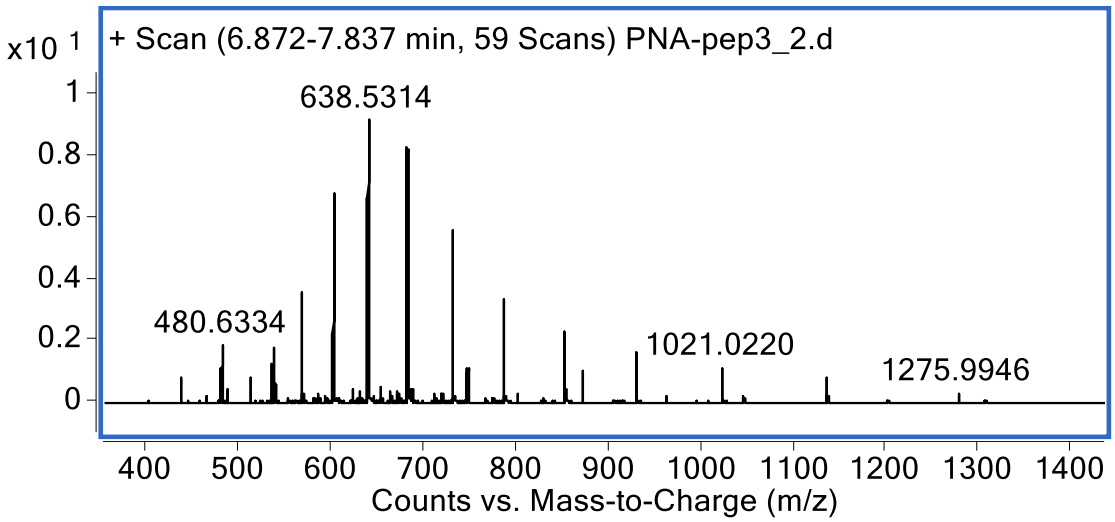


PNA-Mach2 (Method B)
Mass Expected: 11375.5 Da
Mass Observed: 11374.9 Da

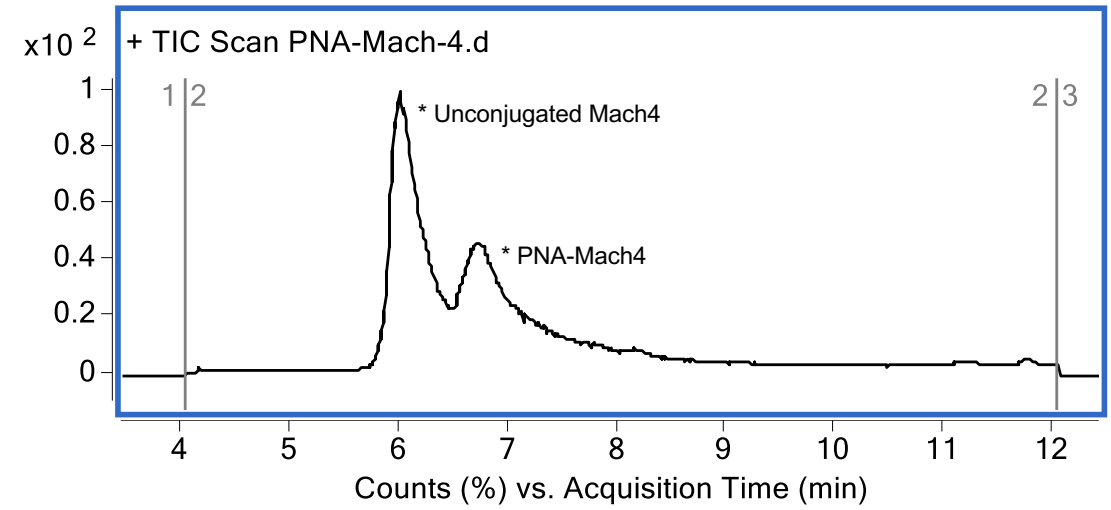


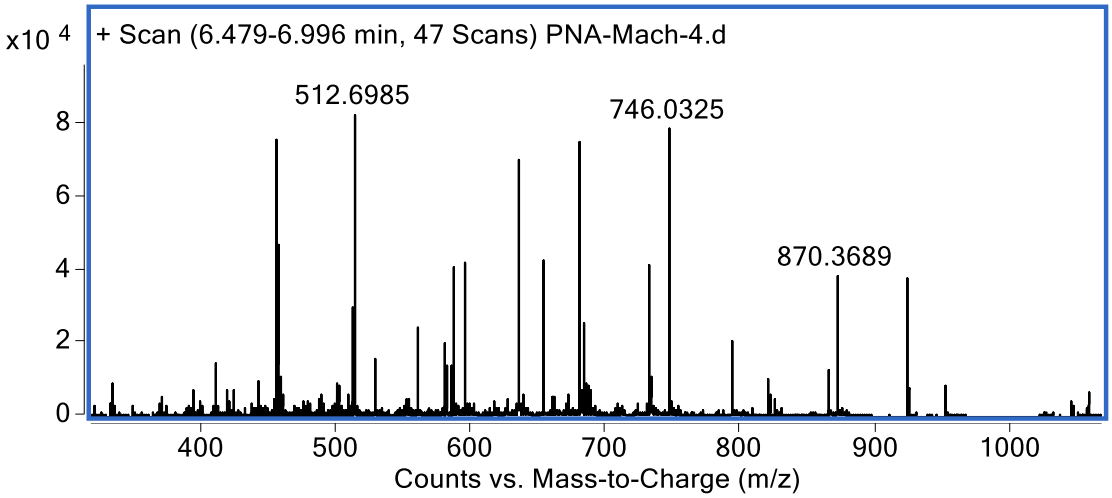
PNA-Mach3 (Method A)
Mass Expected: 10200.0 Da
Mass Observed: 10200.6 Da



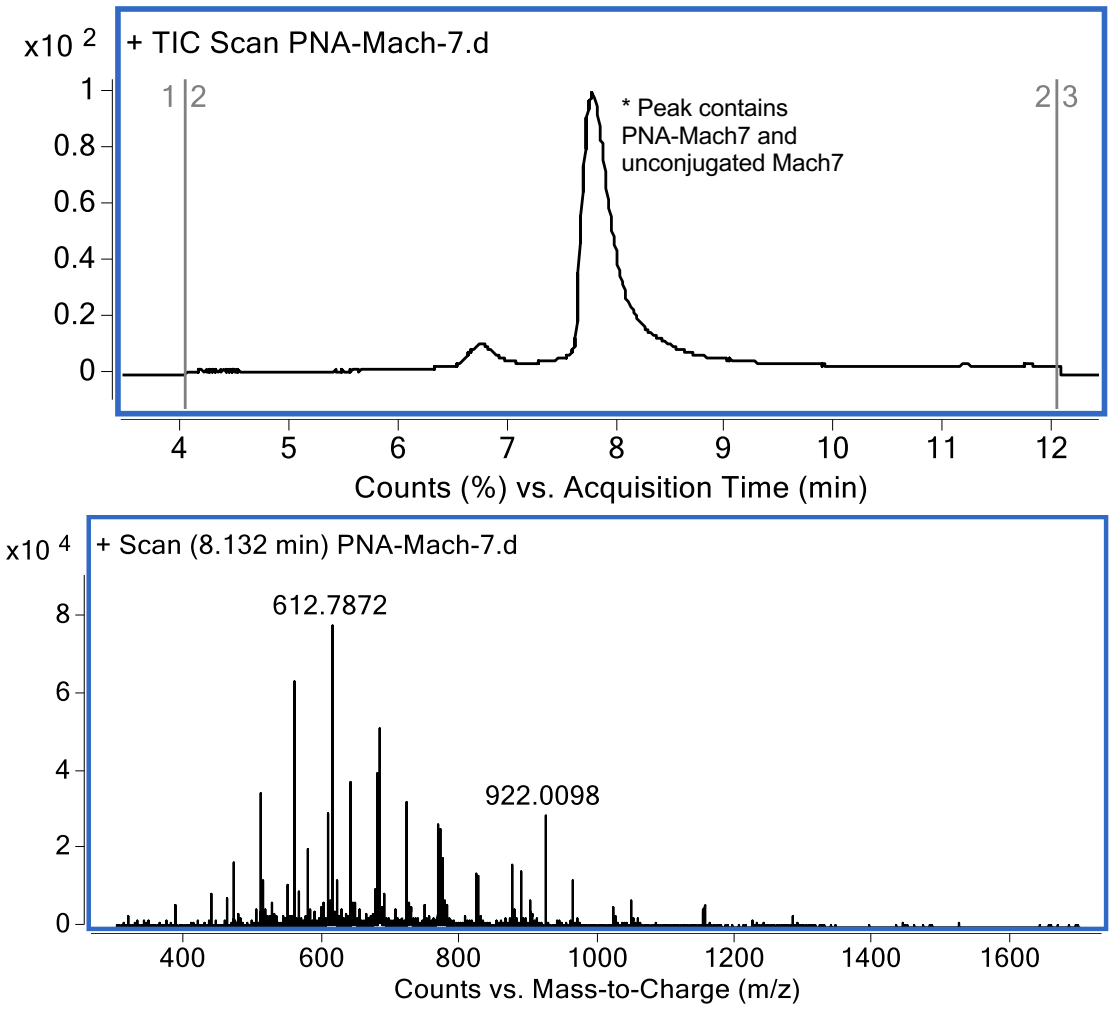


PNA-Mach4 (Method B)
Mass Expected: 9498.4 Da
Mass Observed: 9497.8 Da

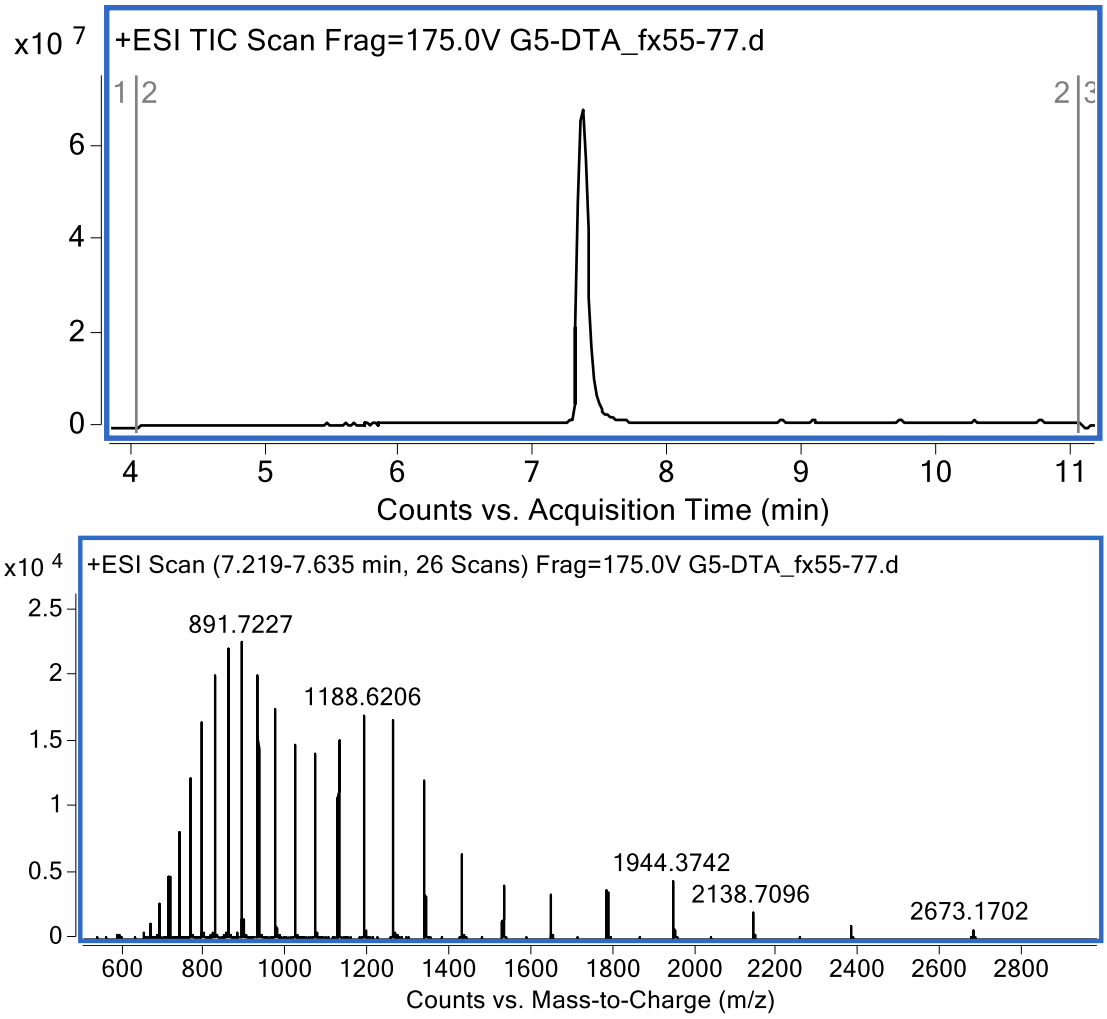




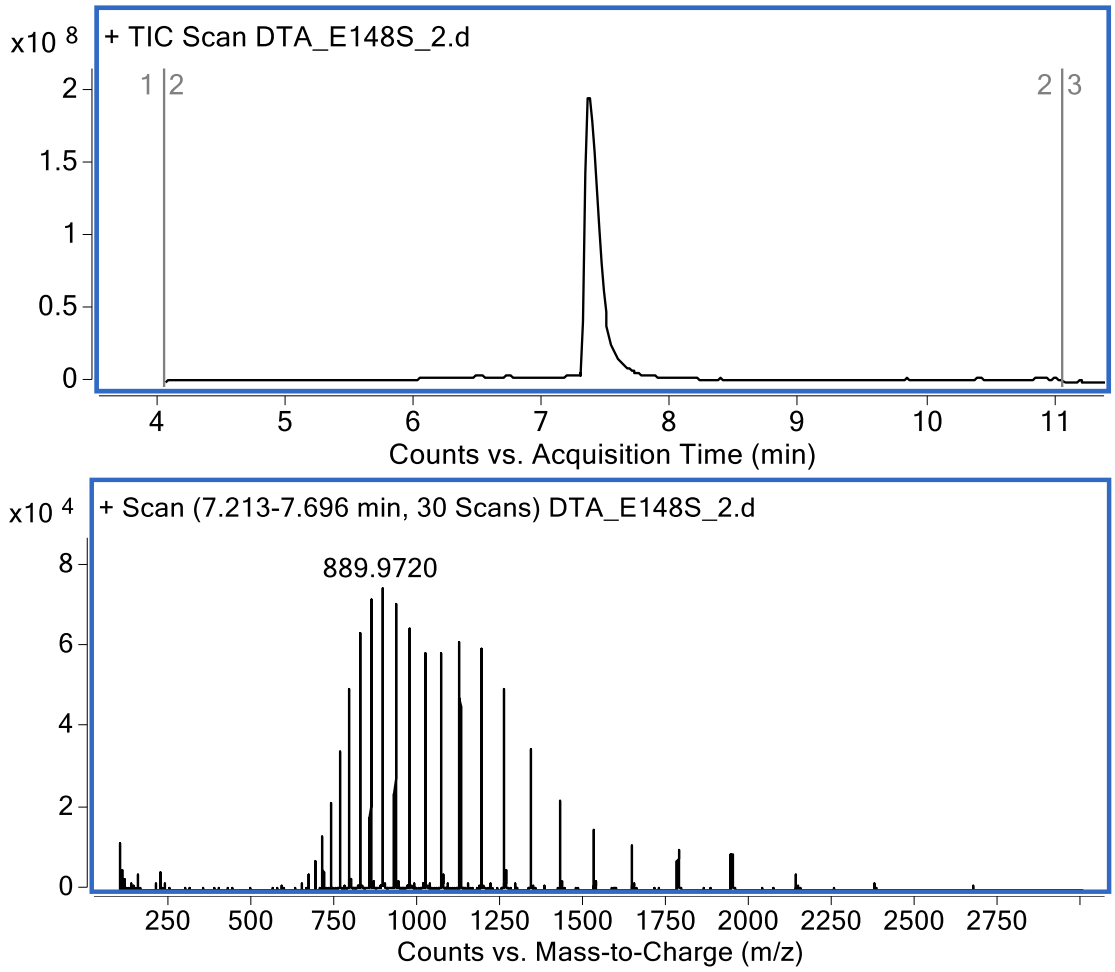
PNA-Mach7 (Method B)
Mass Expected: 11521.8 Da
Mass Observed: 11521.3 Da



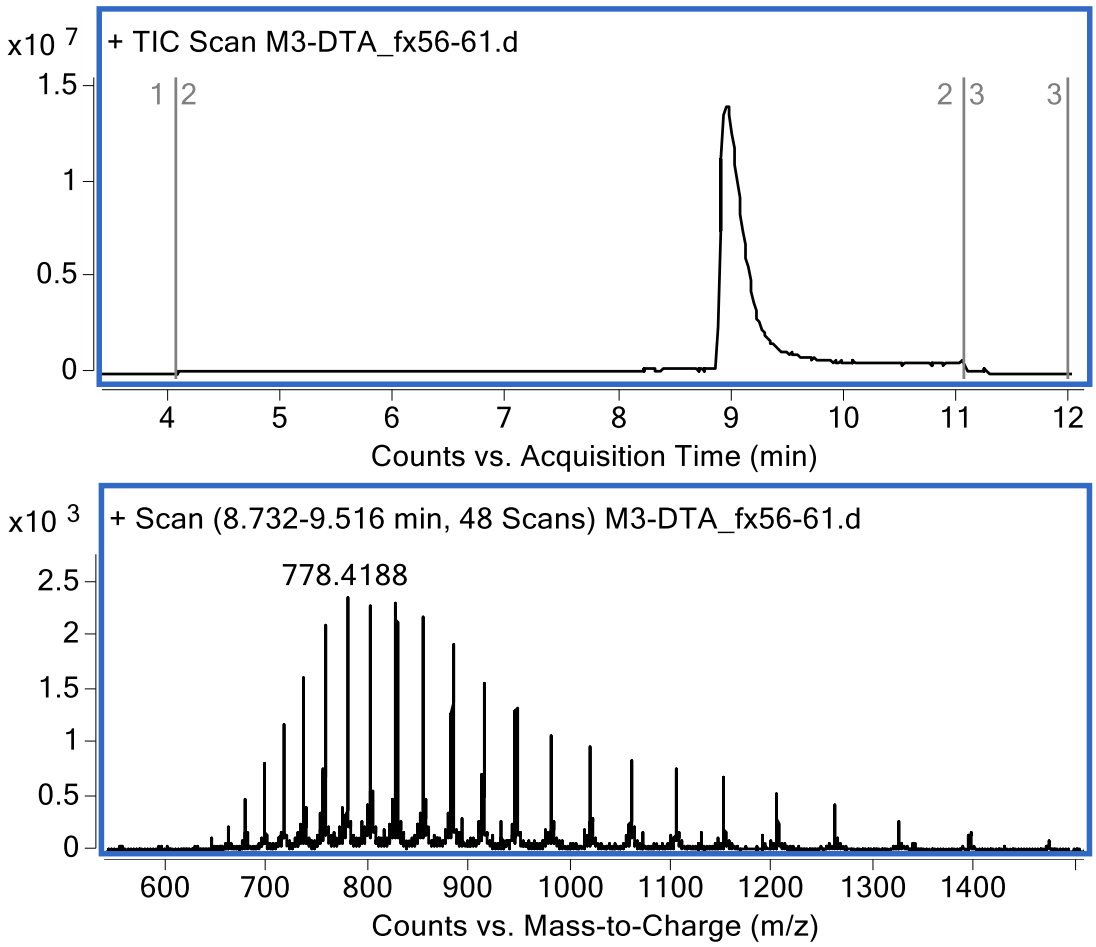
G5-DTA(C186S) (Method A)
Mass Expected: 21376.8 Da
Mass Observed: 21377.2 Da



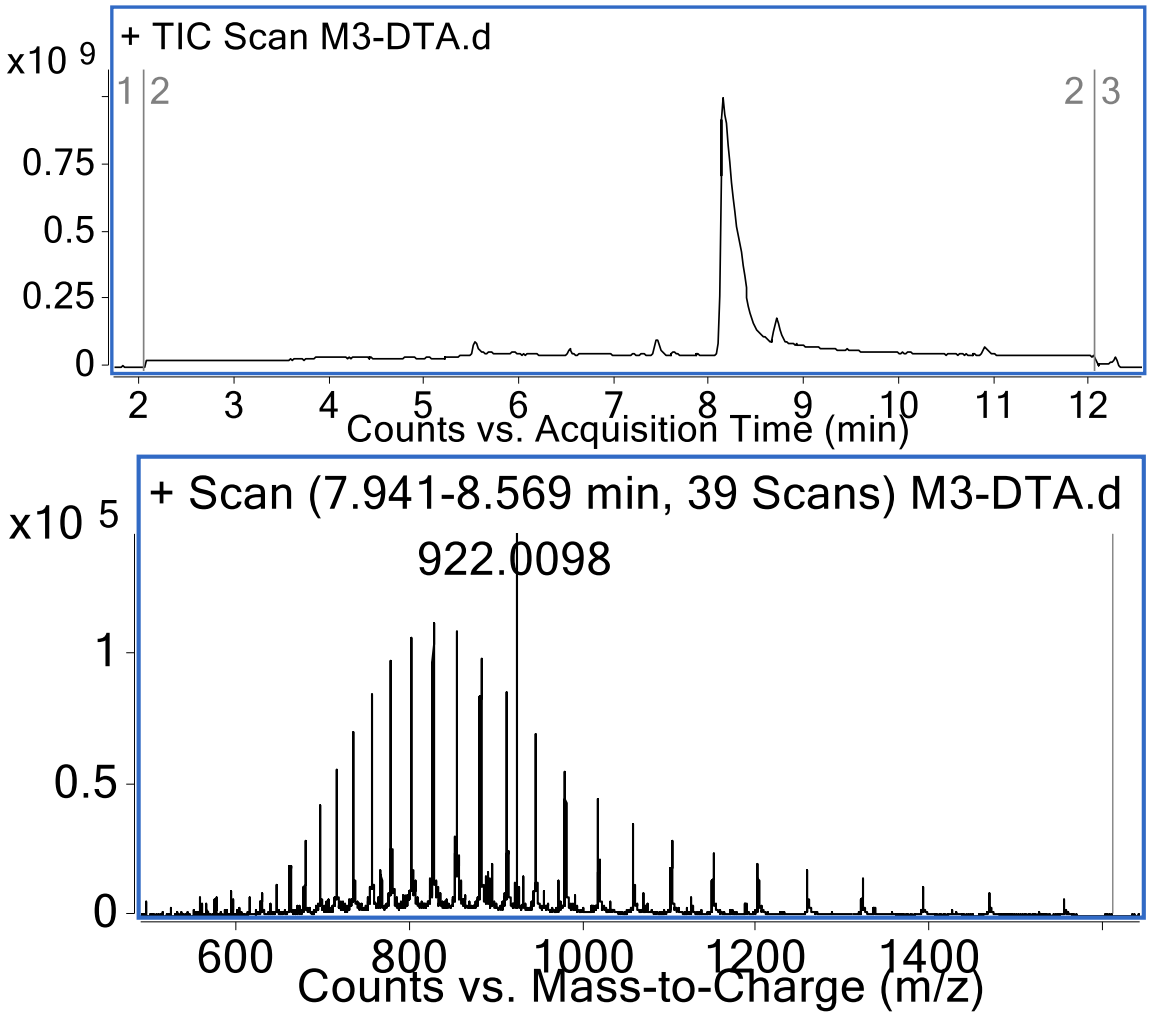
G5-DTA(C186S, E148S) (Method A)
Mass Expected: 21334.6 Da
Mass Observed: 21335.3 Da



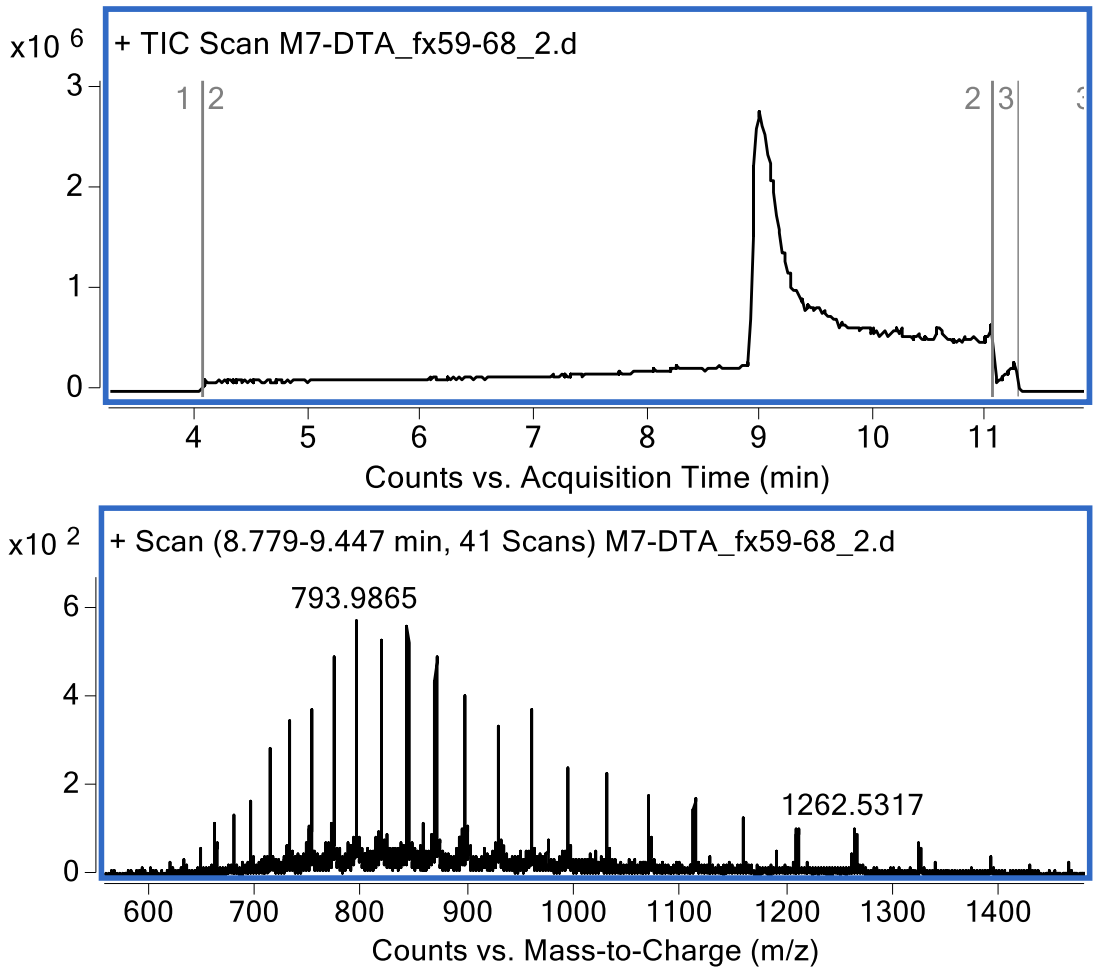
Mach3-DTA(C186S) (Method A)
Mass Expected: 26428.7 Da
Mass Observed: 26432.0 Da



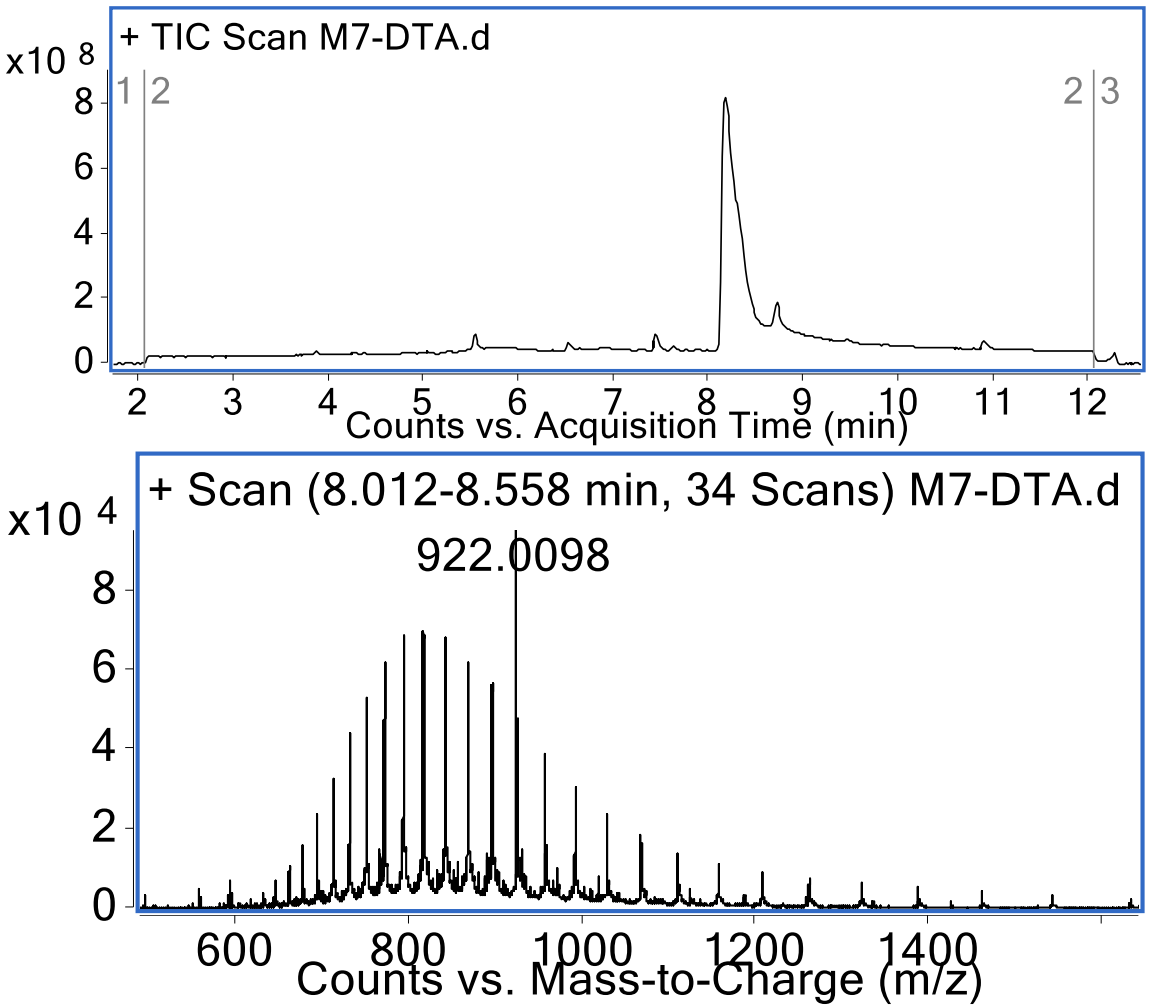
Mach3-DTA(C186S, E148S) (Method B)
Mass Expected: 26386.7 Da
Mass Observed: 26388.2 Da



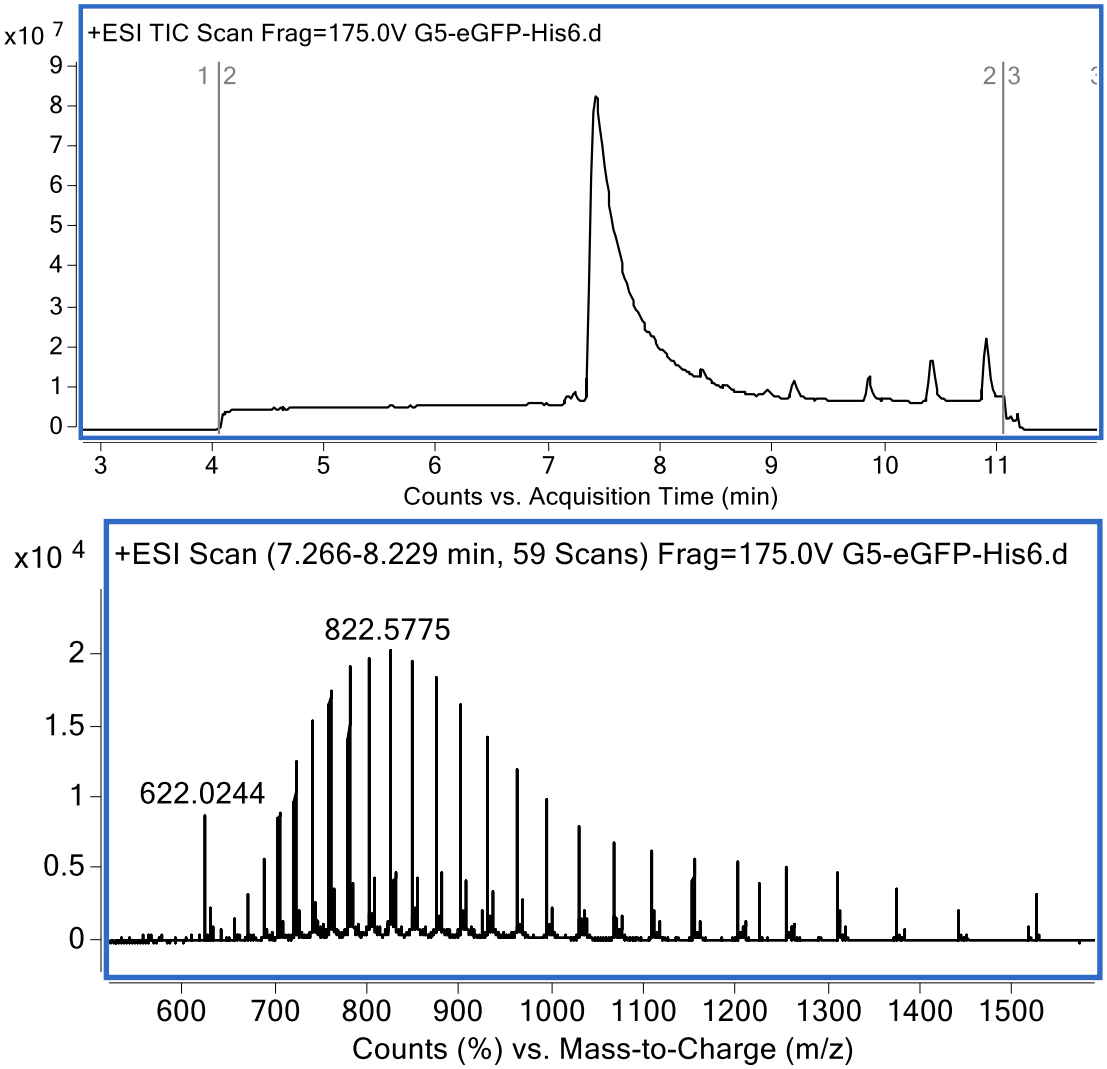
Mach7-DTA(C186S) (Method A)
Mass Expected: 27750.5 Da
Mass Observed: 27755.1 Da



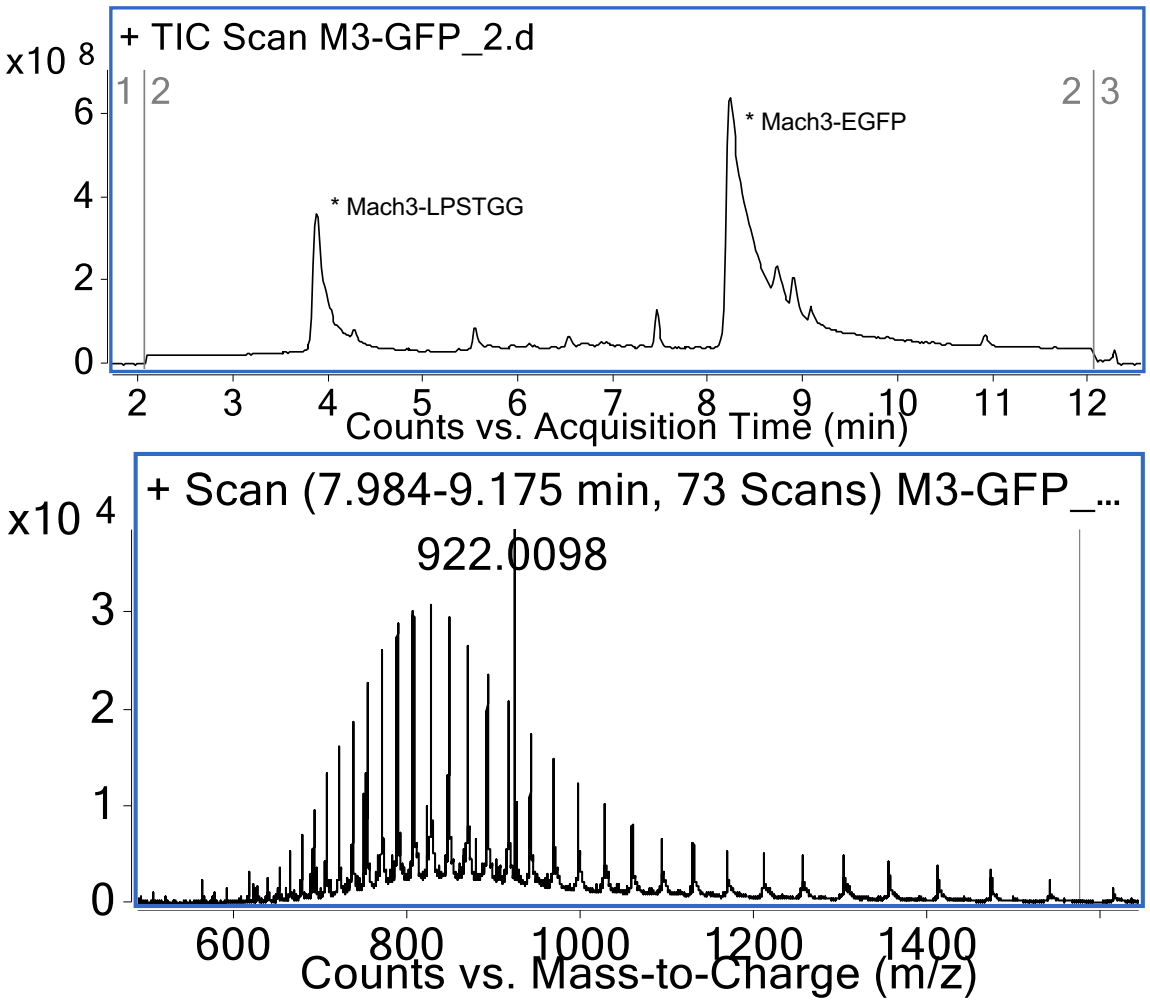
Mach7-DTA(C186S, E148S) (Method B)
Mass Expected: 27708.5 Da
Mass Observed: 27710.1 Da



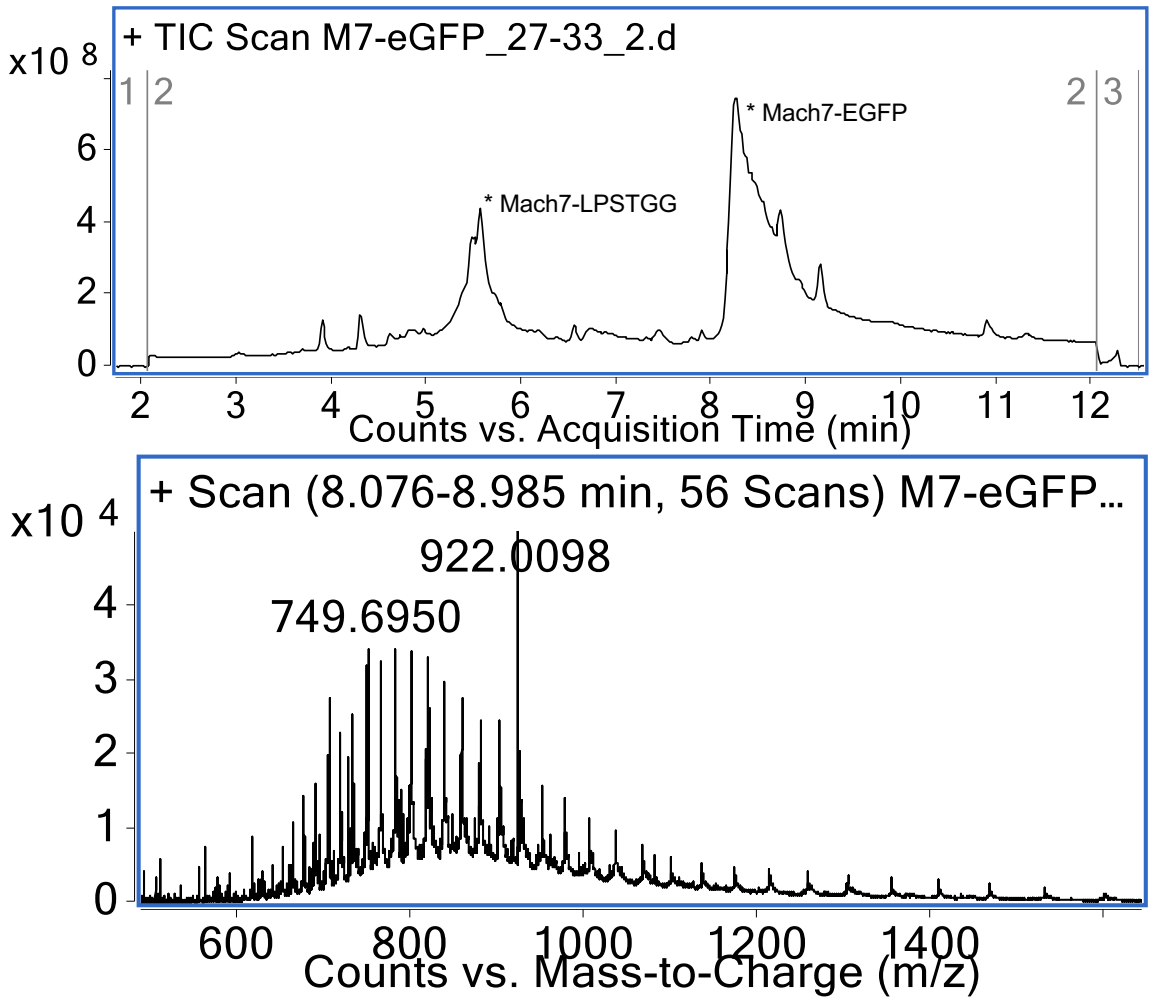
G5-EGFP (Method A)
Mass Expected: 28754.4 Da
Mass Observed: 28754.8 Da



Mach3-EGFP (Method B)
Mass Expected: 33806.5 Da
Mass Observed: 33807.3 Da



Mach7-EGFP (Method B)
Mass Expected: 35128.3 Da
Mass Observed: 35130.3 Da



2.9 Appendix III: Topological Fingerprints

CB_Index = Condensed Bit-vector index, used in the figures in Manuscript and SI
TF_Index = Topological Fingerprint index, for the corresponding CB_index

All ON bits out of the 2048-bits have been represented in the following set of figures. The radius of exploration goes from 0 (atom, itself) to 3 nearest neighbors. The coloring scheme denotes the node atom in blue, atoms which are a part of an aromatic ring in yellow, connected neighbors as a part of the topological exploration in black, and the unexplored neighboring atoms and nodes in gray.

CB_Index	TF_Index	CB_Index	TF_Index	CB_Index	TF_Index	CB_Index	TF_Index
1	1	**51**	585	**101**	1114	**151**	1693

2	11	**52**	623	**102**	1117	**152**	1719
3	22	**53**	625	**103**	1127	**153**	1731
4	27	**54**	650	**104**	1139	**154**	1736
5	32	**55**	667	**105**	1141	**155**	1737
6	67	**56**	671	**106**	1143	**156**	1750
7	70	**57**	680	**107**	1145	**157**	1751
8	74	**58**	708	**108**	1152	**158**	1752
9	79	**59**	713	**109**	1158	**159**	1754
10	80	**60**	724	**110**	1171	**160**	1758
11	119	**61**	727	**111**	1185	**161**	1773
12	132	**62**	739	**112**	1199	**162**	1778
13	140	**63**	742	**113**	1213	**163**	1783
14	150	**64**	745	**114**	1221	**164**	1785
15	173	**65**	759	**115**	1226	**165**	1791
16	197	**66**	776	**116**	1258	**166**	1794
17	204	**67**	784	**117**	1259	**167**	1805
18	220	**68**	785	**118**	1267	**168**	1840
19	222	**69**	786	**119**	1268	**169**	1844
20	227	**70**	806	**120**	1283	**170**	1847
21	229	**71**	807	**121**	1287	**171**	1849
22	231	**72**	831	**122**	1290	**172**	1873
23	272	**73**	857	**123**	1301	**173**	1876
24	280	**74**	878	**124**	1307	**174**	1879
25	283	**75**	889	**125**	1313	**175**	1882
26	289	**76**	894	**126**	1325	**176**	1898
27	293	**77**	900	**127**	1349	**177**	1910
28	294	**78**	926	**128**	1357	**178**	1911
29	295	**79**	931	**129**	1380	**179**	1912
30	305	**80**	955	**130**	1388	**180**	1917
31	310	**81**	966	**131**	1427	**181**	1926
32	321	**82**	971	**132**	1431	**182**	1928
33	328	**83**	981	**133**	1451	**183**	1937
34	329	**84**	983	**134**	1452	**184**	1946
35	362	**85**	989	**135**	1459	**185**	1947
36	364	**86**	1014	**136**	1462	**186**	1969
37	368	**87**	1017	**137**	1507	**187**	1970
38	376	**88**	1019	**138**	1517	**188**	2006
39	378	**89**	1022	**139**	1544	**189**	2013
40	389	**90**	1027	**140**	1547	**190**	2022
41	394	**91**	1028	**141**	1558	**191**	2042
42	412	**92**	1031	**142**	1564		
43	420	**93**	1034	**143**	1573		
44	425	**94**	1057	**144**	1601		
45	473	**95**	1066	**145**	1602		
46	482	**96**	1072	**146**	1607		
47	545	**97**	1082	**147**	1633		
48	553	**98**	1088	**148**	1656		
49	561	**99**	1104	**149**	1661		
50	575	**100**	1110	**150**	1685		

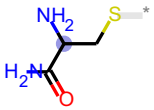
Linker 2

1

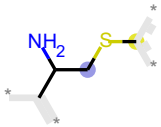
80

116

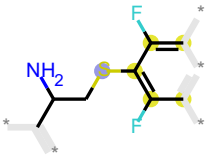
140



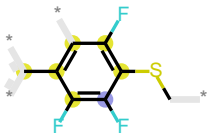
210



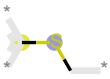
222



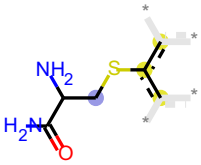
302



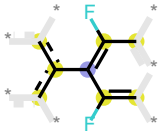
319



341



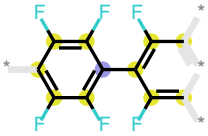
480



551

650

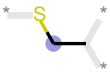
699



773

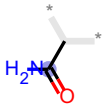
786

807

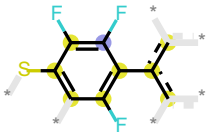


1093

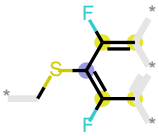
1171



1212

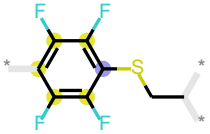


1234

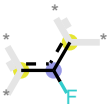


1305

1380

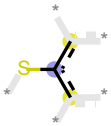


1436

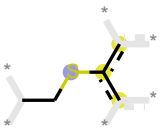


1445

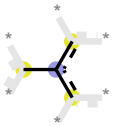
1654



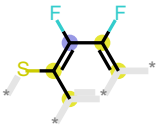
1716



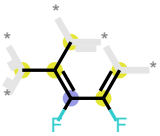
1733



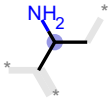
1747



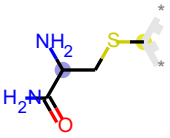
1821



1848



1849



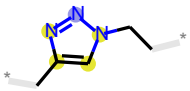
1855

1917

1928

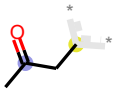
Linker 3

1



72

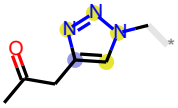
80



109

124

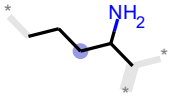
140



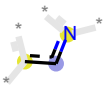
186

325

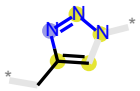
378



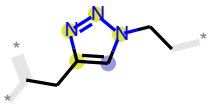
412



464

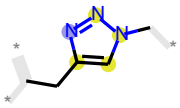


488



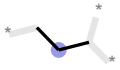
567

650

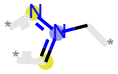


669

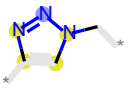
724



739



747

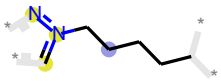


781

786

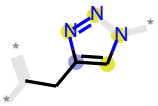
807

893



913

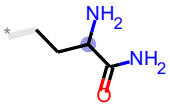
935



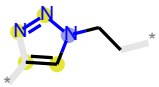
980

1017

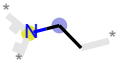
1057



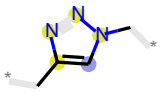
1127



1141



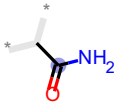
1145



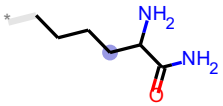
1155

1164

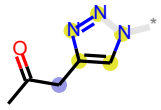
1171



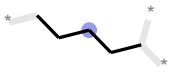
1212



1214

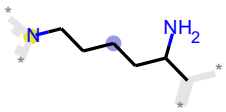


1259



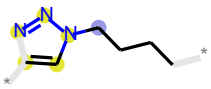
1283

1380

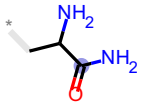


1536

1550

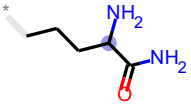


1582



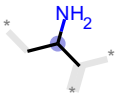
1654

1689



1728

1783

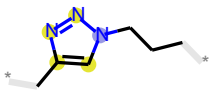


1849

1873

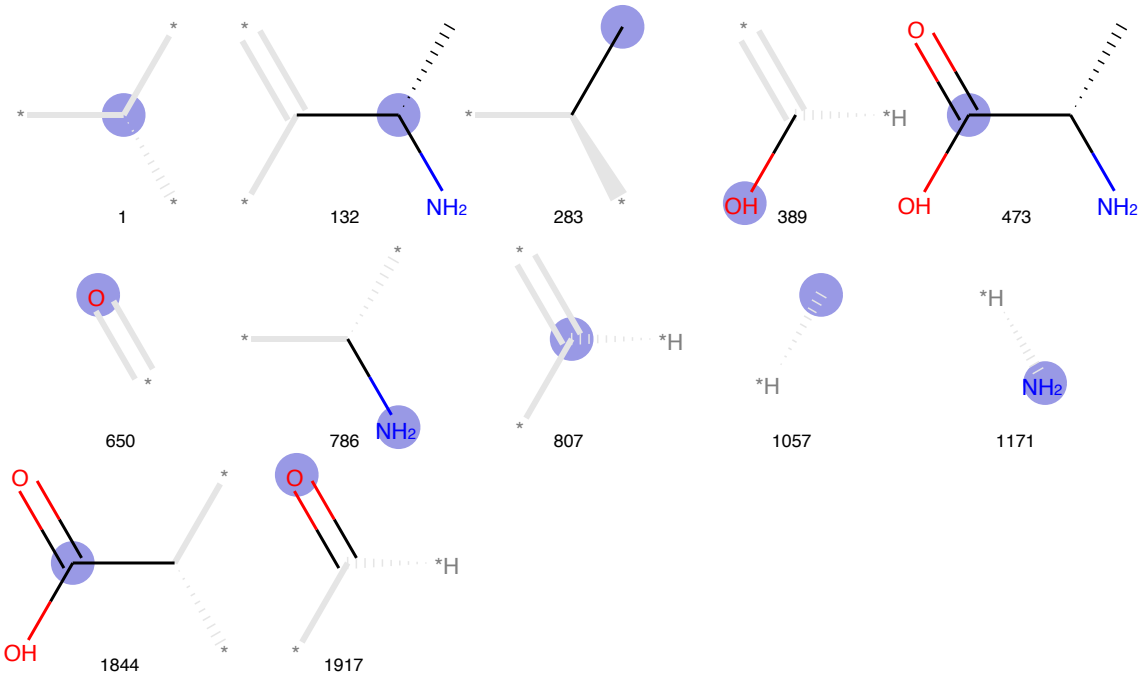
1911

1917



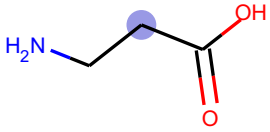
1971

Alanine


1
132
NH2
283
OH
389
*H
O
OH
473
NH2
O
650
786
NH2
807
*H
*H
1057
*H
NH2
1171
O
OH
1844
O
*H
1917

Beta-Alanine

80



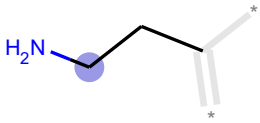
173

389

650

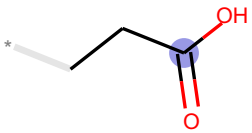
807

981

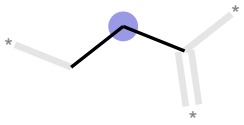


1088

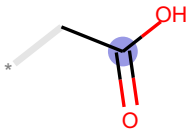
1171



1287

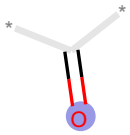


1564



1737

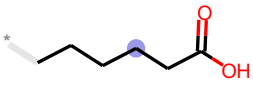
1840



1917

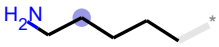
Aminohexanoic acid

80



295

389

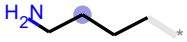


561

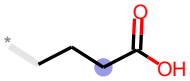
650

807

981



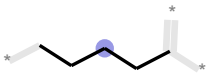
1082



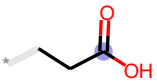
1110

1143

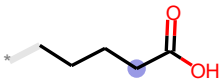
1171



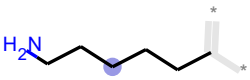
1267



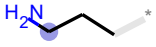
1287



1301

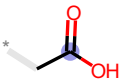


1462

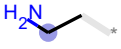


1517

1564



1737



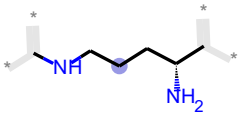
1840

1911

1917

Arginine

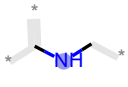
1



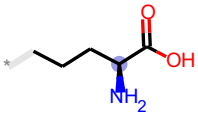
67

80

140

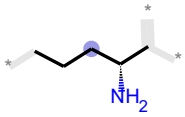


197



289

389



412

623

650

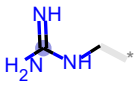
667

708

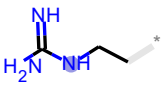
739

786

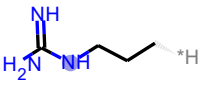
807



894

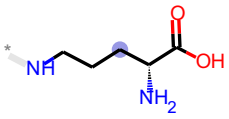


983



1027

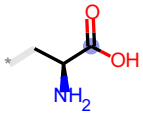
1104



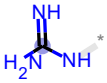
1127

1152

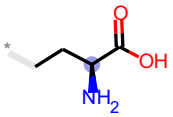
1171



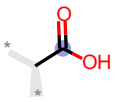
1427



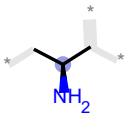
1451



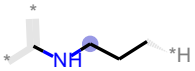
1791



1844



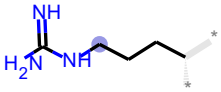
1849



1876

1911

1917

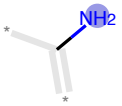


2042

Asparagine

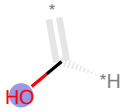
1

80



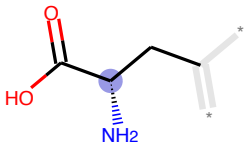
140

376

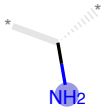


389

650



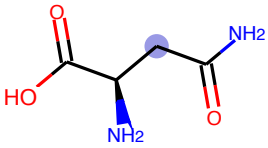
776



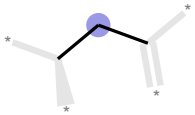
786

807

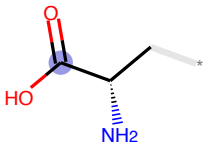
1171



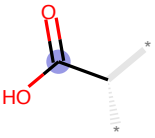
1213



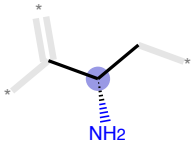
1290



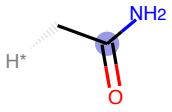
1427



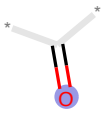
1844



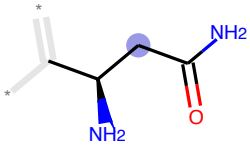
1849



1898

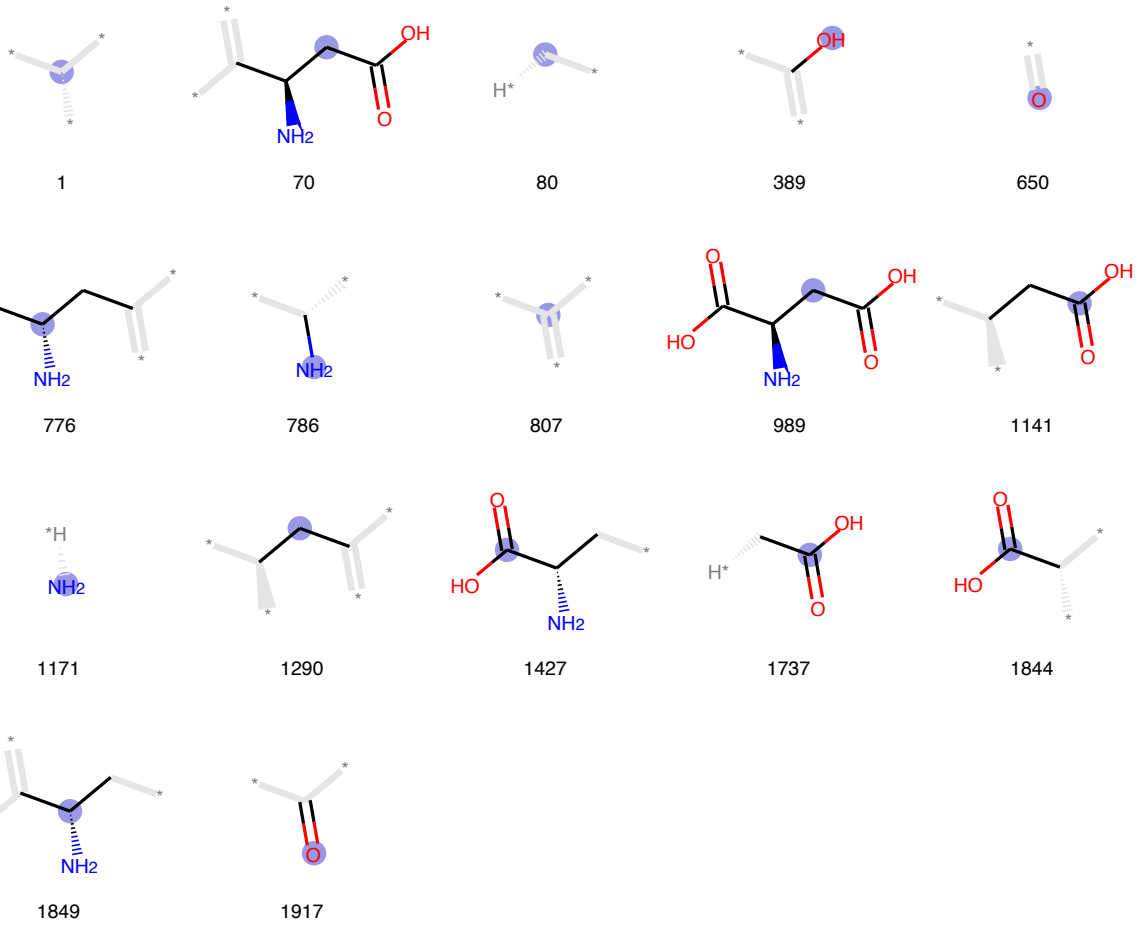


1917



1946

Aspartic acid

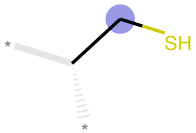


Cysteine

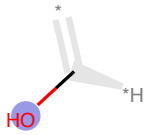
1

80

229

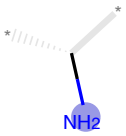


321



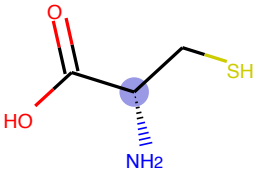
389

650



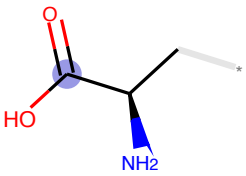
786

807



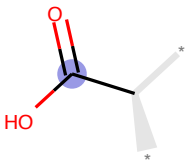
1072

1171

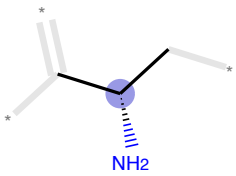


1427

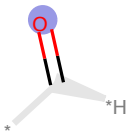
1459



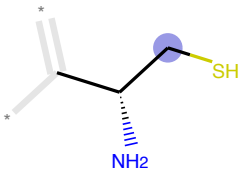
1844



1849



1917

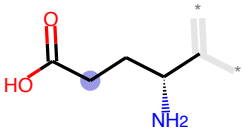


1926

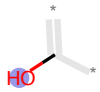
Glutamic acid

1

80



293



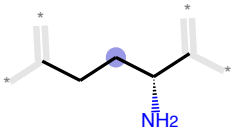
389

650

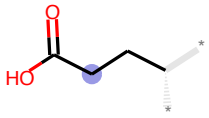
739

786

807

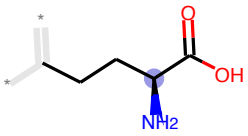


900

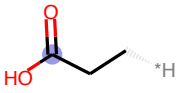


955

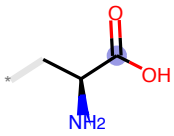
1171



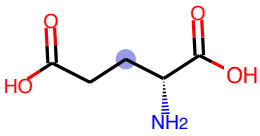
1258



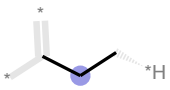
1287



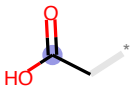
1427



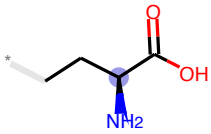
1547



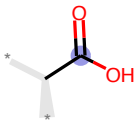
1564



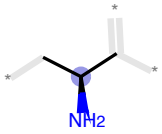
1737



1791



1844



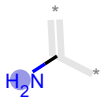
1849

1917

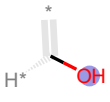
Glutamine

1

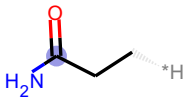
80



140

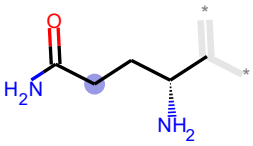


389



545

650

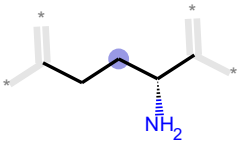


671

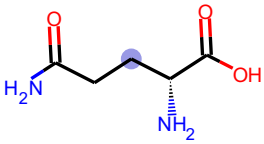
739

786

807

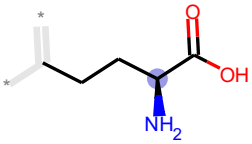


900

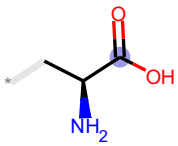


971

1171

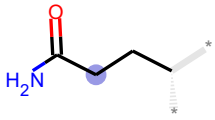


1258

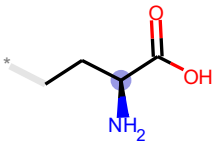


1427

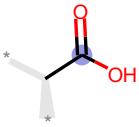
1564



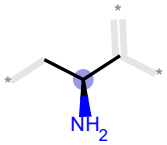
1752



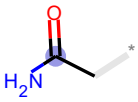
1791



1844



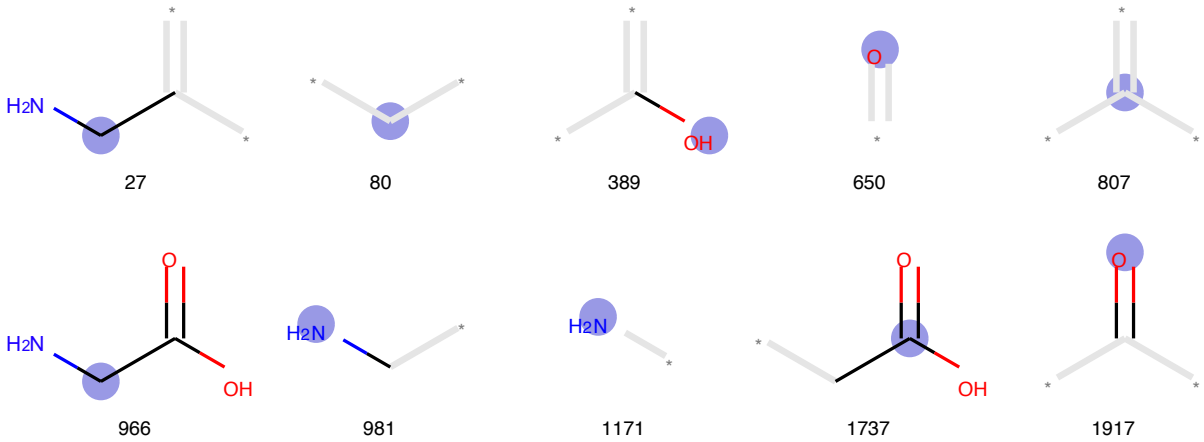
1849



1898

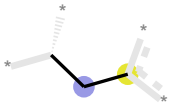
1917

Glycine


H2N
27
80
OH
389
O
650
807
O
H2N
OH
966
H2N
981
H2N
1171
O
OH
1737
O
1917

Histidine

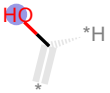
1



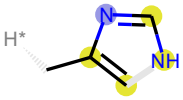
79

80

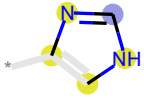
378



389



575

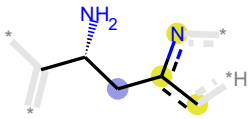


625

650

713

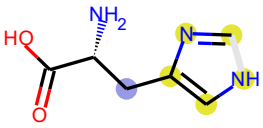
724



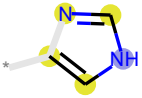
785

786

807



889

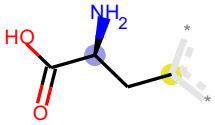


931

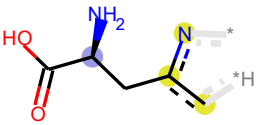
1114

1145

1171

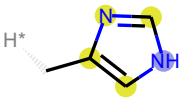


1221

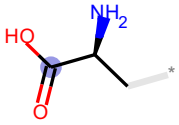


1259

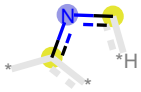
1380



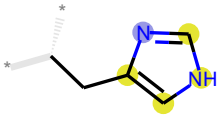
1388



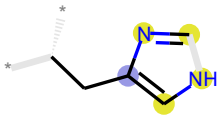
1427



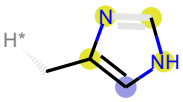
1452



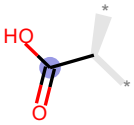
1601



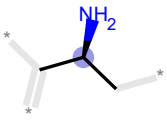
1633



1758

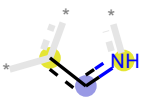


1844

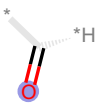


1849

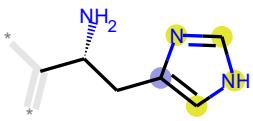
1873



1879

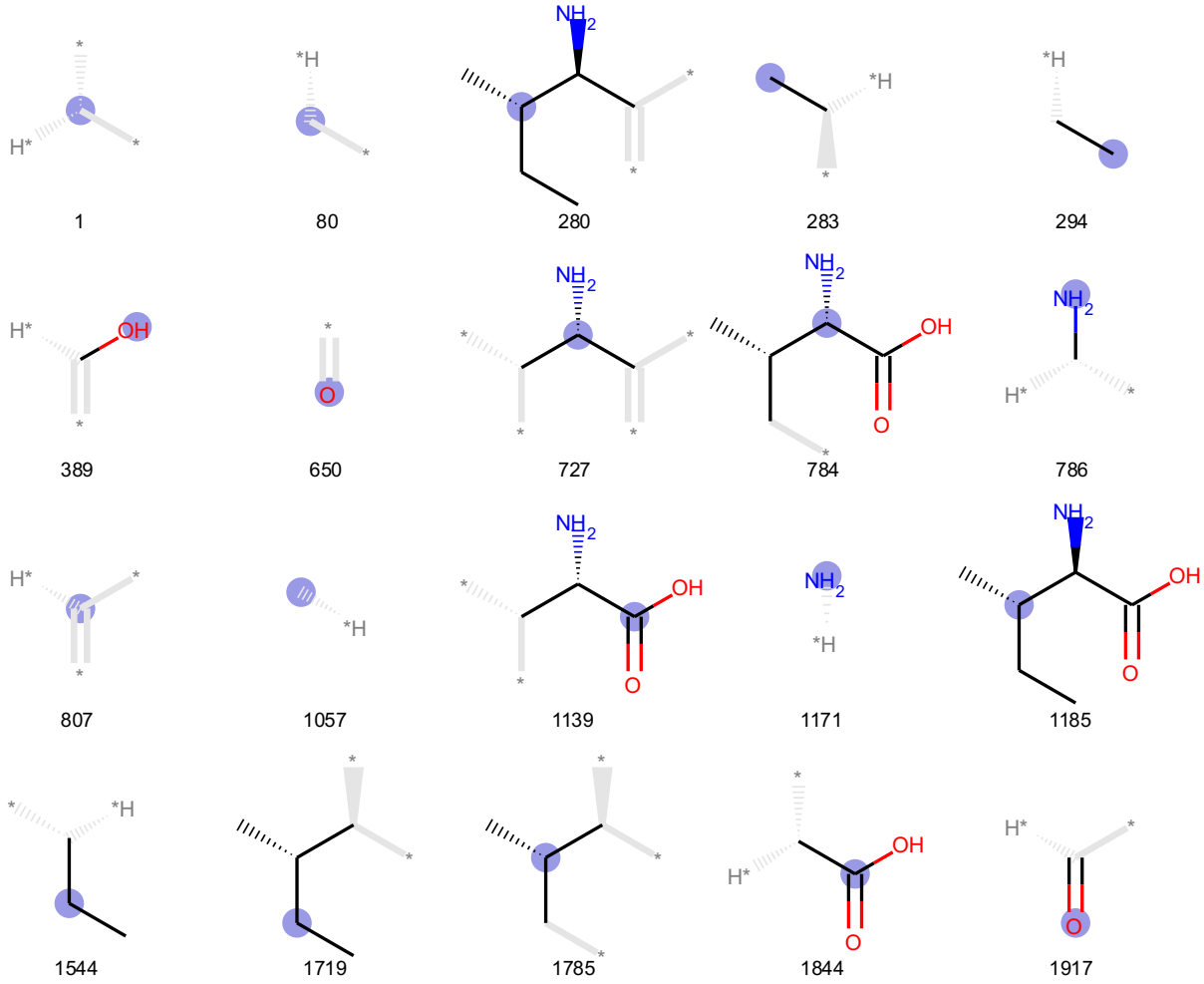


1917



1969

Isoleucine

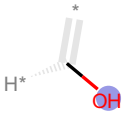

1
80
280
283
294
389
650
727
784
786
807
1057
1139
1171
1185
1544
1719
1785
1844
1917

Leucine

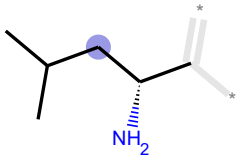
1

80

283

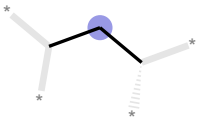


389

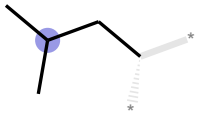


425

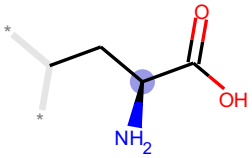
650



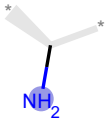
680



759

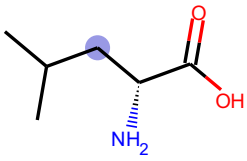


776



786

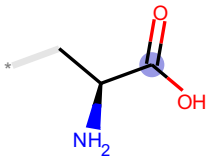
807



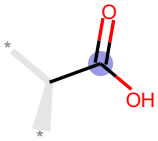
878

1057

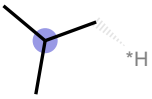
1171



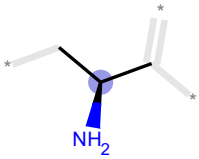
1427



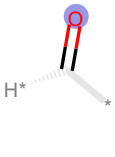
1844



1847



1849

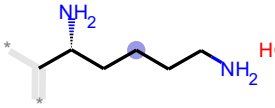


1917

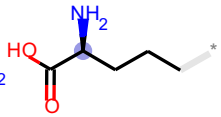
Lysine

1

80

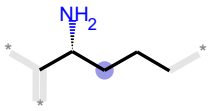


220



289

389



412

650

739

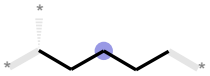
786

807

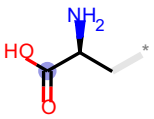
981

1082

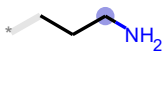
1171



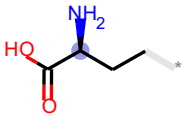
1283



1427

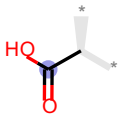


1517

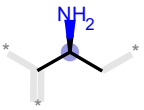


1791

1840



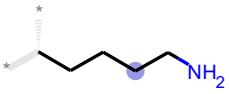
1844



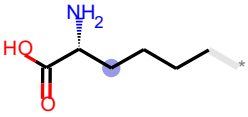
1849

1911

1917



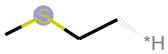
1928



2006

Methionine

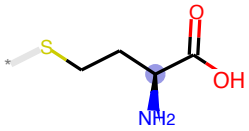
1



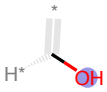
11

80

116



368



389

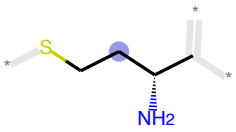
394

650

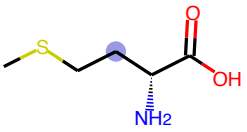
739

786

807



1021

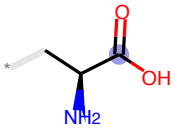


1034

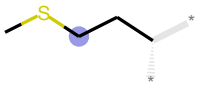
1057

1171

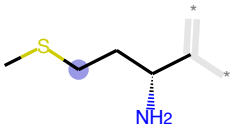
1182



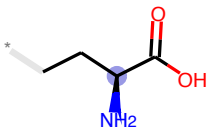
1427



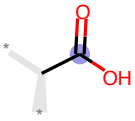
1558



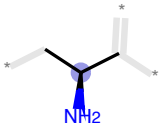
1661



1791



1844



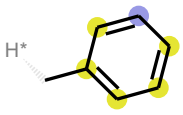
1849

1917

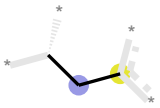
1942

Phenylalanine

1

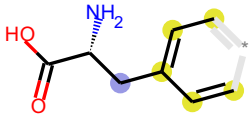


32

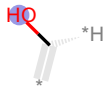


79

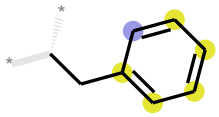
80



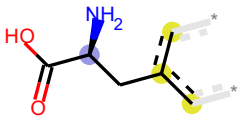
150



389



420

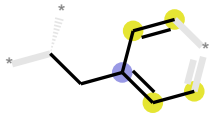


585

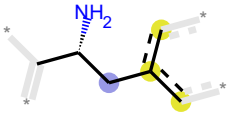
650

786

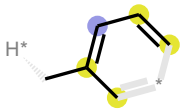
807



857



1017

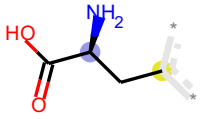


1066

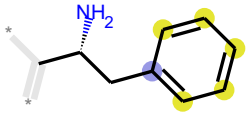
1088

1171

1199

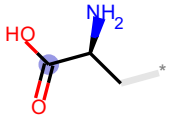


1221



1307

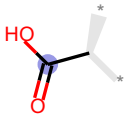
1380



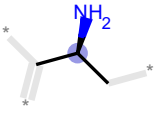
1427

1750

1754



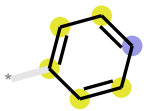
1844



1849

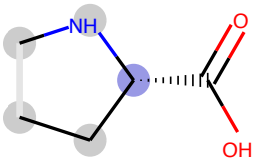
1873

1917

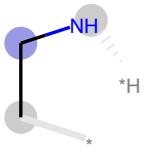


1947

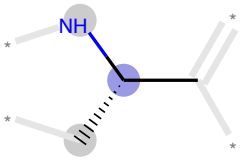
Proline



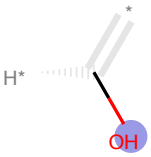
74



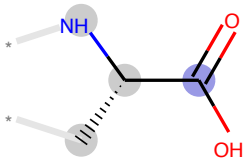
305



362

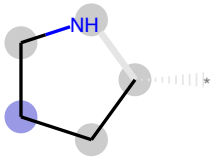


389



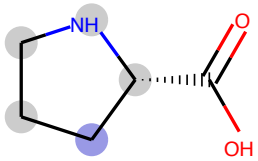
553

650



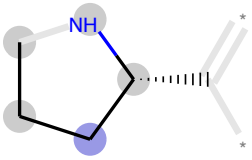
742

807



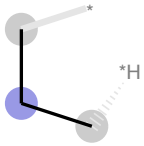
831

926

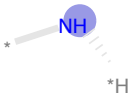


1014

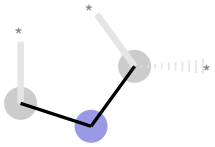
1019



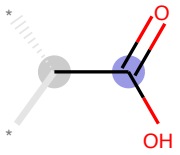
1028



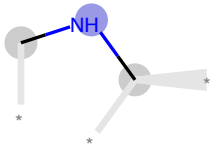
1114



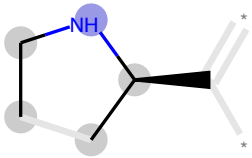
1325



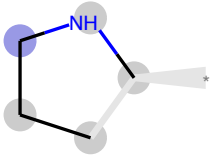
1431



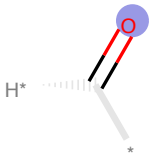
1507



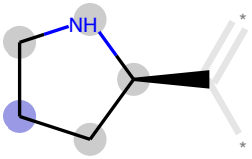
1736



1912

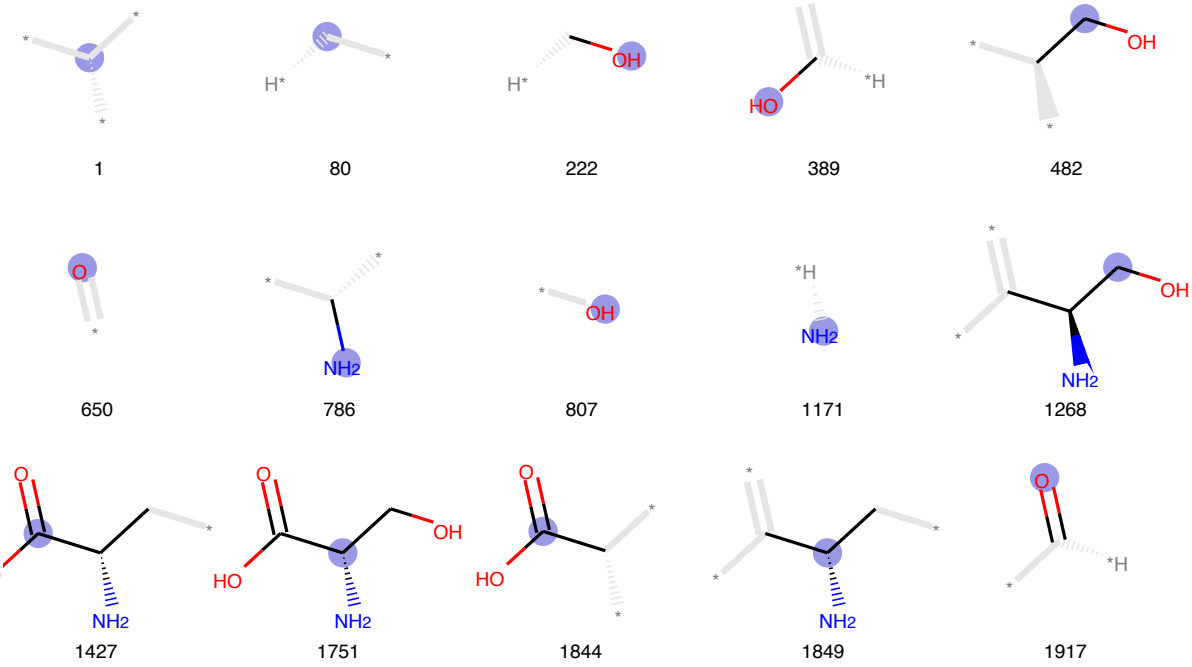


1917



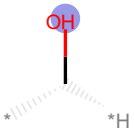
2022

Serine

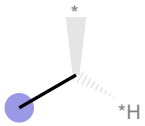


Threonine

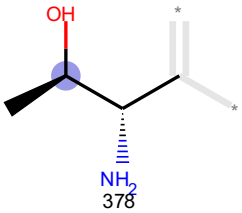
1



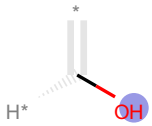
227



283

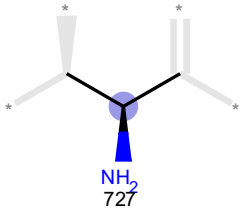


378

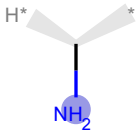


389

650



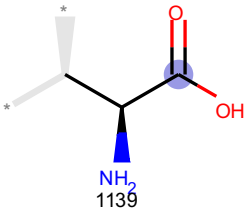
727



786

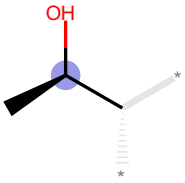
807

1057

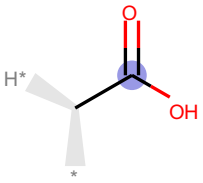


1139

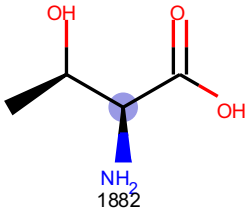
1171



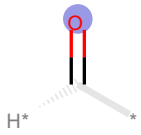
1693



1844



1882



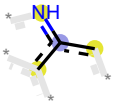
1917

Tryptophan

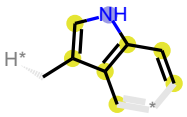
1

79

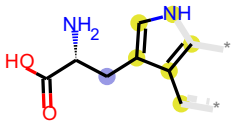
80



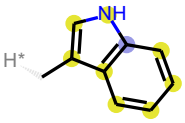
119



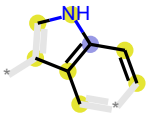
140



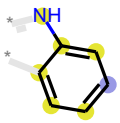
204



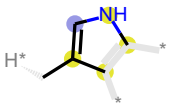
272



328



329

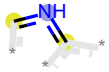


364

389

650

786

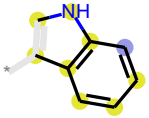


806

807

1088

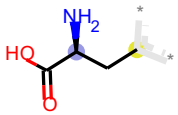
1114



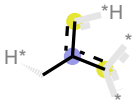
1158

1171

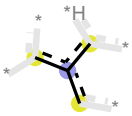
1199



1221

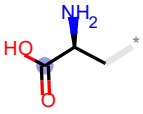


1349

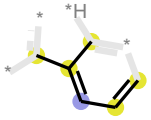


1357

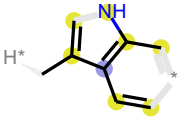
1380



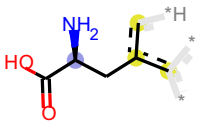
1427



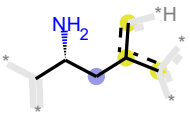
1573



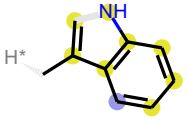
1607



1656

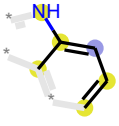


1685

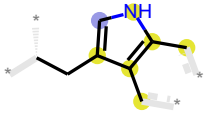


1731

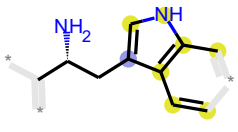
1750



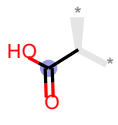
1783



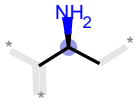
1794



1805



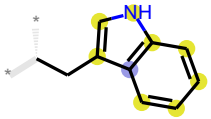
1844



1849

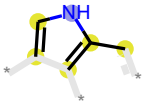
1873

1879

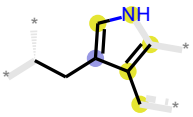


1910

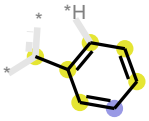
1917



1937



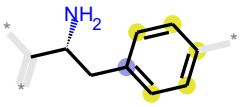
1970



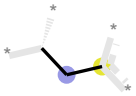
2013

Tyrosine

1

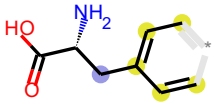


22

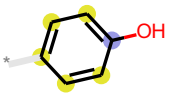


79

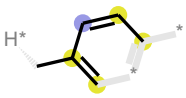
80



150

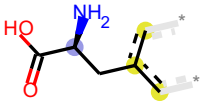


231



310

389



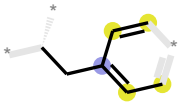
585

650

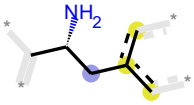
745

786

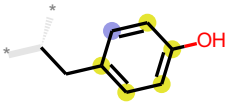
807



857

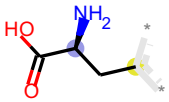


1017

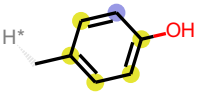


1117

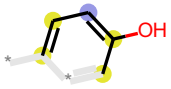
1171



1221

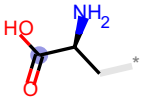


1226



1313

1380

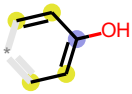


1427

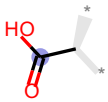
1602

1750

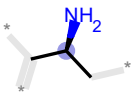
1754



1778



1844



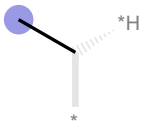
1849

1873

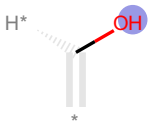
1917

Valine

1

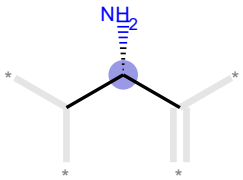


283



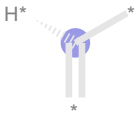
389

650

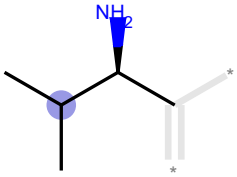


727

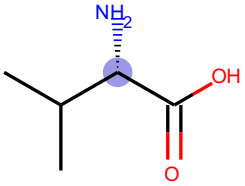
786



807

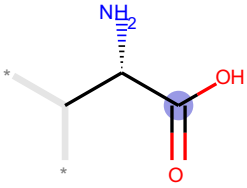


1022



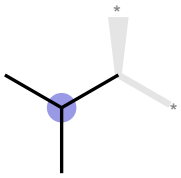
1031

1057

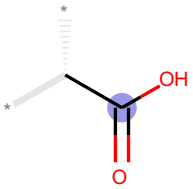


1139

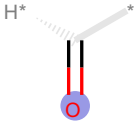
1171



1773



1844



1917

References

(1) Exploring chemical space: Can AI take us where no human has gone before? https://cen.acs.org/physical-chemistry/computational-chemistry/Exploring-chemical-space-AI-take/98/i13 (accessed 2020 -05 -12).

(2) Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; Volkov, Y.; Zholus, A.; Shayakhmetov, R. R.; Zhebrak, A.; Minaeva, L. I.; Zagribelnyy, B. A.; Lee, L. H.; Soll, R.; Madge, D.; Xing, L. Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors. https://doi.org/10.1038/s41587-019-0224-x.

(3) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackerman, Z. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180* (4), 688–702.

(4) Spänig, S.; Heider, D. Encodings and Models for Antimicrobial Peptide Classification for Multi-Resistant Pathogens. *BioData Min* **2019**, *12*, 7. https://doi.org/10.1186/s13040-019-0196-x.

(5) Witten, J.; Witten, Z. Deep Learning Regression Model for Antimicrobial Peptide Design. *bioRxiv* **2019**, 692681. https://doi.org/10.1101/692681.

(6) Liu, G.; Zeng, H.; Mueller, J.; Carter, B.; Wang, Z.; Schilz, J.; Horny, G.; Birnbaum, M. E.; Ewert, S.; Gifford, D. K. Antibody Complementarity Determining Region Design Using High-Capacity Machine Learning. *Bioinformatics* **2020**, *36* (7), 2126–2133. https://doi.org/10.1093/bioinformatics/btz895.

(7) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS central science* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.

(8) Su, R.; Hu, J.; Zou, Q.; Manavalan, B.; Wei, L. Empirical Comparison and Analysis of Web-Based Cell-Penetrating Peptide Prediction Tools. *Brief. Bioinform* **2019**.

(9) Sanders, W. S.; Johnston, C. I.; Bridges, S. M.; Burgess, S. C.; Willeford, K. O. Prediction of Cell Penetrating Peptides by Support Vector Machines. *PLoS computational biology* **2011**, *7* (7), e1002101.

(10) Manavalan, B.; Subramaniyam, S.; Shin, T. H.; Kim, M. O.; Lee, G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *Journal of Proteome Research* **2018**, *17* (8), 2715–2726. https://doi.org/10.1021/acs.jproteome.8b00148.

(11) Crook, Z. R.; Nairn, N. W.; Olson, J. M. Miniproteins as a Powerful Modality in Drug Development. *Trends in Biochemical Sciences* **2020**, *45* (4), 332–346. https://doi.org/10.1016/j.tibs.2019.12.008.

(12) Beaulieu, M.-E.; Jauset, T.; Massó-Vallés, D.; Martínez-Martín, S.; Rahl, P.; Maltais, L.; Zacarias-Fluck, M. F.; Casacuberta-Serra, S.; Pozo, E. S. del; Fiore, C.; Foradada, L.; Cano, V. C.; Sánchez-Hervás, M.; Guenther, M.; Sanz, E. R.; Oteo, M.; Tremblay, C.; Martín, G.; Letourneau, D.; Montagne, M.; Alonso, M. Á. M.; Whitfield, J. R.; Lavigne, P.; Soucek, L. Intrinsic Cell-Penetrating Activity Propels Omomyc from Proof of Concept to Viable Anti-MYC Therapy. *Science Translational Medicine* **2019**, *11* (484). https://doi.org/10.1126/scitranslmed.aar5012.

(13) Juliano, R. L. The Delivery of Therapeutic Oligonucleotides. *Nucleic acids research* **2016**, *44* (14), 6518–6548.
(14) Slastnikova, T. A.; Ulasov, A. V; Rosenkranz, A. A.; Sobolev, A. S. Targeted Intracellular Delivery of Antibodies: The State of the Art. *Frontiers in pharmacology* **2018**, *9*.
(15) Miersch, S.; Sidhu, S. S. Intracellular Targeting with Engineered Proteins. *F1000Research* **2016**, *5*.
(16) Trenevska, I.; Li, D.; Banham, A. H. Therapeutic Antibodies against Intracellular Tumor Antigens. *Frontiers in immunology* **2017**, *8*, 1001.
(17) Fu, A.; Tang, R.; Hardie, J.; Farkas, M. E.; Rotello, V. M. Promises and Pitfalls of Intracellular Delivery of Proteins. *Bioconjugate chemistry* **2014**, *25* (9), 1602–1608.
(18) Illien, F.; Rodriguez, N.; Amoura, M.; Joliot, A.; Pallerla, M.; Cribier, S.; Burlina, F.; Sagan, S. Quantitative Fluorescence Spectroscopy and Flow Cytometry Analyses of Cell-Penetrating Peptides Internalization Pathways: Optimization, Pitfalls, Comparison with Mass Spectrometry Quantification. *Scientific Reports* **2016**, *6*, 36938. https://doi.org/10.1038/srep36938.
(19) Wolfe, J. M.; Fadzen, C. M.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Perfluoroaryl Bicyclic Cell-Penetrating Peptides for Delivery of Antisense Oligonucleotides. *Angewandte Chemie* **2018**, *130* (17), 4846–4849.
(20) Betts, C.; Saleh, A. F.; Arzumanov, A. A.; Hammond, S. M.; Godfrey, C.; Coursindel, T.; Gait, M. J.; Wood, M. J. A. Pip6-PMO, a New Generation of Peptide-Oligonucleotide Conjugates with Improved Cardiac Exon Skipping Activity for DMD Treatment. *Molecular Therapy-Nucleic Acids* **2012**, *1*, e38.
(21) Boisguérin, P.; Deshayes, S.; Gait, M. J.; O'donovan, L.; Godfrey, C.; Betts, C. A.; Wood, M. J. A.; Lebleu, B. Delivery of Therapeutic Oligonucleotides with Cell Penetrating Peptides. *Advanced drug delivery reviews* **2015**, *87*, 52–67.
(22) Chery, J. RNA Therapeutics: RNAi and Antisense Mechanisms and Clinical Applications. *Postdoc Journal* **2016**, *4* (7), 35–50.
(23) Mendell, J. R.; Rodino-Klapac, L. R.; Sahenk, Z.; Roush, K.; Bird, L.; Lowes, L. P.; Alfano, L.; Gomez, A. M.; Lewis, S.; Kota, J. Eteplirsen for the Treatment of Duchenne Muscular Dystrophy. *Annals of neurology* **2013**, *74* (5), 637–647.
(24) Moulton, J.; Jiang, S. Gene Knockdowns in Adult Animals: PPMOs and Vivo-Morpholinos. *Molecules* **2009**, *14* (3), 1304–1323.
(25) McClorey, G.; Banerjee, S. Cell-Penetrating Peptides to Enhance Delivery of Oligonucleotide-Based Therapeutics. *Biomedicines* **2018**, *6* (2). https://doi.org/10.3390/biomedicines6020051.
(26) Inc, S. T. Sarepta Therapeutics Announces Positive Clinical Results from MOMENTUM, a Phase 2 Clinical Trial of SRP-5051 in Patients with Duchenne Muscular Dystrophy Amenable to Skipping Exon 51 http://www.globenewswire.com/news-release/2020/12/07/2140613/0/en/Sarepta-Therapeutics-Announces-Positive-Clinical-Results-from-MOMENTUM-a-Phase-2-Clinical-Trial-of-SRP-5051-in-Patients-with-Duchenne-Muscular-Dystrophy-Amenable-to-Skipping-Exon-5.html (accessed 2020 -12 -07).
(27) Cardozo, A. K.; Buchillier, V.; Mathieu, M.; Chen, J.; Ortis, F.; Ladrière, L.; Allaman-Pillet, N.; Poirot, O.; Kellenberger, S.; Beckmann, J. S.; Eizirik, D. L.; Bonny, C.; Maurer, F. Cell-Permeable Peptides Induce Dose- and Length-Dependent Cytotoxic Effects.

Biochimica et Biophysica Acta (BBA) - Biomembranes **2007**, *1768* (9), 2222–2234. https://doi.org/10.1016/j.bbamem.2007.06.003.
(28) Fadzen, C. M.; Holden, R. L.; Wolfe, J. M.; Choo, Z.-N.; Schissel, C. K.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. *Biochemistry* **2019**, *58* (38), 3980–3989.
(29) Wolfe, J. Peptide Conjugation to Enhance Oligonucleotide Delivery. Massachusetts Institute of Technology 2018.
(30) Wolfe, J. M. Peptide Conjugation to Enhance Oligonucleotide Delivery. *PhD Thesis, Massachusetts Institute of Technology* **2018**.
(31) Wei, L.; Tang, J.; Zou, Q. SkipCPP-Pred: An Improved and Promising Sequence-Based Predictor for Predicting Cell-Penetrating Peptides. *BMC genomics* **2017**, *18* (7), 742.
(32) Pandey, P.; Patel, V.; George, N. V; Mallajosyula, S. S. KELM-CPPpred: Kernel Extreme Learning Machine Based Prediction Model for Cell-Penetrating Peptides. *Journal of proteome research* **2018**, *17* (9), 3214–3222.
(33) Chen, B.; Khodadoust, M. S.; Olsson, N.; Wagar, L. E.; Fast, E.; Liu, C. L.; Muftuoglu, Y.; Sworder, B. J.; Diehn, M.; Levy, R.; Davis, M. M.; Elias, J. E.; Altman, R. B.; Alizadeh, A. A. Predicting HLA Class II Antigen Presentation through Integrated Deep Learning. *Nature Biotechnology* **2019**, *37* (November). https://doi.org/10.1038/s41587-019-0280-2.
(34) Lee, E. Y.; Wong, G. C. L.; Ferguson, A. L. Machine Learning-Enabled Discovery and Design of Membrane-Active Peptides. *Bioorganic and Medicinal Chemistry* **2018**, *26* (10), 2708–2718. https://doi.org/10.1016/j.bmc.2017.07.012.
(35) A Dobchev, D.; Mager, I.; Tulp, I.; Karelson, G.; Tamm, T.; Tamm, K.; Janes, J.; Langel, U.; Karelson, M. Prediction of Cell-Penetrating Peptides Using Artificial Neural Networks. *Current computer-aided drug design* **2010**, *6* (2), 79–89.
(36) Jearawiriyapaisarn, N.; Moulton, H. M.; Buckley, B.; Roberts, J.; Sazani, P.; Fucharoen, S.; Iversen, P. L.; Kole, R. Sustained Dystrophin Expression Induced by Peptide-Conjugated Morpholino Oligomers in the Muscles of Mdx Mice. *Molecular Therapy* **2008**, *16* (9), 1624–1629.
(37) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5* (2), 107–113. https://doi.org/10.1021/c160017a018.
(38) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754. https://doi.org/10.1021/ci100050t.
(39) Moniz, J. R. A.; Krueger, D. Nested LSTMs. *arXiv:1801.10308* **2018**.
(40) Agrawal, P.; Bhalla, S.; Usmani, S. S.; Singh, S.; Chaudhary, K.; Raghava, G. P. S.; Gautam, A. CPPsite 2.0: A Repository of Experimentally Validated Cell-Penetrating Peptides. *Nucleic acids research* **2015**, *44* (D1), D1098–D1103.
(41) Winkler, W. E. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. **1990**.
(42) Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision* **2017**, 618–626. https://doi.org/10.1109/ICCV.2017.74.

(43) McCloskey, K.; Taly, A.; Monti, F.; Brenner, M. P.; Colwell, L. J. Using Attribution to Decode Binding Mechanism in Neural Network Models for Chemistry. *PNAS* **2019**, *116* (24), 11624–11629. https://doi.org/10.1073/pnas.1820657116.
(44) Sanchez-Lengeling, B.; Wei, J. N.; Lee, B. K.; Gerkin, R. C.; Aspuru-Guzik, A.; Wiltschko, A. B. Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules. *arXiv:1910.10685 [physics, stat]* **2019**.
(45) Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Liu, S.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by Automated Flow Chemistry. *Science* **2020**, *368* (6494), 980–987. https://doi.org/10.1126/science.abb2491.
(46) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *Journal of molecular biology* **1990**, *215* (3), 403–410.
(47) Miller, S. E.; Tsuji, K.; Abrams, R. P. M.; Burke, T. R.; Schneider, J. P. Uncoupling the Folding-Function Paradigm of Lytic Peptides to Deliver Impermeable Inhibitors of Intracellular Protein–Protein Interactions. *J. Am. Chem. Soc.* **2020**, *142* (47), 19950–19955. https://doi.org/10.1021/jacs.0c07921.
(48) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G. Antisense and Antigene Properties of Peptide Nucleic Acids. *Science* **1992**, *258* (5087), 1481–1485.
(49) Choe, S.; Bennett, M. J.; Fujii, G.; Curmi, P. M. G.; Kantardjieff, K. A.; Collier, R. J.; Eisenberg, D. The Crystal Structure of Diphtheria Toxin. *Nature* **1992**, *357* (6375), 216.
(50) Wilson, B. A.; Reich, K. A.; Weinstein, B. R.; Collier, R. J. Active-Site Mutations of Diphtheria Toxin: Effects of Replacing Glutamic Acid-148 with Aspartic Acid, Glutamine, or Serine. *Biochemistry* **1990**, *29* (37), 8643–8651.
(51) Calis, J. J. A.; Maybeno, M.; Greenbaum, J. A.; Weiskopf, D.; De Silva, A. D.; Sette, A.; Keşmir, C.; Peters, B. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Computational Biology* **2013**, *9* (10). https://doi.org/10.1371/journal.pcbi.1003266.
(52) Abes, S.; Moulton, H. M.; Clair, P.; Prevot, P.; Youngblood, D. S.; Wu, R. P.; Iversen, P. L.; Lebleu, B. Vectorization of Morpholino Oligomers by the (R-Ahx-R)4 Peptide Allows Efficient Splicing Correction in the Absence of Endosomolytic Agents. *Journal of Controlled Release* **2006**, *116* (3), 304–313. https://doi.org/10.1016/j.jconrel.2006.09.011.
(53) Cerrato, C. P.; Künnapuu, K.; Langel, Ü. Cell-Penetrating Peptides with Intracellular Organelle Targeting. *Expert opinion on drug delivery* **2017**, *14* (2), 245–255.
(54) Nischan, N.; Herce, H. D.; Natale, F.; Bohlke, N.; Budisa, N.; Cardoso, M. C.; Hackenberger, C. P. R. Covalent Attachment of Cyclic TAT Peptides to GFP Results in Protein Delivery into Live Cells with Immediate Bioavailability. *Angewandte Chemie International Edition* **2015**, *54* (6), 1950–1953.
(55) Mijalis, A. J.; Thomas III, D. A.; Simon, M. D.; Adamo, A.; Beaumont, R.; Jensen, K. F.; Pentelute, B. L. A Fully Automated Flow-Based Approach for Accelerated Peptide Synthesis. *Nature chemical biology* **2017**, *13* (5), 464.
(56) Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by Automated

Flow Chemistry. *ChemRxiv. Preprint.* **2020**, No. 2. https://doi.org/10.26434/chemrxiv.11833503.v1.
(57) Sazani, P.; Gemignani, F.; Kang, S.-H.; Maier, M. A.; Manoharan, M.; Persmark, M.; Bortner, D.; Kole, R. Systemically Delivered Antisense Oligomers Upregulate Gene Expression in Mouse Tissues. *Nature biotechnology* **2002**, *20* (12), 1228–1233.
(58) Clark, S.; Hayes, P. SigOpt WebPage.
(59) Landrum, G. RDKit: Open-Source Cheminformatics. *RDKit: Open-source cheminformatics* **2006**.
(60) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (null), 2825–2830.
(61) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; KDD '16; Association for Computing Machinery: New York, NY, USA, 2016; pp 785–794. https://doi.org/10.1145/2939672.2939785.
(62) Head, T.; Kumar, M.; Nahrstaedt, H.; Louppe, G.; Shcherbatyi, I. *Scikit-Optimize/Scikit-Optimize*; Zenodo, 2020. https://doi.org/10.5281/zenodo.4014775.
(63) Gautam, A.; Chaudhary, K.; Kumar, R.; Sharma, A.; Kapoor, P.; Tyagi, A.; Raghava, G. P. S.; Open source drug discovery consortium. In Silico Approaches for Designing Highly Effective Cell Penetrating Peptides. *Journal of Translational Medicine* **2013**, *11* (1), 74. https://doi.org/10.1186/1479-5876-11-74.
(64) Wei, L.; Tang, J.; Zou, Q. SkipCPP-Pred: An Improved and Promising Sequence-Based Predictor for Predicting Cell-Penetrating Peptides. *BMC Genomics* **2017**, *18* (7), 742. https://doi.org/10.1186/s12864-017-4128-1.
(65) Wei, L.; Xing, P.; Su, R.; Shi, G.; Ma, Z. S.; Zou, Q. CPPred-RF: A Sequence-Based Predictor for Identifying Cell-Penetrating Peptides and Their Uptake Efficiency. *J. Proteome Res.* **2017**, *16* (5), 2044–2053. https://doi.org/10.1021/acs.jproteome.7b00019.
(66) Manavalan, B.; Subramaniyam, S.; Shin, T. H.; Kim, M. O.; Lee, G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *J. Proteome Res.* **2018**, *17* (8), 2715–2726. https://doi.org/10.1021/acs.jproteome.8b00148.
(67) Qiang, X.; Zhou, C.; Ye, X.; Du, P.; Su, R.; Wei, L. CPPred-FL: A Sequence-Based Predictor for Large-Scale Identification of Cell-Penetrating Peptides by Feature Representation Learning. *Brief Bioinform* **2020**, *21* (1), 11–23. https://doi.org/10.1093/bib/bby091.
(68) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS Cent Sci* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.

3 Deep Learning Enables Discovery of a Short Nuclear Targeting Peptide for Efficient Delivery of Antisense Oligomers

The work presented in this chapter has been reproduced from the following publication with permission from the American Chemical Society:

Lopez-Vidal, E. M.*; **Schissel, C. K.*;** Mohapatra, S.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Discovery of a Short Nuclear Targeting peptide for PMO delivery using Machine Learning. *JACS Au* **2021** (Accepted)

**These authors contributed equally to this work.*

3.1 Introduction

Delivery of therapeutic macromolecules to intracellular targets is a major challenge. For example, phosphorodiamidate morpholino oligomers (PMO, >6,500 Da) are synthetic antisense oligonucleotides that can bind strongly to nucleic acids. The first approved PMO therapy Exondys 51 (eteplirsen) was a breakthrough for treatment of Duchenne muscular dystrophy (DMD) in 2016 and acts via corrective splicing of mRNA.[1] To access its genetic target and elicit its therapeutic effect, PMO must reach the nucleus of cells in target tissues. Although a few antisense oligonucleotides have been approved by the Food and Drug Administration (FDA), a major obstacle for clinical advancement of these synthetic biopolymers is their poor cell permeability.[2] As a consequence, clinical use of PMOs without delivery vehicles requires large administered doses.[3] Traditional approaches to deliver PMO using liposomes and nanoparticles rarely advance to clinical work, often suffering from poor endosomal escape or significant toxicity.[4] Other methods of macromolecule delivery exist with varying efficiencies.[2,5–7] However, delivery of PMO by covalent attachment to cell-penetrating peptides (CPPs) has been studied broadly and has recently shown promise in clinical trials. SRP-5051, a cell-penetrating peptide attached to eteplirsen, led to higher tissue exposure, exon-skipping, and dystrophin production in patients taking a monthly dose compared to patients taking weekly doses of Exondys 51.[8]

The use of CPPs is a promising mode of macromolecule drug delivery. CPPs are short peptides of cationic, amphipathic or hydrophobic nature that facilitate intracellular delivery of cargoes that are otherwise non-cell penetrant, such as large and charged hydrophilic biomolecules.[9–12] CPPs can enhance delivery through covalent linkages or through the formation of noncovalent nanoparticle complexes in the case of negatively charged oligonucleotides.[13] The known sequence-activity motifs for CPPs largely rely on the guanidinium group of arginine (Arg).[14] While polyarginine peptides have been shown to promote cell uptake with greater efficiency than other cationic residues due to membrane affinity and difference in effective protonation,[15,16] these sequences often remain trapped in the endosomes.[17] A promising polyarginine peptide, Bpep, triggers endosomal escape by interspacing arginine residues with long alkyl 6-aminohexanoic acid and beta-alanine residues.[18] Still, one of the key challenges to clinical translation of polyarginine PMO-CPPs is their demonstrated *in vivo* toxicity caused by the peptide portion.[19,20] While a linear relationship has been shown between number of arginine residues and LD50 in mice, there is a stark difference between observed toxicity in vitro and in vivo.[19,21] It is thought that the systemic

toxicity induced by nonaarginine at 5 μmol/kg doses may be due to mast cell degranulation caused by the positive charges.[22,23] Even so, recent clinical trial results show promise for the future of PMO-CPPs in patients.[8] Although CPPs have been extensively studied, structure-activity relationships remain obscure due to inconsistent findings, making rational design of unique highly active CPPs challenging.[11] These experimental inconsistencies are due in part to the dependence of CPP performance on cargo, cell type and treatment conditions, meaning there is no universally applicable CPP. For example, studies in our laboratory have demonstrated that the cell-penetrating ability of common CPPs differs when bound to a cyanine dye versus a PMO drug, with no discernible trend.[9] Therefore, to design unique CPPs that can deliver a particular macromolecular cargo and that do not rely on arginine, that cargo must be included.

One promising method for connecting such gaps in experimental space is the use of machine learning. Recent studies using a variety of strategies have been proven useful in the quantitative activity prediction for the design of antimicrobial peptides.[24,25] Design of CPPs by machine learning has typically been limited to the use of binary classifiers because the training data are of inadequate size and from non-standardized experiments.[9,26–28] Our group has recently developed a machine learning strategy that combined a large standardized dataset with deep learning to simultaneously design nuclear-targeting miniproteins and quantitatively predict their activity.[12] This approach took advantage of the availability of a collection of hundreds of chimeric PMO-CPPs for the deep learning model training and used fingerprint representations to include unnatural amino acids aiding in cytosolic delivery[18], producing highly active miniproteins (40-80 residues). Most importantly, this machine learning framework was able to extrapolate activity data beyond the range of the training data set, suggesting that it could be leveraged to design shorter CPPs that distill sequences down to their critical motifs. Short peptides would provide advantages over longer sequences, with lower molecular weight, and are of particular interest to the broad community to enhance the internalization of macromolecules with minimal modification.[29]

Here we report the use of our deep learning model repurposed for the discovery of an 18-mer CPP that is able to deliver PMO with the same efficiency as the widely used polyarginine peptide, Bpep.[18] Our peptide, **P6**, contains only one arginine residue and improves the delivery of PMO by 25-fold compared to PMO alone. Sequence-activity studies show the reliance on an N-terminal pentalysine motif, the single arginine residue, and the C-terminal cysteine residue. Peptide **P6** has a wide therapeutic index *in vitro* and was able to deliver an anionic enzyme to the cytosol. Finally,

PMO-P6 was able to restore protein synthesis *in vivo* after a single intravenously-administered dose without evidence of kidney toxicity. These results demonstrate that we can use a machine learning model trained with long PMO-CPPs to design optimized, short CPPs (<20 residues) with minimal arginine content that are able to deliver PMO to the nucleus of cells with high efficiency and minimal toxicity in animals.

3.2 Results

3.2.1 Machine learning-based design

To generate our sequences, we repurposed the machine learning model previously developed for the design of nuclear-targeting abiotic miniproteins in order to optimize short CPPs.[12] In brief, this strategy includes generation of starting sequences, a predictor of activity, and an optimizer to improve the activity of the sequence (Fig. 3.1A). The model was trained using a chimeric library composed of 600 unique PMO-peptide conjugates, and tested in an activity-based assay where delivery of PMO to the nucleus of cells results in fluorescence, measured by flow cytometry (Fig. 3.1B-C).[30] The model optimizes sequences by increasing predicted activity while minimizing length and arginine content to mitigate potential cell toxicity.[31] The output is hundreds of de novo designed sequences containing unnatural amino acids with a broad spectrum of predicted activity. Our previous report resulted in high activity sequences reaching 40-80 residues in length. In this work we demonstrate that the machine learning strategy can be used to predict high activity, short sequences (<20 residues) for the delivery of PMO.

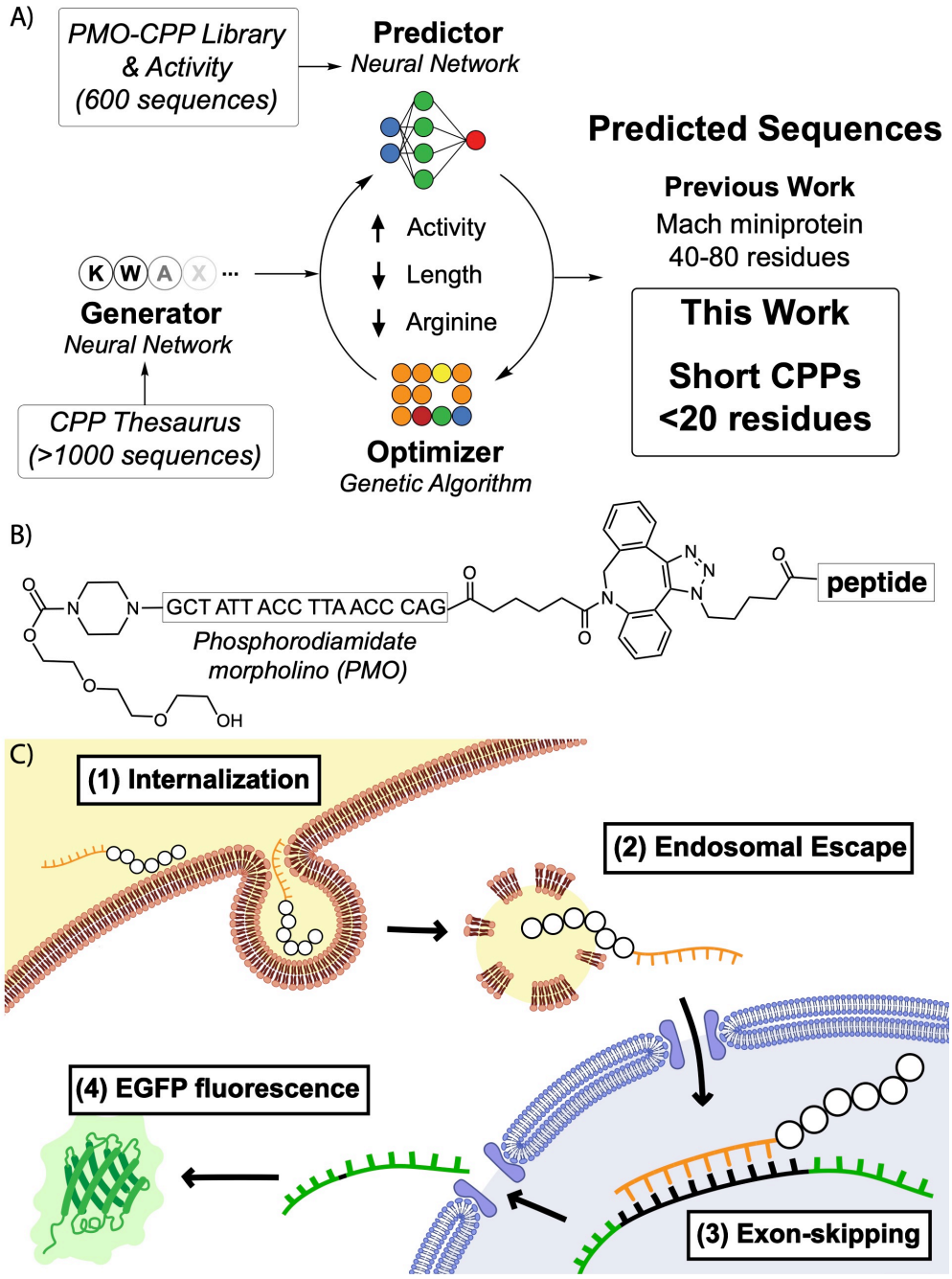


Figure 3.1 Deep learning model trained on long (>20 residues) sequences can be repurposed to predict high-activity short (<20 residues) sequences. (A) Scheme of the machine learning-based CPP design approach. A deep learning triad trained with a combinatorial library of PMO-CPPs was first developed to generate unique miniproteins. In this work, the model is adapted to predict short CPPs that retain high PMO delivery activity. (B) Structure of PMO-CPP conjugates studied. PMO IVS2-654 was attached to the N-terminus of peptides via strained alkyne-azide cycloaddition. (C) Schematic of the activity-based in vitro assay used to quantify PMO delivery activity. In order to produce fluorescence output, the PMO-peptide must enter the cell, escape the endosome, and localize to the nucleus where the PMO achieves exon-skipping activity.

3.2.2 Data augmentation improves extrapolation for shorter CPPs

Because the original model was trained using a dataset composed mostly of long (>40 residues) sequences, we used data augmentation to circumvent the sequence length imbalance, and extrapolate predicted activities in the short (<20 residues) CPP sequence space. The dataset obtained from our previous experimental studies had a mean sequence length of 41 residues, with 92% of sequences being longer than 20 residues.[9,12] In contrast, here we intended to design sequences with 20 residues or less (Fig. 3.2A). To achieve this goal, in the training set we replicated sequences with 20 residues or less 9 times, thereby showing the model 10x sequences in the desired design space (Fig. 3.2B). This augmentation ensured that the model is trained over a similar sequence space as desired, which contains shorter length from the augmented dataset and higher activity from the longer sequences already present in the dataset.[32,33] The caveat with this approach is that the diversity of sequences in the shorter search space has not changed, although the length domain has been shifted down towards shorter residues.

With less than 10% of sequences in the training dataset in the desired design space and most with low experimental mean fluorescence intensity (MFI), data augmentation slightly improved extrapolation of predicted activities for sequences with fewer than 20 residues. The predicted peptides were validated experimentally, as described below, and their experimental activities were compared to the model's predicted activity to judge the prediction accuracy (Fig. 3.2C).

We then performed retrospective in silico experiments and noted that the convolutional neural network (CNN) model performed the best amongst all tested models on the dataset of experimentally validated sequences (Table 3.1, Figs 3.3). For unbiased root mean square error (uRMSE) on the experiment dataset, we noted that simpler models had slightly better values than CNN models, but the ensemble variance denoted by the error bars in the parity plots were significantly large, thereby rendering the mean uRMSE values unreliable.

Data augmentation to artificially inflate the number of short sequences present in the dataset appeared to improve prediction of short sequences. A significantly larger training dataset, containing a greater diversity of sequences and activities among short (<20 residue) peptides, would aid the effort of enhancing predictive models to extrapolate activity prediction, as observed from a retrospective active learning study (Fig 3.5). To verify this, we trained a series of machine learning models with different data augmentation. To find the best model architecture for activity prediction, we tested 11 different model architectures, from simpler models, such as lasso and

random forest regressor, to convolutional neural networks, with both 1x and 10x data augmentation. The sequences were featurized using matrices of monomer fingerprints as previously described. For the training, we split the dataset into 60% training, 20% validation to update the weights during the training, and held out 20% of the dataset to evaluate the model post-training.[12] For robust predictions, we performed model ensembling, with a total of 25 models for each model architecture (5 random weight initializations x top 5 hyperparameter sets).

The predicted peptides were validated experimentally, and their experimental activities were compared to the model's predicted activity to judge the prediction accuracy. We noted that the CNN model indeed performed the best amongst all other models on the dataset of sequences that were experimentally tested with 0.148 R2 score and 0.512 Pearson's correlation. For uRMSE on the experiment dataset, we noted that simpler models had slightly better values than CNN models, but the error bars in the parity plots were significantly large, thereby rendering the mean uRMSE values unreliable.

10x augmented convolutional neural network (CNN) and 1x multi-layer perceptron (MLP) models performed the best amongst all augmentation and model architectures. On the test dataset, 1x MLP outperformed all models on unitless root mean squared error (uRMSE, RMSE scaled to standard deviation 1) and Pearson's correlation metrics, while 10x CNN had the highest R2 value. Given the better extrapolation ability of CNN as compared to MLP, we hypothesized that CNN would be a better choice for generation of new sequences.

Data augmentation results in worse performance of simpler models, while improving the performance of CNN models (Table 3.2, Fig 3.4). We noted that simpler model architectures such as Gaussian process and support vector regression overfit on the training data, and performed poorly on the test and experiment datasets, as noted from the tabulated metrics and significantly large error bars in the parity plots. This behavior may be attributed to the method of training, where the simpler models are trained in one-shot by showing all the data in the training set including the augmented sequences which bias the error correction during training. On the other hand, CNN models are trained by showing data in smaller batches. We hypothesize that the batched training results in weights of different layers being constrained to be updated to predict for the desired design space as they see similar datapoints across all batches.

For the CNN models, we tested different data augmentations ranging from 1x to 20x, and note that 10x was the optimal augmentation on the basis of balancing the distribution of sequences of

different lengths in the training dataset and performance on the held-out test dataset (Table 3.3, Fig 3.5). With increasing augmentation of sequences with 20 residues or less, we noted that overfitting caused the metrics in the validation dataset to improve. On the held-out test dataset, uRMSE worsened with augmentation, from 0.73 of the standard deviation of training data for 1x, to 0.75 for 10x, 0.84 for 15x, and 0.89 for 20x CNN models. For R^2 score and Pearson's correlation metrics, 10x CNN was seen to be the best model. As performance on the test dataset demarcates the performance of the model for sequences of all lengths and denotes robust training, we chose 10x CNN model for further usage. The model choice was validated by metrics noted in the experiment dataset, where the 10x CNN model was a close second for uRMSE and Pearson's correlation, and had the highest R^2 score amongst other CNN models.

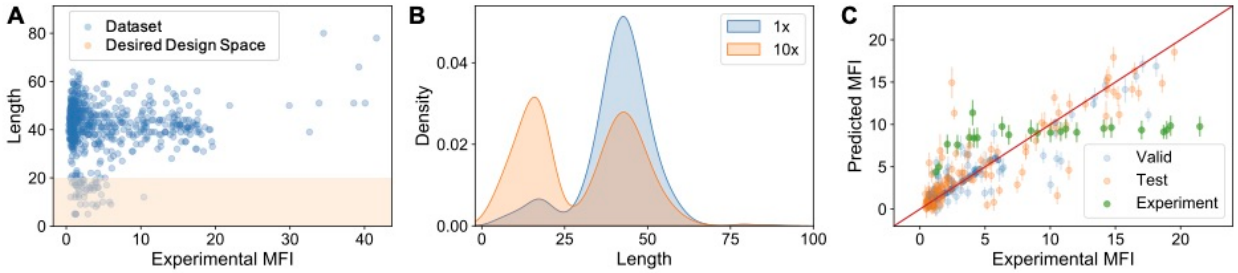


Figure 3.2 Data augmentation improves extrapolation of activity for sequences with 20 residues or less. (A) Majority of sequences in the training dataset, 601 out of 653, are longer than 20 residues (scatter plot). The desired design space of sequences with 20 or less residues is denoted in orange. (B) The distribution of lengths of sequences as-is and with 10x augmentation of sequences with 20 residues or less shows the change from a nearly normal distribution with a mean of 41 residues and a small peak at 23 residues, to a bimodal distribution. (C) The 10x CNN model ensemble predicts mean fluorescence intensity (MFI) for sequences in the validation (blue), test (orange) and experiment (green) datasets with reasonable accuracy. The points indicate the mean value of the predicted MFI from all the models in the ensemble, and the error bars denote the standard deviation of prediction.

Table 3.1 Metrics for validation, test and experiment datasets for simpler models trained with 1x augmentation, and CNN models with 1x and 10x augmentation.[a]

Model	Validation			Test			Experiment		
	uRMSE	R2	P C	uRMSE	R2	P C	uRMSE	R2	P C
CNN-1x	0.572	0.669	0.835	0.728	0.533	0.747	1.127	-0.226	0.486
CNN-10x	0.378	0.853	0.927	0.745	0.690	0.832	1.395	0.148	0.512
GP	0.099	0.990	0.995	0.672	-0.141	0.809	0.932	-9.607	0.502
Gradient Boosting	0.125	0.983	0.993	0.629	0.433	0.809	1.116	-25.518	0.336
KNN	0.341	0.842	0.943	0.672	0.263	0.779	1.404	-66.896	0.356
Lasso	0.684	-2.384	0.893	0.838	-4.971	0.759	1.047	-20.082	0.315
MLP	0.168	0.967	0.987	0.580	0.505	0.842	0.958	-6.926	0.410
Random Forest	0.222	0.937	0.979	0.692	0.222	0.762	1.098	-11.305	0.077
Ridge	0.305	0.882	0.954	0.666	0.220	0.786	1.046	-2.912	0.353
SGD	0.869	-29.192	0.832	0.970	-45.410	0.733	1.265	-1106	0.331
SVR	0.278	0.868	0.985	0.756	-1.383	0.775	0.928	-6.169	0.429
XGBoost	0.365	0.736	0.965	0.734	-0.374	0.743	1.055	-9.107	0.244

[a]Abbreviations: GP - Gaussian Process, KNN - k-nearest neighbors, MLP - multi-layer perceptron, SGD - stochastic gradient descent, SVR - support vector regression, XGBoost - extreme gradient boosting.

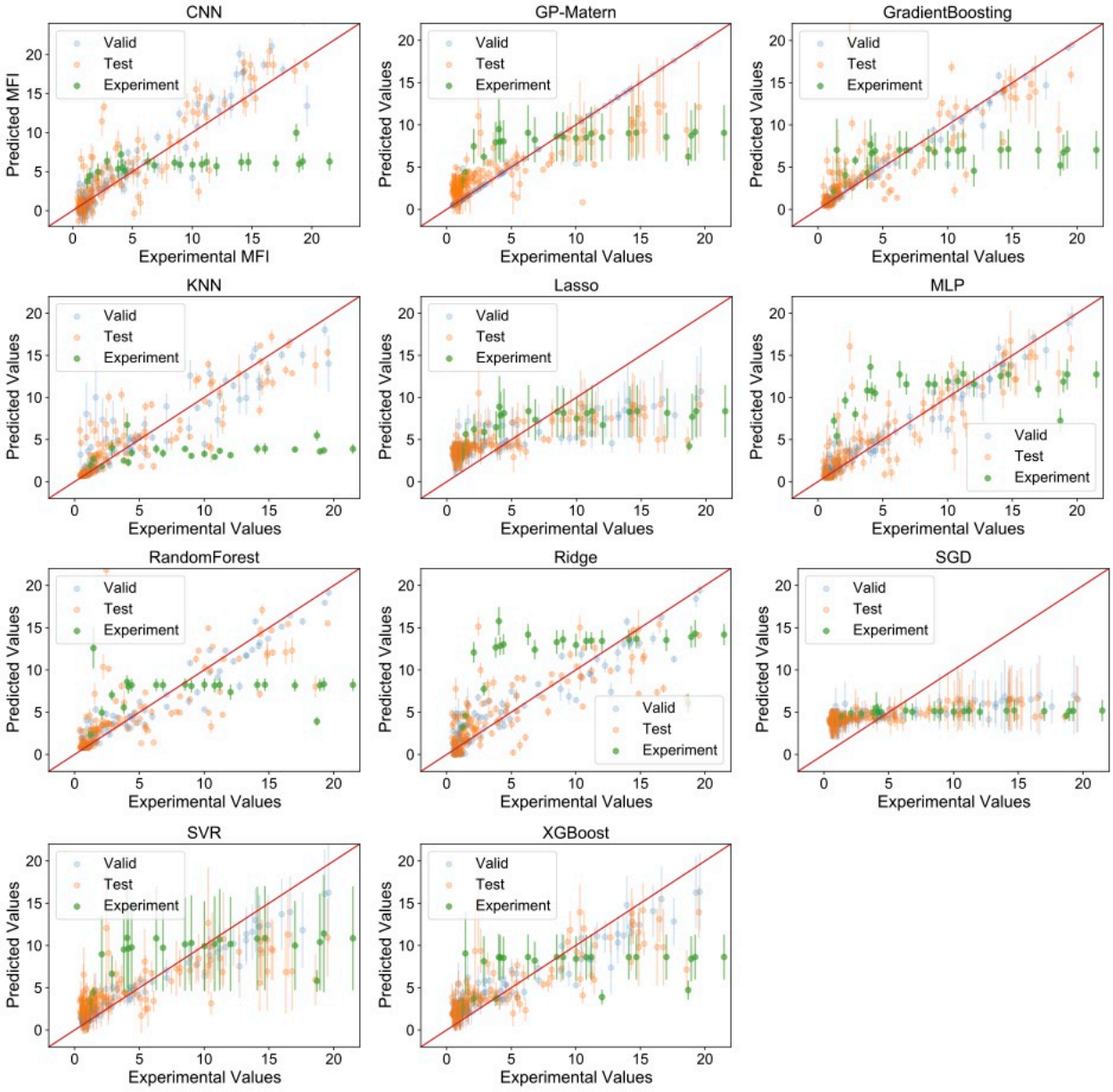


Figure 3.3 Parity plots with predicted and experimental values for validation, test and experiment datasets for simpler models and CNN models with 1x. Datapoints belonging to the validation, test and experiment sets are in different colors, as noted in the legend. The model architecture corresponding to each subplot has been noted in the title. The points indicate the mean value of the predicted MFI from all the models in the ensemble, and the error bars denote the standard deviation. Abbreviations: GP-Matern - Gaussian Process with Matern kernel, KNN - k-nearest neighbors, MLP - multi-layer perceptron, SGD - stochastic gradient descent, SVR - support vector regression, XGBoost - extreme gradient boosting.

Table 3.2 Metrics for validation, test and experiment datasets for simpler models trained with 10x augmentation, and CNN models with 1x and 10x augmentation.[a]

Model	Validation			Test			Experiment		
	uRMSE	R2	P C	uRMSE	R2	P C	uRMSE	R2	P C
CNN-1x	0.572	0.669	0.835	0.728	0.533	0.747	1.127	-0.226	0.486
CNN-10x	0.378	0.853	0.927	0.745	0.690	0.832	1.395	0.148	0.512
GP	0.094	0.991	0.995	0.814	0.075	0.816	1.184	-9.469	0.479
Gradient Boosting	0.114	0.986	0.994	0.806	0.398	0.799	1.298	-16.597	0.406
KNN	0.249	0.913	0.975	0.850	0.310	0.773	1.894	-18.284	0.105
Lasso	0.721	-4.624	0.908	1.072	-6.721	0.775	1.418	-43.690	0.411
MLP	0.096	0.990	0.995	0.730	0.562	0.838	1.211	-4.440	0.488
Random Forest	0.282	0.877	0.972	0.883	0.092	0.755	1.295	-8.578	0.466
Ridge	0.265	0.912	0.965	0.839	0.215	0.783	1.216	-3.298	0.381
SGD	0.886	-49.522	0.823	1.243	-59.090	0.704	1.645	-1434	0.487
SVR	0.286	0.848	0.988	0.957	-1.472	0.769	1.154	-5.489	0.444
XGBoost	0.356	0.733	0.971	0.931	-0.323	0.730	1.278	-7.528	0.285

[a]Abbreviations: GP - Gaussian Process, KNN - k-nearest neighbors, MLP - multi-layer perceptron, SGD - stochastic gradient descent, SVR - support vector regression, XGBoost - extreme gradient boosting.

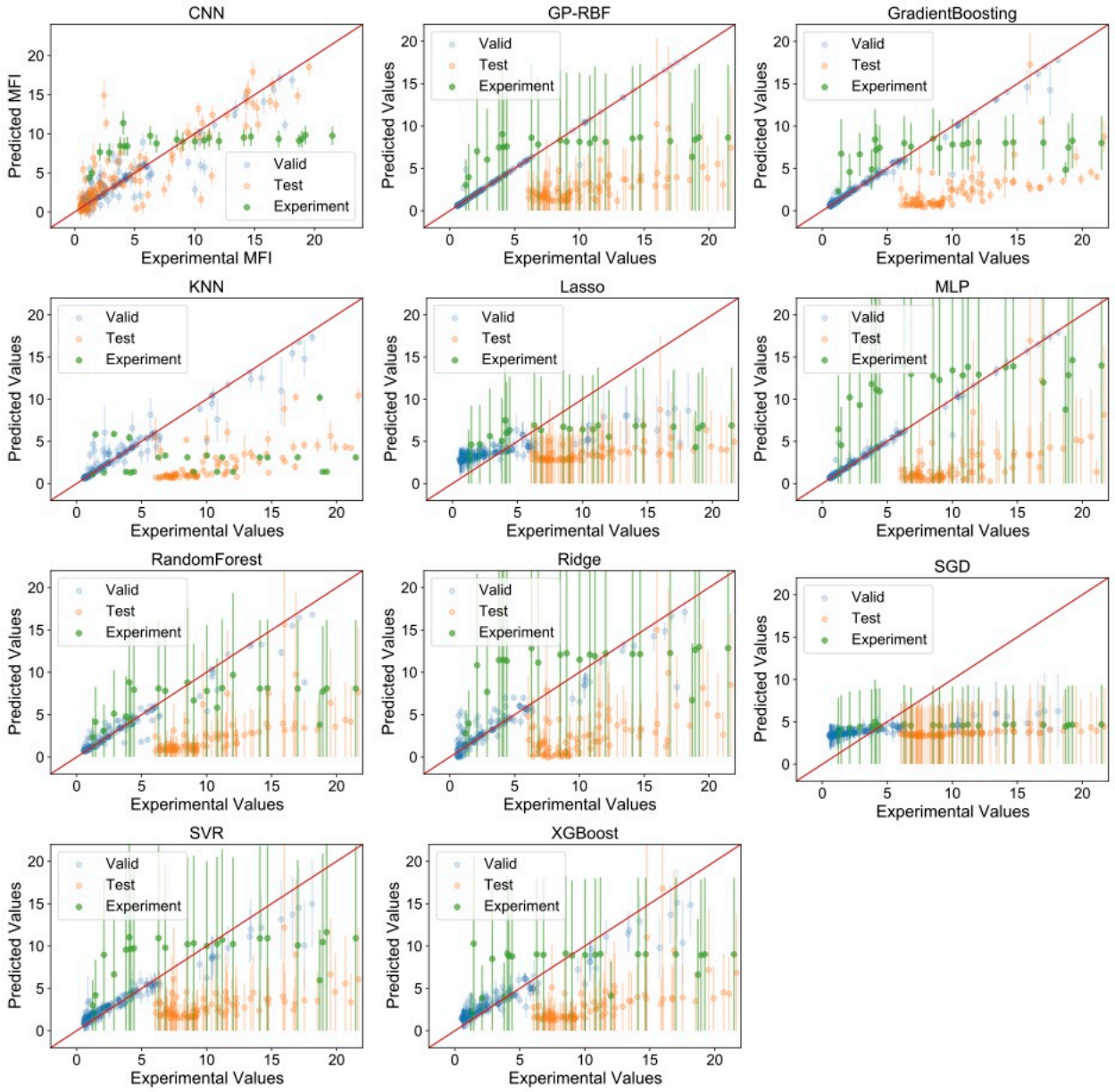


Figure 3.4 Parity plots with predicted and experimental values for validation, test and experiment datasets for simpler models and CNN models with 10x augmentation. The model architecture corresponding to each subplot has been noted in the title. The points indicate the mean value of the predicted MFI from all the models in the ensemble, and the error bars denote the standard deviation. Abbreviations: GP-RBF - Gaussian Process with radial basis function kernel, KNN - k-nearest neighbors, MLP - multi-layer perceptron, SGD - stochastic gradient descent, SVR - support vector regression, XGBoost - extreme gradient boosting.

Table 3.3 Metrics for validation, test and experiment datasets for CNN models with 1x, 10x 15x and 20x augmentation. [a]

Augment	Valid			Test			Experiment		
	uRMSE	R2	PC	uRMSE	R2	PC	uRMSE	R2	PC
1x	0.572	0.669	0.835	0.728	0.533	0.747	1.127	-0.226	0.486
10x	0.378	0.853	0.927	0.745	0.690	0.832	1.395	0.148	0.512
15x	0.418	0.867	0.935	0.837	0.673	0.823	1.739	0.111	0.499
20x	0.323	0.896	0.948	0.888	0.681	0.827	1.983	0.121	0.562

[a] The top 2 metrics for each column are noted in red and blue, respectively. 10x augmentation performs really well for test and experiment datasets. This augmentation is also noted to be the most optimal choice based on the overfitting on the train and valid datasets for 20x augmentation.

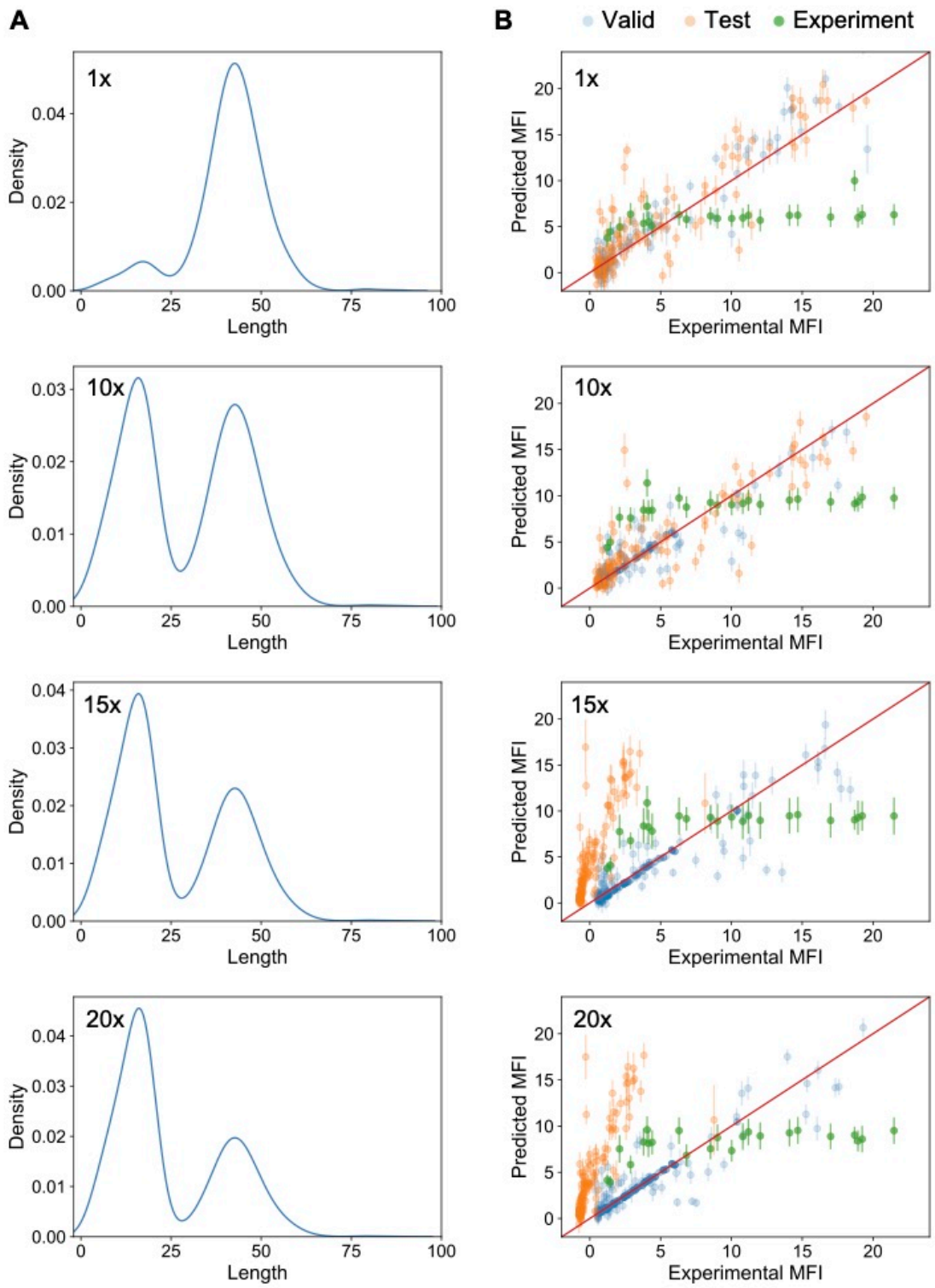


Figure 3.5 Performance of CNN models trained with different data augmentation.(A) The distribution of lengths of sequences for training datasets with different data augmentations. (B) Parity plots with predicted and experimental MFI values for validation, test and experiment datasets for CNN models with different data augmentation. The data augmentation has been noted in the respective plots. Datapoints belonging to the valid, test and experiment sets are in different colors, as noted in the legend.

We also evaluated retrospectively how active learning would impact prediction accuracy for the new sequences. To start with, we had 653 sequences in the training dataset and 19 sequences, with 20 residues or less, which were experimentally validated. For the first generation of active learning models, we added 1 sequence to the training data set and tested on 18. Subsequently, in each generation, we added 3 sequences, with (4, 15), (7, 12), and ultimately (16, 3) in the (train, test) data sets. For each generation, we used the top 5 model architectures from the 10x augmentation run, and followed a similar training methodology in terms of hyperparameters and random seeds for weight initialization. Given that we had an external test set, we trained with 80:20 train:validation set. To simulate a real-world scenario, all sequences from the previous generation were carried on to the next generation, and 3 random sequences were added to the training data set in each generation. We performed 3 such active learning simulations, varying the order in which the sequences are added to the train set, and averaged the results across all of them.

The test set performance improved every time new sequences were added to the training data set. However, the performance on the validation data set worsened slightly towards the end, owing to significantly more short CPP sequences than longer sequences, similar to results observed for 20x augmentation. The improving model accuracy on the test data set indicates that active learning with new sequences can help in sequential optimization of the model to predict sequences in desired design space (Fig 3.6).

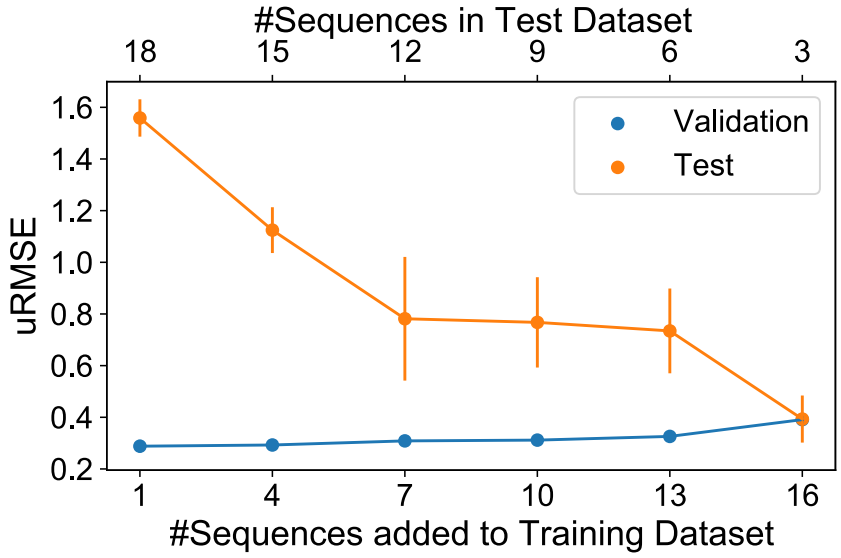


Figure 3.6 Model performance on the held out test dataset measured as uRMSE of short CPP sequences improves as more short sequences are added to the training dataset.

3.2.3 Validation of predictions

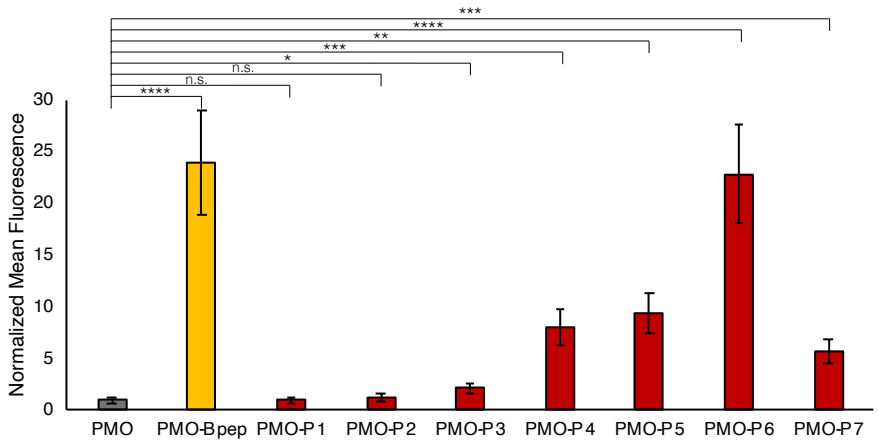
We had the model generate the highest predicted activity sequences with increasing lengths (from 5 to 20 residues), from which we selected thirteen sequences to analyze and seven to validate experimentally (Fig. 3.7). These sequences contain 0 or 1 arginine residues and have predicted activities ranging from 4 to 11-fold increase with respect to unconjugated PMO. Of note, the sequences were cationic and contained several lysine residues as well as the extended alkyl backbone unnatural residues beta-alanine (B) and 6-aminohexanoic acid (X). A clear trend of the predicted activity increasing with length can be seen. Notably, the predicted sequences contained similar common motifs, indicating that the model consistently identifies these motifs as potential drivers of high predicted activity. We then selected seven of these sequences to validate experimentally, by ranking them by length and selecting every other sequence (**P1**-**P7**, Fig 3.7B). These sequences were synthesized via semi-automated fast-flow solid-phase peptide synthesis (SPPS) and coupled with 5-azidopentanoic acid to the N-terminus before cleavage and purification as previously reported. Separately, dibenzocyclooctyne (DBCO) was coupled to the 3' end of PMO anti-IVS2-654 before the two constructs were combined into the PMO-CPP conjugate via strain-promoted azide-alkyne cycloaddition.

PMO-CPP constructs were then validated with the HeLa 654 enhanced green fluorescent protein (EGFP) assay as previously described (Figure 1C).[9,10,12] HeLa 654 are cells stably transfected with an EGFP coding sequence interrupted by an intron from the human β-globin gene (IVS2-654) containing a mutation that alters the normal pre-mRNA splice site. The change in splicing leads to retention of an unnatural mRNA fragment in the spliced EGFP mRNA and the translation of a nonfluorescent form of EGFP. The PMO anti-IVS2-654 hybridizes to the mutant β-globin exon in the stably transfected HeLa cells, altering gene splicing and leading to full-length EGFP expression. The amount of PMO delivered is therefore correlated to the amount of functional EGFP expressed, quantified via flow cytometry.[34] This assay directly informs on the amount of PMO that is internalized into the nucleus of these cells. As positive control we used PMO conjugated to Bpep, a high-arginine content peptide with the sequence RXRRBRRXRRBR studied extensively for PMO delivery.[18]

The selected constructs exhibited a length-dependent increase in activity up to the 18-mer, **P6**. Analysis of the exon skipping activity measured in the EGFP assay performed with three different biological replicates (Fig. 3.7A, Fig. 3.8-3.10) indicated that **PMO-P6** is the conjugate that

displays the highest activity, with a 23-fold increase relative to unconjugated PMO. This activity is comparable to the activity of the standard conjugate **PMO-Bpep** (24-fold increase) while containing only 1 arginine residue. When compared to the library dataset, it is clear that P6 stands out as being shorter with fewer arginine residues, while having greater activity than any sequence in the library dataset (Figure 3.7C). While having minimal arginine residues, the predicted sequences are still polycationic and all contain an oligo-lysine N-terminus. The second and third best performers among the conjugates of PMO with these 7 predicted peptides were **PMO-P4** and **PMO-P5** with 8-fold and 9-fold increase in activity relative to unconjugated PMO, respectively. The rest of the conjugates showed a 6-fold increase or lower. Interestingly, we observed a similar trend to the predicted activity, of increasing activity with increasing length, with the exception of the 20-mer **PMO-P7** which had moderate 6-fold activity. This observation suggests that while machine learning models have the potential to predict high activity sequences, random errors leading to false positives are common.

A)



B)

Peptide	Sequence	# Arg Residues	Length	Net charge	Predicted Activity	Experimental Activity
PMO-Bpep	RXRRBRRXRRBR	8	12	8	9	24
PMO-P1	KXKXT	0	5	2	4	1
-	KXKHQQGP	0	8	3	4	-
PMO-P2	KXKHQQQXK	0	9	4	5	1
-	KXKHQQGKKT	0	10	5	5	-
PMO-P3	KXKKQQGKKKHR	1	12	8	8	2
-	HKKKQQGKKKHRW	1	13	9	8	-
PMO-P4	HKKKKQBKKKHRWP	1	14	10	9	8
-	KKKKKQBKKKHRWPM	1	15	10	9	-
PMO-P5	KKKKKQBKKKHRWPMG	1	16	10	9	9
-	KKKKKQGKKHRWPMGG	1	17	9	5	-
PMO-P6	**KKKKKQBKKKHRWPKXXC**	**1**	**18**	**11**	**10**	**23**
-	KKKKNQBKKKHRWPMKXCP	1	19	10	10	-
PMO-P7	KKKKNQBKKKHRWPMKXCPQ	1	20	10	11	6

C)

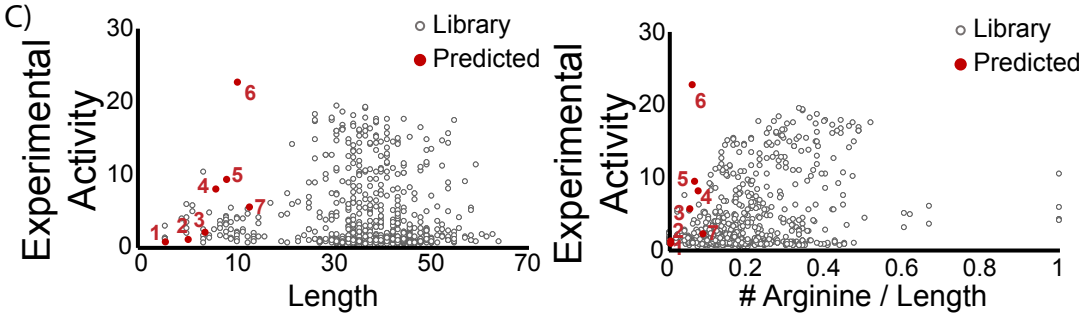


Figure 3.7 Predicted PMO-peptide conjugates enhance PMO delivery in a length- and sequence-dependent manner. (A) Delivery activities of predicted PMO-peptide conjugates **PMO-P1** to **PMO-P7** compared to **PMO-Bpep**, measured by the EGFP assay and calculated as fluorescence relative to PMO alone. (B) List of predicted peptides. Sequences selected represent those with the highest predicted activity of each length. Peptides selected for experimental validation are named **P1-P7** and were attached to PMO IVS2-654 and tested in the EGFP assay. Bpep was used as a positive control for comparison. Predicted and experimental activities are shown as relative to PMO alone. (C) Experimental PMO delivery activity of the library peptides (grey) and validated predicted peptides (red) versus length and arginine content relative to length. Bars in (A) represent group mean ± SD, n = 3. Each sample was measured at a concentration of 5 μM. Shown here is one representative biological replicate, which was repeated twice with similar results, as shown in the Supporting Information. (*p<0.01, **p<0.001, ***p<0.0001, ****p<0.00001 compared to PMO. n.s. = not significant as determined by student's unpaired two-tailed t-test).

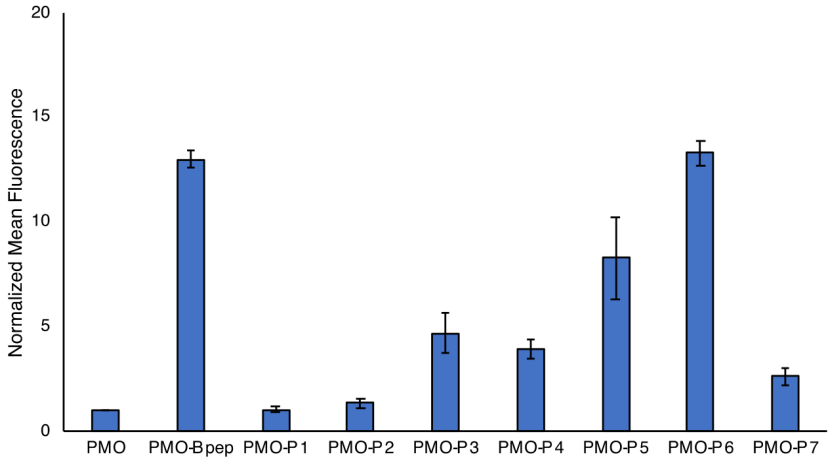


Figure 3.8 Biological replicate 1 of Activity (eGFP assay) of the PMO-P1- P7 measured in one biological replicate at a concentration of 5 µM for each PMO-peptide conjugate. The eGFP fluorescence was normalized with respect to the PMO alone condition.

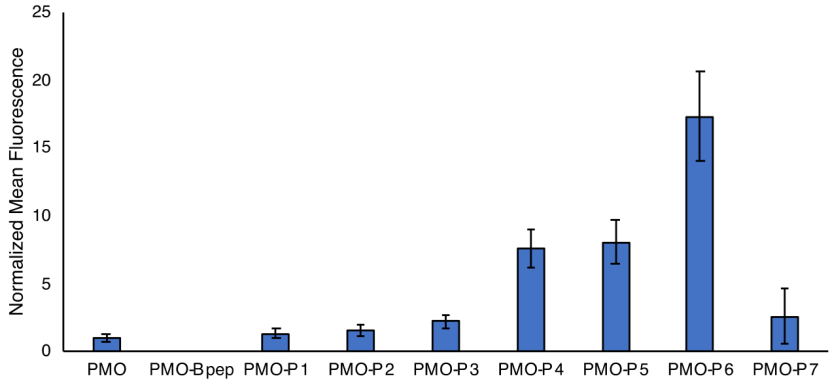


Figure 3.9 Biological replicate 2 of Activity (eGFP assay) of the PMO- P1- P7 measured in one biological replicate at a concentration of 5 µM for each PMO-peptide conjugate. The eGFP fluorescence was normalized with respect to the PMO alone condition.

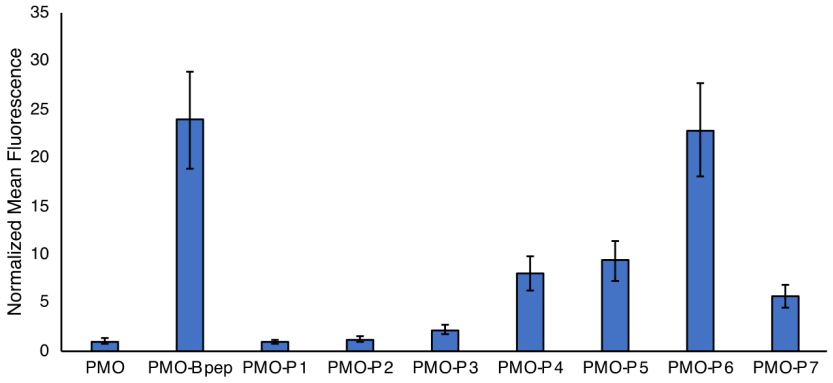
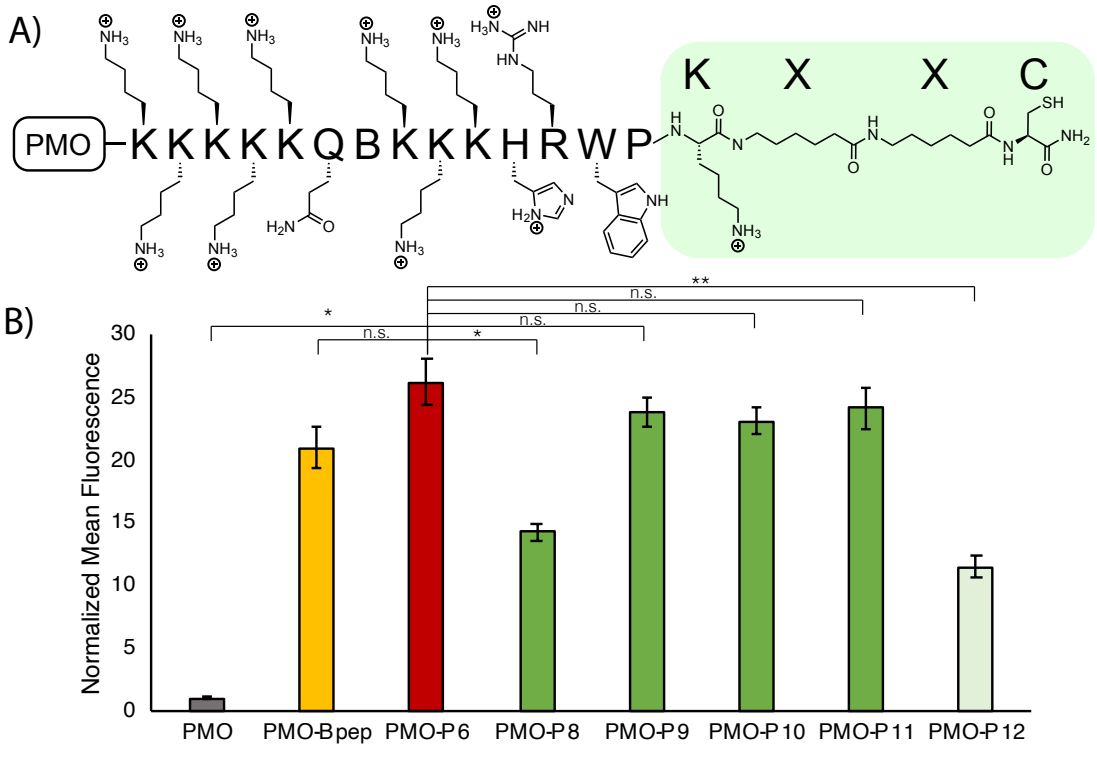


Figure 3.10 Biological replicate 3 of Activity (eGFP assay) of the PMO-P1- P7 measured in one biological replicate at a concentration of 5 µM for each PMO-peptide conjugate. The eGFP fluorescence was normalized with respect to the PMO alone condition.

3.2.4 Alanine mutations reveal sequence-activity relationships

We employed alanine scanning to reveal that the C-terminal cysteine residue of **P6** is partially responsible for the peptide's high experimental activity. The sequences of **P6** and **P5**, the best and second-best performers, respectively, are identical except for a few amino acids located towards the C-terminus; **P6** contains a KXXC motif while **P5** contains a MG motif. Despite this similarity, **P6** has a nearly 3-fold higher activity than **P5.** We hypothesized that the presence of the KXXC motif, or specific residues within this motif, could be responsible for the activity of **P6**. To test this hypothesis, we substituted each one of the residues in the KXXC motif with alanine residues, as well as deleted the KXXC motif completely (**P8-P12** in Fig 3.11). After conjugation of each **P6** analog to PMO we carried out the Hela 654 EGFP assay for activity profiling of PMO delivery. Activities of **PMO-P8** to **PMO-P11** demonstrated that of the alanine mutations of the different positions in the KXXC motif, only the mutation of cysteine at the final position led to a significant, but not complete, decrease in activity. This observation suggests that while not solely responsible for the high activity observed for **PMO-P6,** the C-terminal cysteine contributes to uptake. Still curious about the presence of the KXXC motif, which is not found in the lower-performing **PMO-P5**, we made a truncated version of **P6** lacking the KXXC C-terminus (**PMO-P12**, Fig 3.11). We observed a decrease in activity of the truncated sequence similar to that of **PMO-P8** and only slightly higher than **PMO-P5**, indicating again that the cysteine residue is the most critical C-terminal residue.



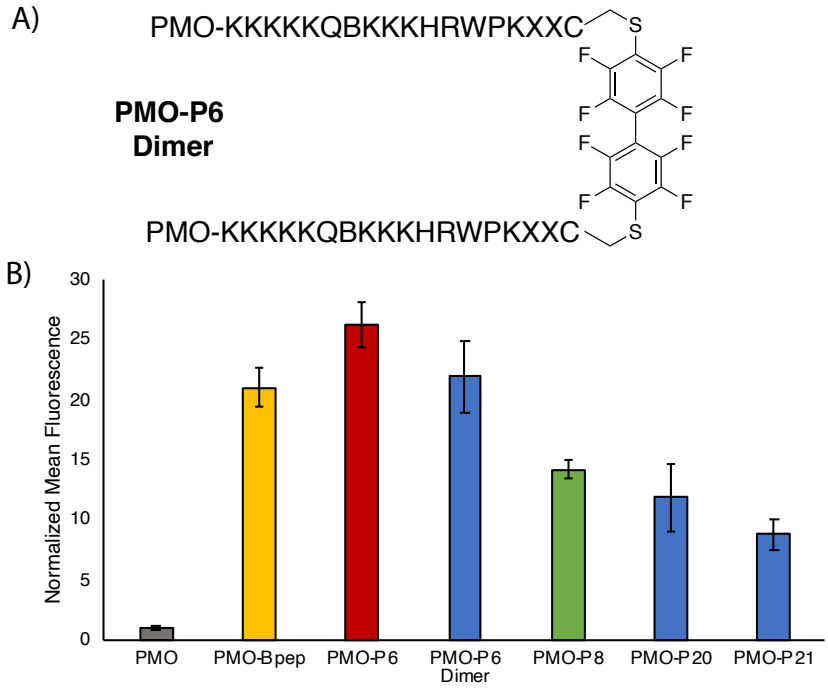
C)

Peptide	Sequence	Net Charge	Activity
PMO-Bpep	RXRRBRRXRRBR	8	21
PMO-P6	KKKKKQBKKKHRWPKXXC	11	26
PMO-P8	KKKKKQBKKKHRWPKXX**A**	11	14
PMO-P9	KKKKKQBKKKHRWPKX**A**C	11	24
PMO-P10	KKKKKQBKKKHRWPK**A**XC	11	23
PMO-P11	KKKKKQBKKKHRWP**A**XXC	10	24
PMO-P12	KKKKKQBKKKHRWP	10	11

Figure 3.11 Cys-containing aminohexanoic acid C-terminus has mild effect on activity of PMO-P6. (A) The construct **P6**, highlighting the KXXC C-terminal motif. B = beta-alanine, X = 6-aminohexanoic acid. (B) Shown are activities of various mutated sequences as determined by EGFP assay and reported as fluorescence relative to PMO alone. Shown is a representative experiment, with biological replicates shown in the Supporting Information. (C) Table of Ala-mutated sequences corresponding to activities shown in B) as well as net charge of each sequence and activity relative to PMO alone. Bars represent group mean ± SD (n = 3, *p < 0.01, **p < 0.001, compared to **PMO-P6**; n.s. = not significant; determined using unpaired two-tailed student's t-test).

To further examine the role of the C-terminal cysteine residue, we made several analogs of P6. First, we made a dimeric form to mimic disulfide formation by conjugating two copies of P6 through a S_NAr reaction using decafluorobiphenyl (Fig 3.12). Synthesis of the deca-linked P6 dimer was carried out as previously described.[35] Next, we mutated the C-terminal cysteine residue to a methionine (P20) or deleted the residue altogether (P21) and compared the activities to the C18A mutant, P8. These analogs were then evaluated using the HeLa 654 assay and their activities were normalized to PMO alone. We found that the activity of the P6 dimer was similar to that of P6, whereas the activities of P20 and P21 were similar to

that of P8. This suggests that it is possible that the disulfide dimer is the active form of P6 once inside cells, although thiol-mediated uptake could still be playing a role as well. Clearly the analogs which lack the C-terminal cysteine exhibit a decrease, but not complete ablation, of delivery activity. Therefore, the thiol residue is necessary but not sufficient for the observed activity.



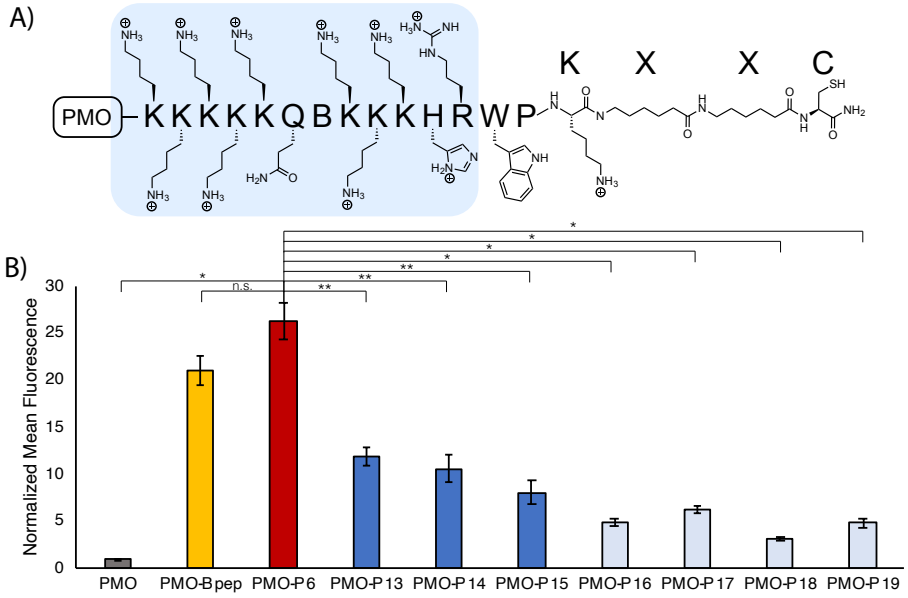
C)

Peptide	Sequence	Net Charge	Activity
PMO-Bpep	RXRRBRRXRRBR	8	24
PMO-P6	KKKKKQBKKKHRWPKXXC	11	26
PMO-P6 Dimer	KKKKKQBKKKHRWPKXXC-(deca)-CXXKPWRHKKKBQKKKK	22	22
PMO-P8	KKKKKQBKKKHRWPKXXA	11	14
PMO-P20	KKKKKQBKKKHRWPKXXM	11	12
PMO-P21	KKKKKQBKKKHRWPKXX	11	9

Figure 3.12 Evaluation of role of C-terminal cysteine residue. (A) Structure of PMO-P6 dimer, where C-terminal cysteine residues are connected via decafluorobiphenyl. (B) Shown are activities of various mutated sequences as determined by EGFP assay and reported as fluorescence relative to PMO alone. (C) Table of sequences corresponding to activities shown in B) as well as net charge of each sequence and activity relative to PMO alone. Bars represent group mean ± SD, N = 3.

We then hypothesized that the cationic motifs may be responsible for the enhanced activity observed for **PMO-P6,** including the pentalysine chain present at the N-terminus, or the single arginine residue. Therefore, we substituted one, two or three lysine for alanine residues, as well as the single arginine residue (**P13-P19**, Fig 3.13-3.15). We observed significant impact on activity when lysine residues within the N-terminal half were substituted with alanine residues. Mutation of more than one lysine within the motif reduced the activity to near 5-fold (**PMO-P16** to **PMO-P18**) as did mutation of the only arginine residue (**PMO-P19**). Together, these

observations suggest that while no single residue is responsible for the high activity of **P6**, the cationic residues may contribute the most to the activity. Nonetheless, the dramatic difference in activity between **P5** and **P6** despite their cationic sequence similarity suggests that the high activity of **P6** is dependent on its unique sequence.



C)

Peptide	Sequence	Net Charge	Activity
PMO-Bpep	RXRRBRRXRRBR	8	21
PMO-P6	KKKKKQBKKKHRWPKXXC	11	26
PMO-P13	K**A**KKKQBKKKHRWPKXXC	10	12
PMO-P14	KKKKKQBK**A**KHRWPKXXC	10	11
PMO-P15	KKK**A**KQBKKKHRWPKXXC	10	8
PMO-P16	K**A**KKKQBK**A**KHRWPKXXC	9	5
PMO-P17	K**A**K**A**KQBKKKHRWPKXXC	9	6
PMO-P18	K**A**K**A**KQBK**A**KHRWPKXXC	8	3
PMO-P19	KKKKKQBKKKH**A**WPKXXC	10	5

Figure 3.13 Sequence-activity studies reveal dependence of P6 PMO delivery efficacy on the Lysine residues. (A) The construct **P6**, highlighting cationic N-terminal motif. B = beta-alanine, X = 6-aminohexanoic acid. (B) Shown are activities of various mutated sequences as determined by EGFP assay and reported as fluorescence relative to PMO alone. Shown is a representative assay, with biological replicates shown in the Supporting Information. (C) Table of Ala-mutated sequences corresponding to activities shown in (B) as well as net charge of each sequence and activity relative to PMO alone. Bars represent group mean ± SD (n = 3, *p < 0.01, **p < 0.001, compared to **PMO-P6**; n.s. = not significant as determined by unpaired two-tailed student's t-test).

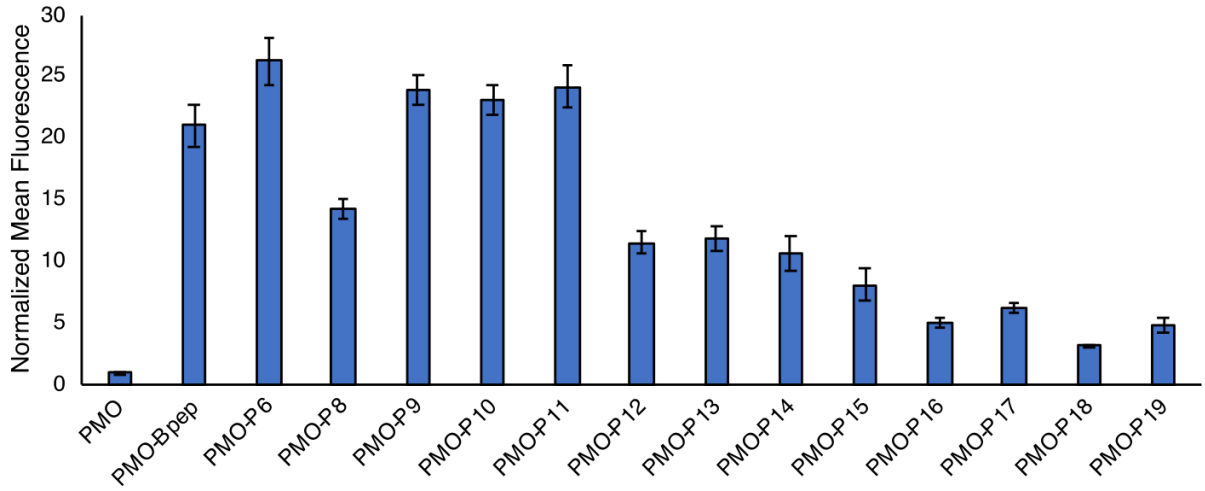


Figure 3.14 Biological replicate 1 of Activity (eGFP assay) of the PMO-P6- P19 measured in one biological replicate at a concentration of 5 µM for each PMO-peptide conjugate. The eGFP fluorescence was normalized with respect to the PMO alone condition.

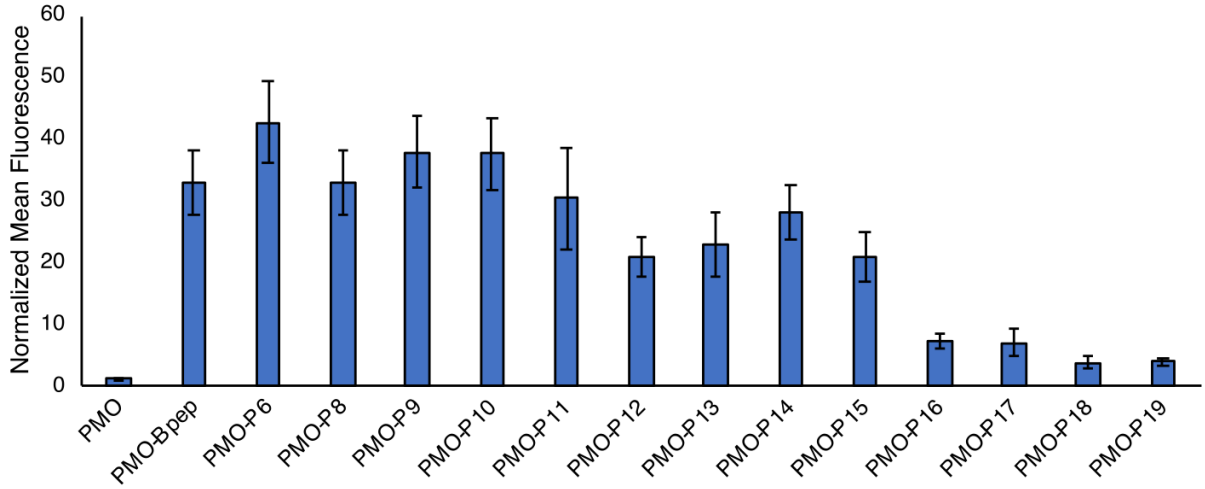


Figure 3.15 Biological replicate 2 of Activity (eGFP assay) of the PMO-P6- P19 measured in one biological replicate at a concentration of 5 µM for each PMO-peptide conjugate. The eGFP fluorescence was normalized with respect to the PMO alone condition.

3.2.5 Concentration response activity and toxicity

PMO-P6 and its truncated derivative, **PMO-P12**, have concentration-dependent activities while remaining nontoxic *in vitro*. We investigated these two sequences further by measuring concentration-response for both activity and toxicity *in vitro* (Figure 3.16A-B). The half maximal effective concentration (EC50) was calculated by measuring the EGFP fluorescence in HeLa 654 cells of each conjugate along a range of concentrations (between 0.1 and 100 µM). The EC50 value of **PMO-P6** was 4 µM, and activity ultimately reached 40-fold over PMO at 25 µM. The EC50 of

PMO-P12 was 7 μM, and activity ultimately reached 30-fold at 50 μM. Therefore **PMO-P6** remained slightly more active than its truncated derivative **PMO-P12**.

We then measured concentration-response of toxicity for **PMO-P6** and **PMO-P12**. Membrane toxicity was assessed using the lactate dehydrogenase (LDH) release assay, where cytosolic LDH detected in the cell supernatant indicates membrane permeabilization. Toxicity is shown as a percentage relative to cells fully lysed with the detergent sodium dodecyl sulfate (SDS). We first assayed the cell supernatant of the HeLa 654 cells used in the concentration-response activity assay and found no toxicity at the concentrations tested, indicating that the observed delivery activity was not due to membrane permeabilization (Fig 3.17). We then analyzed a wider concentration range with the LDH assay in renal epithelial cells (TH1 RPTEC), since polycationic sequences sometimes result in renal toxicity in vivo.[36] We tested concentration ranges of 1 and 200 μM of **PMO-P6** and between 5 and 400 μM of **PMO-P12** (Fig. 3.16A-B). Median lethal concentration (LC50) for **PMO-P6** was observed at 100 μM, and at concentrations under 50 μM the toxicity is negligible. LC50 for **PMO-P12** was 220 μM, and at concentrations below 100 μM the toxicity is negligible. These data demonstrate a large concentration window between activity and toxicity of these constructs *in vitro*, with **PMO-P12** having a slightly larger in vitro window of 30, compared to 25 for **PMO-P6**. Elimination of the KXXC motif at the C-terminus of **PMO-P6** slightly decreases activity, indicating that while **PMO-P12** is less toxic, the machine learning-designed **PMO-P6** is more effective.

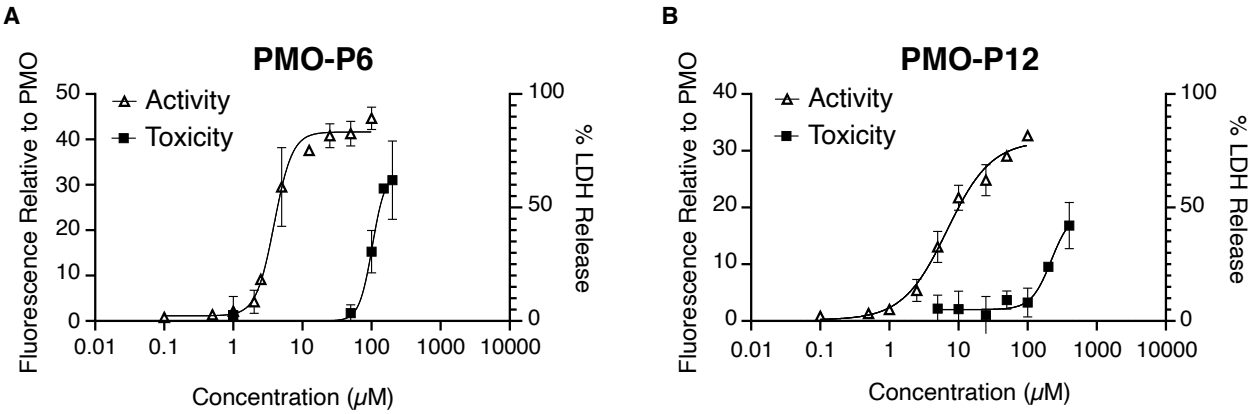


Figure 3.16 PMO-P6 and PMO-P12 induce concentration-dependent activity and have low toxicity in vitro. Activity was determined using the EGFP assay at varying concentrations in HeLa 654 cells and is represented as fluorescence relative to PMO alone. Toxicity was determined using the LDH release assay at varying concentrations in TH1 RPTEC cells, and is represented as % LDH release relative to cells fully lysed with SDS. (A) For **PMO-P6** EC50 = 4 μM and LC50 = 100 μM. (B) For **PMO-P12** EC50 = 7 μM and LC50 = 220 μM. Each point represents group mean ± SD, n = 6.

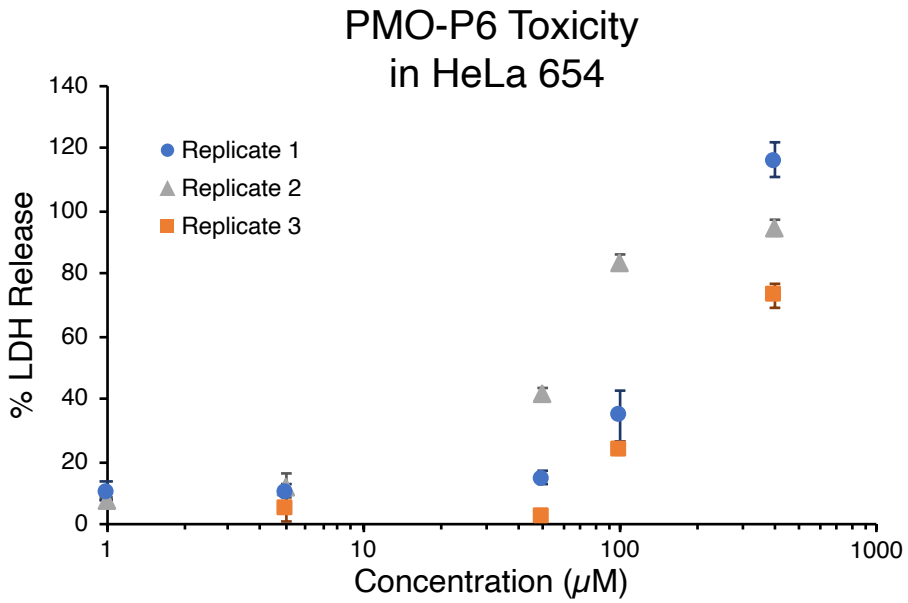


Figure 3.17 Dose-response toxicity of PMO-P6 in HeLa 654 cells. Each color corresponds to individual biological replicate.

3.2.6 Mechanism of uptake

We then investigated the mechanism of cell entry for **PMO-P6** and **PMO-P12** using chemical endocytosis inhibitors. For uptake of PMO-CPPs at physiologically-relevant concentrations the primary mechanism of internalization is energy-dependent and multiple endocytic mechanisms can take place.[10,12] However, there is no main mechanism for cell entry of peptides or PMO-peptide conjugates, and activity is highly dependent on the treatment concentrations, cell line and cargo attached. Therefore, we investigated the mechanism of internalization of **PMO-P6** and **PMO-P12** in HeLa 654 cells using a panel of chemical endocytosis inhibitors: chlorpromazine, cytochalasin D, wortmannin, EIPA and Dynasore. Chlorpromazine inhibits clathrin-mediated endocytosis by preventing the assembly and disassembly of clathrin lattices on cell surfaces and on endosomes. Cytochalasin D inhibits phagocytosis and macropinocytosis; wortmannin alters various endocytosis pathways by inhibiting phosphatidylinositol kinases; 5-(*N*-ethyl-*N*-isopropyl) amiloride (EIPA) inhibits macropinocytosis; and Dynasore, a dynamin inhibitor which also inhibits clathrin-mediated endocytosis.[37]

The mechanism of cell uptake of **PMO-P6** and **PMO-P12** is possibly endocytosis. The experiment was conducted in a pulse-chase format in which HeLa 654 cells were preincubated with the inhibitors. After 30 min, **PMO-P6** or **PMO-P12** were added to each well at a concentration of 5 μM. After incubation at 37 °C for 3 hours the treatment media was replaced

with fresh media and the cells were allowed to grow for another 22 hours at 37 °C. Sample preparation and flow cytometry were then performed as usual. The activities of **PMO-P6** and **PMO-P12** were reduced by treatment with chlorpromazine, cytochalasin D and wortmannin. Chlorpromazine reduced activity in a concentration-dependent manner, suggesting clathrin-mediated endocytosis may be the dominant uptake mechanism (Fig. 3.18). A similar effect was observed with cytochalasin D and wortmannin, although the higher concentrations of these inhibitors appeared to be toxic to HeLa 654 cells. As observed in previous studies, treatment with EIPA increased activity, indicating that uptake doesn't primarily occur through macropinocytosis and blocking this pathway may increase the uptake via the favored pathway. These results indicate that **PMO-P6** and **PMO-P12** are entering HeLa 654 cells via similar mechanisms, likely receptor-mediated endocytosis.

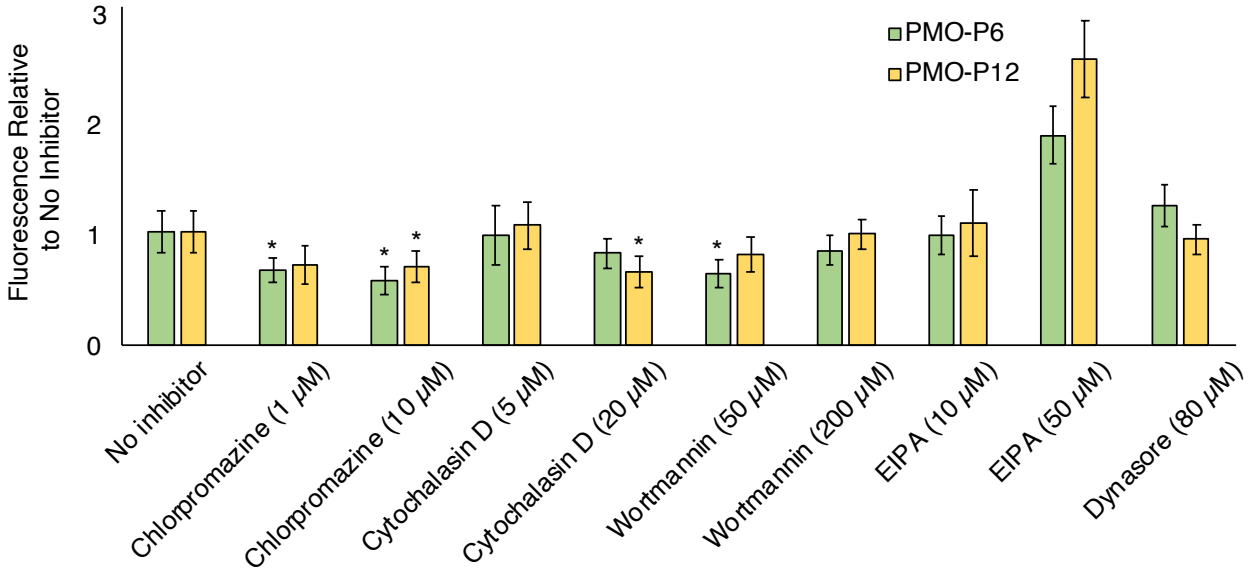


Figure 3.18 PMO-P6 and PMO-P12 enter cells via similar mechanisms. Graph shows the effect of endocytosis inhibitors in PMO delivery using **PMO-P6** and **PMO-P12**. HeLa 654 cells were treated with each inhibitor in a pulse-chase experiment followed by treatment with 5 μM **PMO-P6** or **PMO-P12**. After treatment, the mean fluorescence intensity of EGFP was measured by flow cytometry and normalized with respect to the sample without inhibitor. Bars represent group mean ± SD, n = 3 (*$p < 0.05$, determined using unpaired two-tailed student's t-test).

3.2.7 P6 can deliver other macromolecules to the cytosol of cells

In addition to delivery of antisense oligonucleotides, we found that **P6** can also deliver an anionic, active enzyme. Selected as a model protein, diphtheria toxin A (DTA) is a 21 kDa anionic protein segment of diphtheria toxin containing the catalytic domain of the toxin but lacking the portions for cell entry.[38] Once inside the cytosol, this toxin inhibits protein synthesis and kills the cell. Therefore, its delivery can be measured with a cell proliferation inhibition assay using CellTiter-Glo, which quantifies ATP, a key biomarker of cell viability. **P6-LPSTGG** was synthesized via SPPS, **Gly_5-DTA** was produced via recombinant expression, and the two were combined by sortase A-mediated ligation. **P6-DTA** was incubated with HeLa cells at varying concentrations for 72 h. We found that **P6-DTA** constructs were delivered into the cell cytosol significantly more efficiently (EC50 = 1 μM) than protein alone (EC50 = 4 μM) (Fig. 3.19). Enhanced delivery of this anionic protein suggests that **P6** may be a broadly applicable modification for macromolecule delivery.

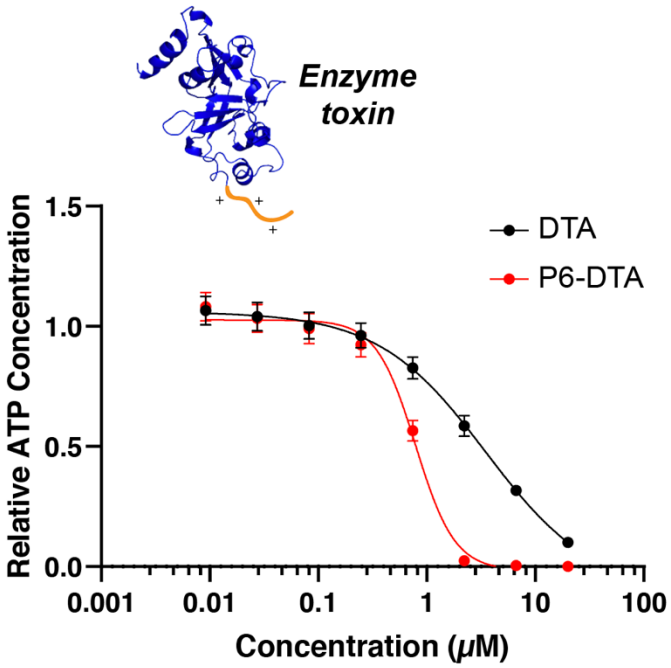


Figure 3.19 P6 delivers active enzyme to the cell cytosol. Shown is a concentration-response curve after HeLa cells were treated with DTA or **P6-DTA** for 72 h. Delivery of the active toxin to the cytosol results in toxicity as measured by the CellTiter-Glo assay, which quantifies adenosine triphosphate (ATP). Each point represents group mean ± SD, n = 3 distinct samples.

3.2.8 In vivo studies

We then found that **PMO-P6** was able to correct aberrant splicing in vivo. We chose for our studies transgenic mice containing the same gene target as the HeLa 654 cells, enabling a direct transition of the PMO-peptide from *in vitro* to *in vivo* studies. Mice were injected with a single dose of **PMO-P6** at three different concentrations (10, 30 and 60 mg/kg). Seven days later, serum and tissues including quadriceps, diaphragm and heart were harvested. We observed a dose-dependent increase in EGFP fluorescence in each muscle tissue examined, including heart (Fig. 3.20A-C). In the case of Duchenne muscular dystrophy, therapies that access cardiac tissue are critical for treatment because the primary cause of death in DMD patients is cardiac arrest. Therefore, the activity observed for **PMO-P6** in the heart is particularly advantageous. Notably, EGFP levels measured in the heart were quite variable between individual mice treated with 30 or 60 mg/kg of **PMO-P6**, whereas the samples from other tissues did not display as high a variability. A possible explanation is the differential oxidative environment in the heart compared to other tissues affecting the local oxidative state of our Cys-containing PMO-peptide, which is known to affect uptake of Cys-containing peptides.[39] Nonetheless, despite this observed variability, **PMO-P6** represents a short, low-arginine content peptide able to deliver PMO to the heart and muscles of mice *in vivo*.

At the same time, we found that **PMO-P6** does not induce kidney toxicity in the mice at the tested doses. *In vivo* use of polycationic CPPs can cause kidney toxicity.[36] We analyzed kidney toxicity by measuring kidney biomarkers isolated from post-treatment serum, specifically blood urea nitrogen (BUN), creatinine and cystatin C. No changes were found in the levels of these three renal function markers 7 days post-treatment which indicates that **PMO-P6** is not toxic to the kidney at the doses tested (Fig. 3.20D-F).

A key reason for the design of low-arginine containing peptides is that high-arginine content PMO-CPPs have demonstrated *in vivo* toxicity.[19] It was observed that reduction in the number of arginine residues confers improved toxicology outcomes but, at the same time, screening studies in *mdx* mice showed a general pattern of reduced dystrophin splice correction activity as arginine residues are removed from the CPP.[40] We show in this study that *in vivo* administration of **PMO-P6** to mice induces a dose-dependent increase in splice correction and EGFP synthesis in quadriceps, diaphragm and heart while having no effect on kidney biomarkers up to a dose of 60

mg/kg. The EGFP mice used in this study are an appropriate model system in that our *in vitro* cell assay can be directly translated to this animal model to validate activity on the same gene target.

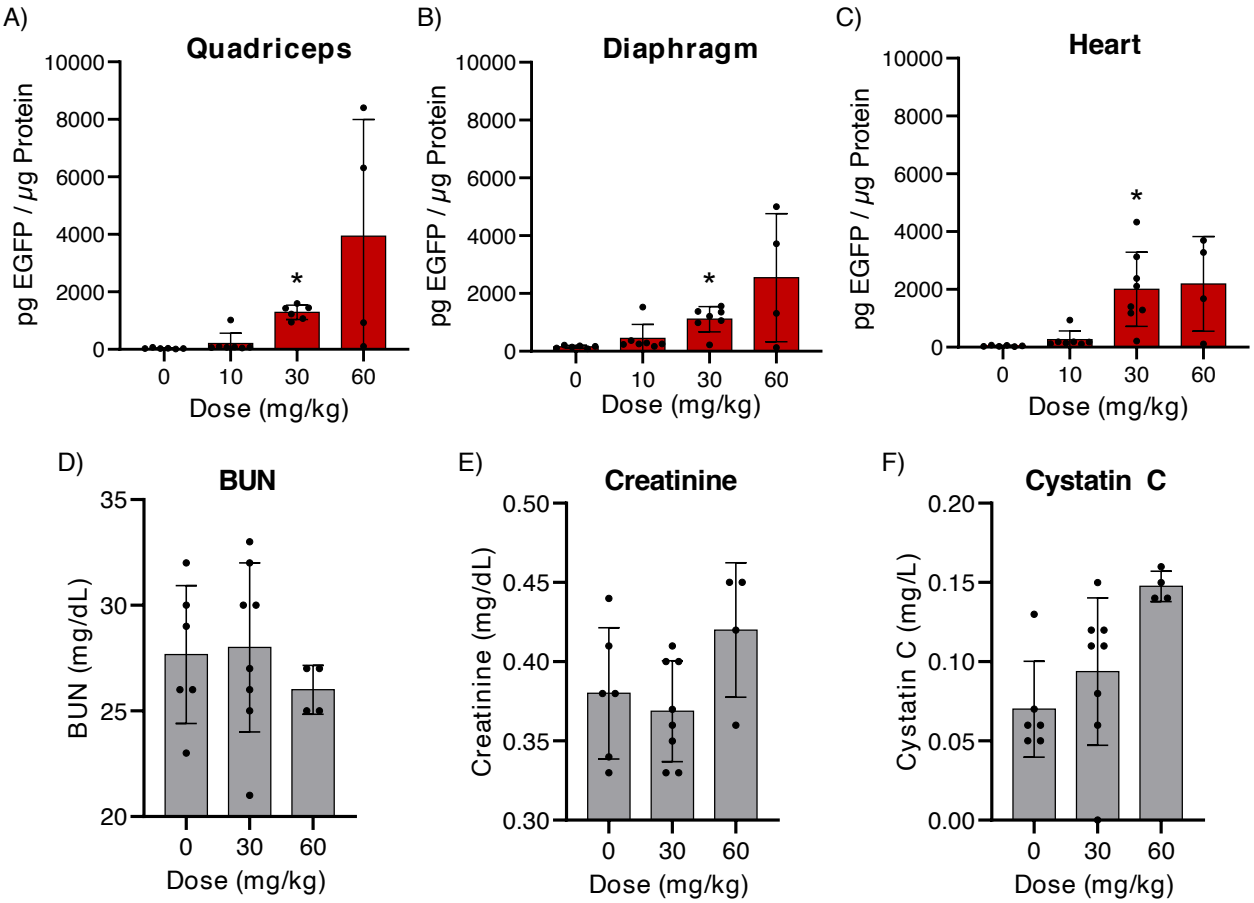


Figure 3.20 PMO-P6 leads to dose-dependent increase in EGFP expression without affecting kidney toxicity biomarkers EGFP expression levels in (A) quadriceps, (B) diaphragm, and (C) heart tissue seven days after a single i.v. injection of various doses of **PMO-P6**. Concentrations of kidney toxicity biomarkers (D) blood urea nitrogen (BUN), (E) creatinine, and (F) cystatin C isolated from serum seven days after a single i.v. injection of various doses of **PMO-P6**. Each bar represents group mean ± SD, dots represent individual samples. Quadriceps: saline (0 mg/kg) n = 6; 10 mg/kg, n = 6; 30 mg/kg, n = 7; 60 mg/kg, n = 4. Diaphragm: saline, n = 6; 10 mg/kg, n = 6; 30 mg/kg, n = 6; 60 mg/kg, n = 4. Heart: saline, n = 6; 10 mg/kg, n = 7; 30 mg/kg, n = 8; 60 mg/kg, n = 4. Statistical significance was determined relative to the 0 mg/kg dose (*$p < 0.01$, unpaired two-tailed student's t-test).

3.3 Discussion

Designing cell-penetrating peptides to deliver a specific macromolecular cargo is a longstanding challenge in chemical biology. Here, a machine learning model trained on longer sequences (>20 residues) drove the discovery of a short, low-arginine content peptide (<20 residues) that efficiently delivers PMO to the nucleus of cells. This cationic 18-mer peptide (**P6**) contains one arginine residue within its sequence and delivers PMO *in vivo*.

It is important to note that the machine learning model used here was trained with a PMO-CPP library composed of long sequences, with the vast majority being linear combinations of known CPPs. However, because our model is able to extrapolate predictions beyond the training set, we were able to generate unique short, low-arginine peptides that were not previously encountered. The challenge of designing effective CPP sequences became apparent during the experimental validation of the predicted peptides. We observed a length-dependent increase in activity up to 18 residues. However, two predicted peptides that differ only in their C-terminal regions, i.e., **P5** (C-terminus: MG) and **P6** (C-terminus: KXXC) displayed significantly different experimental activities, with the former about a third as active as the latter. Data augmentation was able to slightly improve the accuracy of activity prediction, but more work is required to more accurately design and predict unique short peptides for antisense delivery.

We found that the C-terminal cysteine residue boosted activity of **PMO-P6**, but that the cationic N-terminal motif was primarily responsible for its high activity. Mutation of the C-terminal cysteine to alanine resulted in a decrease in activity, but mutations of the other residues within the KXXC motif had no effect. Removal of the entire motif (**PMO-P12**) resulted in similar activity to the C18A mutation, but also led to a slightly larger concentration window between activity and toxicity compared to **P6** in vitro. The cysteine residue could be enhancing uptake by partaking in thiol-mediated uptake,[41] or by forming a more active "chimera" through disulfide formation,[10] although it is clear that this residue is necessary but not sufficient for the activity of **P6**. Further structure-activity studies using analogs of **P6** containing alanine mutations revealed that the pentalysine chain is the primary driver for its high activity, while the C-terminal cysteine residue only moderately boosts the activity. These observations indicate that a combination of length, composition, sequence, and charge influence activity of **P6**. These short peptides demonstrate how minor variations in side chain composition can significantly affect activity, making empirical optimization of short peptides an arduous and unpredictable task.

While **P6** contains a single arginine residue, it is still a polycationic peptide with nine lysine residues. The reason for this may be two-fold; cationic sequences may be ideal for PMO delivery, and the prediction model may have been biased towards polycationic sequences. In the past, our lab has tested hundreds of CPPs for the delivery of PMO, and have found that cationic sequences are in general more effective than other sequences.[9,10,12,35] At the same time, the library that was used to train the machine learning model was composed of cationic sequences, likely biasing the outcomes of prediction. With the constraint of reducing arginine, lysine residues appear to have taken its place. However, we do not expect that the lysine residues have simply 'replaced' arginine, as several past studies have shown that polyarginine peptides outperform polylysine peptides.[42] Here we have found a low-arginine containing peptide with comparable activity to a commonly studied oligoarginine CPP.

The design strategy discussed here is promising for generation of peptide sequences able to deliver macromolecular cargo other than antisense oligonucleotides into cells and may be applicable for design of other functional peptides with a suitable training dataset. Improvement in PMO-CPP library synthesis and testing to access a wider range of sequences and activity may be one strategy to improve machine learning-guided discovery of highly effective peptide sequences.

3.4 Materials & Methods

3.4.1 Materials

H-Rink Amide-ChemMatrix resin was obtained from PCAS BioMatrix Inc. (St-Jean-sur-Richelieu, Quebec, Canada). 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid-hexafluorophosphate (HATU) was purchased from P3 BioSystems. 5-azidopentanoic acid was obtained from Bachem. Fmoc-□-Ala-OH, Fmoc-6-amino hexanoic acid and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from Chem-Impex International. Fmoc-Ala-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Trp(Boc)-OH and Fmoc-Ser(Trt)-OH were purchased from Novabiochem. Peptide synthesis-grade N,N-dimethylformamide (DMF), dichloromethane (DCM), diethyl ether and HPLC-grade acetonitrile were obtained from VWR International. All other reagents, such as dibenzocyclooctyne-acid (DBCO acid), ammonium hydroxide (NH_4OH), piperidine, dimethyl sulfoxide (DMSO), N,N-diisopropylethylamine (DIEA), trifluoroacetic acid (TFA), thioanisole or triisopropylsilane (TIPS), were purchased from Sigma-Aldrich. Milli-Q water was used exclusively. Minimum essential medium (MEM) and tripsin-EDTA (0.25%) were obtained from Gibco. Penicillin Streptomycin was bought from ThermoFisher Scientific. Fetal bovine serum (FBS) was purchased from MilliporeSigma. Phosphate buffered saline (PBS) 1X was obtained from Corning. Propidium iodide was purchased from Invitrogen. DMEM-high glucose was obtained from HyClone. HeLa-654 cells were obtained from the University of North Carolina Tissue Culture Core facility. Human renal proximal tubule epithelial (TH1 RPTEC) cells were bought from Kerafast. PMO IVS2-654 was provided by Sarepta Therapeutics. For the LDH assay, the CytoTox 96□ non-Radioactive cytotoxicity assay kit was purchased from Promega.

3.4.2 Methods for LC-MS analysis

LC-MS analyses were performed on different spectrometers and using different methods.

Agilent 6520 ESI-Q-TOF mass spectrometer equipped with a C18 Zorbax column (300SB C18, 150 mm x 2.1 mm ID, 5 μm 300 Å silica). Mobile phases were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LC-MS method was used for characterization: 1% B from 0 to 2 min, linear ramp from 1% B to 61% B from 2 to 11

min, 61% B to 95% B from 11 to 12 min and finally 3 min of post-time at 1% B for equilibration, flow rate: 0.6 mL/min.

Agilent 6550 ESI-iFunnelQ-TOF mass spectrometer equipped with a C18 column (Phenomenex Luna C18(2), 150 mm x 0.5 mm ID (capillary), 3 μm 100 Å silica). Mobile phases were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LC-MS method was used for characterization: 1% B and ramping to 61% B over 12 minutes. MS was on from 4 to 14 minutes. Flow rate was 50 μL/min.

Agilent 6550 ESI-iFunnelQ-TOF mass spectrometer equipped with a C4 column (Phenomenex Jupiter C4, 150 mm x 1.0 mm ID, 5 μm 300 Å silica). Mobile phases were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LC-MS method was used for characterization: 1% B and ramping to 61% B over 10 minutes. MS was on from 4 to 12 minutes. Flow rate was 100 μL/min.

Chromatograms were obtained using one of the previous methods unless otherwise noted. All data were processed using Agilent MassHunter software package. Y-axis in all chromatograms shown represents total ion current (TIC) unless noted.

3.4.3 Peptide synthesis, purification, and conjugation

Peptides were synthesized using the Fmoc/tBu strategy via continuous fast-flow peptide synthesis, using either fully-automated or semi-automated technology as previously described.[43,44]

Semi-automated fast-flow peptide synthesis

Peptides P**1**- **P19** were synthesized on a 0.1 mmol scale using semi-automated fast-flow peptide synthesis. A 140 mg portion of H-Rink Amid-ChemMatrix was loaded into the reactor and maintained at 70 °C. All reagents and solvents were flowed through a stainless-steel loop maintained at 70 °C before introduction into the reactor. For each coupling, 2.5 mL of HATU (0.38 M in DMF) and amino acid (0.1 mmol) were mixed with 0.5 mL of DIEA and delivered to the reactor at 6 mL/min followed by washing with DMF for 1 minute at 20 mL/min. Fmoc removal was accomplished washing with 20% (v/v) piperidine in DMF for 20 seconds at 20 mL/min. Between each step and at the end of the peptide synthesis, DMF (1 minute at 20 mL/min) was used to wash out the reactor.

Functionalization of the peptides with azide at the N-terminus

To cap the peptides with 5-azidopentanoic acid, 50 mg of each peptide attached to the resin (P**1**- **P7**) were incubated for 1 hour at room temperature with 5-azidopentanoic acid (0.25 mmol) dissolved in 0.625 mL of HATU (0.38 M in DMF) and 0.125 mL of DIEA. For peptides P**8**- **P19**, 150 mg of each peptide attached to the resin were incubated for 1 hour at room temperature with 5-azidopentanoic acid (0.75 mmol) dissolved in 1.875 mL of HATU (0.38 M in DMF) and 0.375 mL of DIEA. After completion of the synthesis, the resins were washed 3 times with DCM and dried under vacuum.

Peptides cleavage and deprotection

After functionalization of 50 mg of the peptide-resin, each peptide was subjected to simultaneous global side-chain deprotection and cleavage from resin by treatment with 1.25 mL of a a mixture of 94% trifluoroacetic acid (TFA), 2.5% thioanisole, 2.5% water and 1% triisopropylsilane (TIPS) (v/v) and the mixture was kept for 2 hours at room temperature.

Afterwards, 10 mL of cold diethyl ether (chilled at -80 °C) were added to precipitate and wash each of the peptides. The crude product was pelleted through centrifugation for 3 minutes at 4,000 rpm and the diethyl ether was decanted. Precipitation with cold diethyl ether (4 mL) and centrifugation was repeated two more times. After the third wash, the pellet was dissolved in a mixture (4 mL) of 50% water, 50% acetonitrile and 0.1% TFA, vortexed and filtered through a fritted syringe to remove the resin and lyophilized.

In the case of the cleavage of 150 mg of functionalized peptide-resin, 4 mL of the same cleavage cocktail were used and after cleavage, the triple quantities of cold diethyl ether were used to precipitate and wash the peptides.

Peptides purification by HPLC

The peptides were dissolved in water and acetonitrile containing 0.1% TFA, filtered through a 0.22 µm nylon filter and purified by mass-directed semi-preparative reversed-phase HPLC. Solvent A was water with 0.1% of TFA and solvent B was acetonitrile with 0.1% of TFA. The peptides were purified on an Agilent Zorbax 300SB-C18 column: 9.4 x 250 mm, 5 µm. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The

purity of the fraction pool was confirmed by LC-MS. For peptides P**1**- **P7**, a linear gradient that changed at a rate of 1%/min was used starting with 1% B and finishing with 61% B over a period of 60 minutes. For peptides P**8**- **P19** the purification started with 1% B and finished with 40% B over a period of 60 minutes.

Peptides purification using reversed-phase flash chromatography (Biotage)

Peptides P**6**, P**8**- **P18** were purified using reversed-phase flash chromatography (Biotage). The peptides were dissolved in water and acetonitrile containing 0.1% TFA and purified using reversed-phase flash chromatography. Solvent A was water with 0.1% of TFA and solvent B was acetonitrile with 0.1% of TFA. The peptides were purified using a SNAP Bio C18 300 Å, 10 g column. A linear gradient from 1% B to 10% B was used during 4 column volumes, 10% B to 50% B during 15 column volumes followed by an equilibration step during 2 column volumes at 50% B. Using UV data of each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

PMO functionalization and purification

PMO IVS2-654 was functionalized with DBCO acid. The DBCO acid (10 mg) was dissolved in 80 µL of amine-free DMF. Then, 74 µL of HBTU (0.4 M in amine-free DMF) and 5.6 µL of DIEA were added to the DBCO acid solution. Separately, 100 mg of PMO IVS2-654 were dissolved in 300 µL of DMSO. The DBCO solution was added to the PMO solution and mixed gently. The reaction was left to react for 30 minutes. Afterwards, the reaction was quenched with 2 mL of water and 4 mL of NH_4OH and left for 20 minutes. The reaction mixture was diluted with 40 mL of water/acetonitrile (95:5) and purified by mass-directed preparative reversed-phase HPLC. Solvent A was water and solvent B was acetonitrile. The PMO functionalized with DBCO was purified on an Agilent Zorbax 300SB-C3 column: 21.2 x 100 mm, 5 µm. A linear gradient was used starting with 2% B and finishing with 60% B over a period of 70 minutes. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

PMO-peptide conjugation using copper-free azide-alkyne click chemistry

Peptide conjugates were synthesized using copper-free azide-alkyne 1,3-dipolar cycloaddition in water. PMO-DBCO was dissolved in water at 5 mM concentration (determined gravimetrically). The peptides were dissolved in water at 5 mM concentration (determined gravimetrically; the molecular weight was calculated to include 1 trifluoroacetate counter ion per lysine and arginine residue). In a microcentrifuge tube, equal volumes of PMO-DBCO solution and peptide solution were mixed (volumes between 12 µL and 100 µL depending on the reaction) and the resulting mixtures were vortexed. The reactions took place within a period of 1 day and they were followed by LC-MS until ~95% consumption of peptide was observed. The PMO-peptide conjugates were lyophilized.

PMO-peptide conjugates purification by HPLC

Each PMO-peptide conjugate was dissolved in water. The solution was filtered through a 0.22 µm nylon filter and purified by mass-directed semi-preparative reversed-phase HPLC. Solvent A was water with 100 mM ammonium acetate (pH 7.2) and solvent B was acetonitrile.

Conjugates **PMO-P1** to **PMO-P7** were dissolved in water and purified on an Agilent Zorbax 300SB-C3 column: 9.4 x 250 mm, 5 µm. A linear gradient that changed at a rate of 1% per min was used starting with 1% B and finishing with 61% B over a period of 60 minutes. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS. PMO-peptide conjugates were obtained as their acetate salts.

Conjugates **PMO-P8** to **PMO-P19** were dissolved in water and purified on an Agilent Zorbax 300SB-C3 column: 2.1 x 150 mm, 5 µm. A linear gradient that changed at a rate of 2% per min was used starting with 1% B and finishing with 61% B over a period of 30 minutes. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS. PMO-peptide conjugates were obtained as their acetate salts.

3.4.4 Cell Assays

Exon skipping activity of the conjugates with the peptides predicted by machine learning and analogs of P6

HeLa654 cells were maintained in MEM (Minimum Essential Medium) supplemented with 10% (v/v) of fetal bovine serum (FBS) and 1% (v/v) of penicillin-streptomycin (Pen Strep) at 37 °C and 5% CO_2. 24 hours before treatment, HeLa654 cells were plated at a density of 8,000 cells per well in a 96-well plate. The next day, fresh 1 mM stocks of each of the PMO-peptide conjugates in PBS (1X) were prepared. The exact concentration of the stock solutions was determined by measuring the absorbance at 260 nm and a PMO extinction coefficient of 168,700 L mol^{-1} cm^{-1}. The growth media was aspirated from the cells and treatment media consisting of each respective conjugate (5 μM concentration) in MEM supplemented with 10% FBS and 1% Pen Strep was added. The cells were incubated with treatment-containing media for 22 hours at 37 □C and 5% CO_2. Next, the treatment media was aspirated. Trypsin-EDTA (0.25%) (20 μL) was added to the cells and incubated for 10 min at 37 °C and 5% CO_2. To quench the trypsin, 80 μL of MEM supplemented with 10% (v/v) FBS and 1% (v/v) Pen Strep was added to each well. The dissociated cells in media were transferred with a multichannel pipette to a polypropylene v-bottom 96-well plate (Falcon) and centrifuged at 500 rcf (relative centrifugal force) for 3 min. The supernatant was removed, the cell pellets were resuspended with 200 μL of PBS (1X), and the plate was centrifuged again at 500 rcf for 3 min. The supernatant was again removed and the pellets were resuspended in 300 μL of PBS (1X) with 2% FBS (v/v) and 0.2% of a solution of propidium iodide in water. Fluorescence intensity was analyzed by flow cytometry on a BD LSRII flow cytometer. Gates were applied to the data to ensure that cells that were highly positive for propidium iodide or exhibited forward/side scatter readings that were sufficiently different from the main cell population were excluded. The mean fluorescence intensity was calculated for each sample and normalized with respect to the fluorescence of the cells treated with unconjugated PMO. For each PMO-peptide conjugate, the mean fluorescence intensity was averaged across biological replicates.

Exon skipping activity of PMO-P6 and PMO-P12 using endocytosis inhibitors

To interrogate endocytic mechanism, we performed pulse-chase experiment with a panel of endocytosis inhibitors including: chlorpromazine, which is demonstrated to interfere with clathrin-

mediated endocytosis; cytochalasin D, which inhibits phagocytosis and macropinocytosis; wortmannin, which alters various endocytosis pathways by inhibiting phosphatidylinositol kinases; EIPA (5-(N-ethyl-N-isopropyl) amiloride), which inhibits macropinocytosis; and Dynasore, which also inhibits clathrin-mediated endocytosis. HeLa654 cells were plated at a density of 8,000 cells per well in a 96-well plate in MEM supplemented with 10% FBS and 1% penicillin-streptomycin. The stock solutions used for each inhibitor were as follows: chlorpromazine - 10 mM in water; cytochalasin D - 10 mM in DMSO; wortmannin - 2 mM in DMSO; EIPA - 10 mM in DMSO; Dynasore - 10 mM in DMSO. The next day, the cells were treated with each inhibitor at the following concentrations: chlorpromazine (10 and 1 mM); cytochalasin D (20 and 5 mM); wortmannin (200 and 50 mM); EIPA (50 and 10 mM); Dynasore (80 mM) in PBS 1X. After 30 min, **PMO-P6** or **PMO-P12** was added to each well at a concentration of 5 μM. The experiment was conducted in a pulse-chase format, and we evaluated the resulting change in PMO activity. HeLa654 cells were preincubated with various inhibitors for 30 minutes. Treatment with **PMO-P6** or **PMO-P12** and incubation at 37 °C and 5% CO_2 for 3 hours was followed by replacement of the media with fresh media (containing neither inhibitor nor PMO-peptide) and the cells were allowed to grow for another 22 hours at 37 °C and 5% CO_2. Sample preparation and flow cytometry was then performed as usual.

LDH assay

This assay measures the amount of lactate dehydrogenase (LDH) released into the cell culture supernatant by damaged cells.[i] Conversion of lactate to pyruvate produces NADH which in turns reduces a yellow tetrazolium salt (iodonitrotetrazolium violet; INT) into a red formazan dye that absorbs at 490 and 492 nm. As a consequence, the amount of LDH in the supernatant is proportional to the amount of formazan and it informs about the number or lysed cells (dead or damaged).

TH1 RPTEC cells were maintained in DMEM-high glucose supplemented with 10% (v/v) FBS and 1% (v/v) Pen Strep at 37 °C and 5% CO_2. Eighteen hours before treatment, TH1 RPTEC cells were plated at a density of 8,000 cells per well in a 96-well plate. The next day, fresh 10 mM stocks of each of PMO-peptide conjugate were prepared in PBS (1X). The concentration of the stocks was determined by measuring the absorbance at 260 nm and using an extinction coefficient of 168,700 L mol^{-1} cm^{-1}. The growth media was aspirated from the cells and treatment media was

added with each respective conjugate at different concentrations (between 1 and 200 □M) in DMEM-high glucose supplemented with 10% FBS and 1% Pen Strep. The cells were incubated with treatment-containing media for 22 hours at 37 °C and 5% CO_2. Next, the supernatant treatment media was transferred to another clear-bottom 96-well plate for the assay. The assay was performed using the CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega) according to the included technical bulletin with the only difference of using half of the specified amounts (25 µL of each supernatant, 25 µL of the LDH Reagent and 25 µL of the stop solution). The absorbance was measured on a BioTek Epoch Microplate Spectrophotometer at 490 nm. The positive and the negative controls correspond to the maximum cell lysis and to the untreated cells respectively. The data were worked up by subtracting the absorbance of untreated cells from all of the treatment conditions, including the cell lysis, and then dividing by the corrected lysis value. The percentage of cytotoxicity was calculated as:

$$\%Cytotoxicity = 100 \times \frac{(ExperimentalLDHRelease - MediumBackground)}{(MaximumLDHReleaseControl - MediumBackground)}$$

3.4.5 Protein Delivery

Gly_5-DTA was recombinantly expressed and purified via P**6-LPSTGG** peptide was synthesized and purified by our standard protocol as described above. **G_5-DTA** (50 µM) was incubated with P**6-DTA** (500 µM) and SrtA* (2.5 µM) for 90 min at 4 °C in SrtA buffer (10 mM $CaCl_2$, 50 mM Tris, 150 mM NaCl, pH 7.5). The reaction was monitored by LCMS and gel electrophoresis. After 90 min, P**6-DTA** conjugate was isolated using HiTrap SP HP cation exchange column (GE Healthcare, UK) in A: 20 mM Tris, 50 mM NaCl, pH 7.5 buffer and B: 10 mM Tris, 1 M NaCl, pH 11, at a gradient of 0-100 over 20 column volumes. Fractions containing the pure product as determined by LCMS and gel electrophoresis were concentrated using a centrifugal filter unit (10K, Millipore).

Cytosolic delivery of P6-DTA was tested in HeLa cells. Cells were plated at 5,000 cells/well in a 96-well plate the day before the experiment. P6-DTA was prepared at varying concentrations in complete media and transferred to the plate, and incubated at 37 °C and 5% CO_2. Cell proliferation was measured after 48 h using the CellTiter-Glo assay, and ATP concentration was reported normalized to the no treatment condition.

3.4.6 In vivo studies

The mice used in the study contain a similar transgene as the HeLa654 cells. EGFP-654 transgenic mice were first obtained from Dr. Ryszard Kole's laboratory.[34] This mouse model ubiquitously expresses EGFP-654 transgene throughout body under chicken β-actin promoter. A mutated nucleotide 654 at intron 2 of human β-globin gene is contained in the EGFP-654 sequence which interrupts EGFP-654 coding sequence and prevents proper translation of EGFP protein. The antisense activity of PMO blocks aberrant splicing and resulted in EGFP expression, the same as in the HeLa 654 assay. In this study, 6- to 8-week-old male EGFP-654 mice bred at Charles River Laboratory were shipped to the vivarium at Sarepta Therapeutics (Cambridge, MA). These mice were group housed with ad libitum access to food and water. All animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Sarepta Therapeutics.

Before injection, the PMO-peptide was confirmed to have minimal endotoxin levels. For the endotoxin assay measurement using the **PMO-P6** lot that was used for animal studies, 0.5 mg of **PMO-P6** as acetate salt were dissolved in 1 mL of PBS (1X). The cartridge used was the 0.01 of the Charles River Endosafe nexgen-PTS. 25 □L of the sample were placed into each of the four sample reservoirs of the cartridge. The lot of **PMO-P6** (63 mg as acetate salt) used for animal studies showed 0.0471 EU/mg (EU refers to Endotoxin Units).

After 3-days of acclimation, mice were randomized into groups to receive a single *i.v.* tail vein injection of either saline or **PMO-P6** at the indicated doses; 0, 10, 30, and 60 mg/kg. Seven days after the injection, the mice were euthanized for serum and tissue sample collection. Quadriceps, diaphragm, heart were rapidly dissected, snap-frozen in liquid nitrogen and stored at -80 °C until analysis.

Serum from all groups were collected seven days post-injection and tested for kidney injury markers using a Vet Axcel Clinical Chemistry System (Alfa Wassermann Diagnostic Technologies, LLC). Specifically, serum BUN, creatinine, and cystatin C levels were measured using ACE® Creatinine Reagent (Alfa Wassermann, Cat# SA1012), ACE® Blood Urea Nitrogen Reagent (Alfa Wassermann, Cat# SA2024) and Diazyme Cystatin C immunoassay (Diazyme Laboratories, Cat# DX133C-K), respectively, per manufacturer's recommendation.

20-25 mg of mouse tissue was homogenized in RIPA buffer (Thermo Fisher, Cat# 89900) with protease inhibitor cocktail (Roche, 04693124001) using a Fast Prep 24-5G instrument (MP Biomedical). Homogenates were centrifuged at 12,000 g for 10 min at 4 °C. The resultant

supernatant lysates were quantified by Pierce BCA Protein Assay Kit (Thermo Fisher, Cat# 23225) and saved for EGFP expression measurement. Specifically, 80 µg of lysates were aliquoted in each well in a black-wall clear-bottom 96-well microplate (Corning). EGFP fluorescent intensity of each sample was measured in duplicates using a SpectraMAx i3x microplate reader (Molecular devices) by default setting. The average EGFP fluorescent intensity of each sample was then plotted against a standard curve constructed by recombinant EGFP protein (Origen, Cat#TP790050) to quantify EGFP protein level per µg protein lysate.

3.4.7 Statistics

Statistical analysis and graphing was performed using Prism (Graphpad) or Excel (Microsoft). Concentration-response curves were fitted using Prism using nonlinear regression. The listed replicates for each experiment indicates the number of distinct samples measured for a given assay. Significance for activities between constructs was determined using a student's two-sided, unpaired t-test.

3.5 Acknowledgements

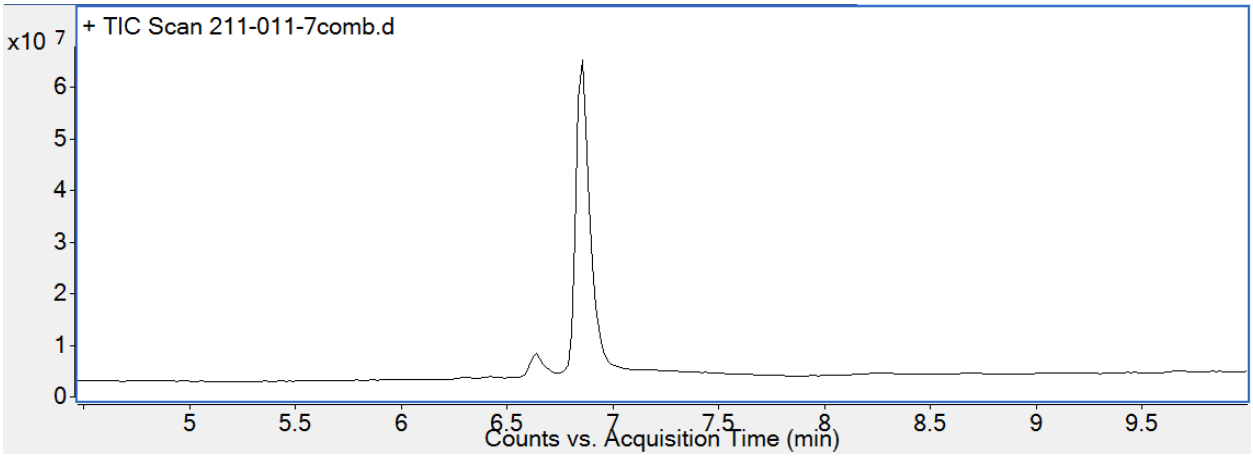
This research was funded by Sarepta Therapeutics. We thank the Swanson Biotechnology Center Flow Cytometry Facility at the Koch Institute for the use of their flow cytometers, and Bryan Mastis and Sam Foley for help with in vivo studies. C.K.S. (4000057398) acknowledges the National Science Foundation Graduate Research Fellowship (NSF Grant No. 1122374) for research support. S.M. acknowledges an MIT Jameel Clinic – AI for Health (J-Clinic) fellowship for research support.

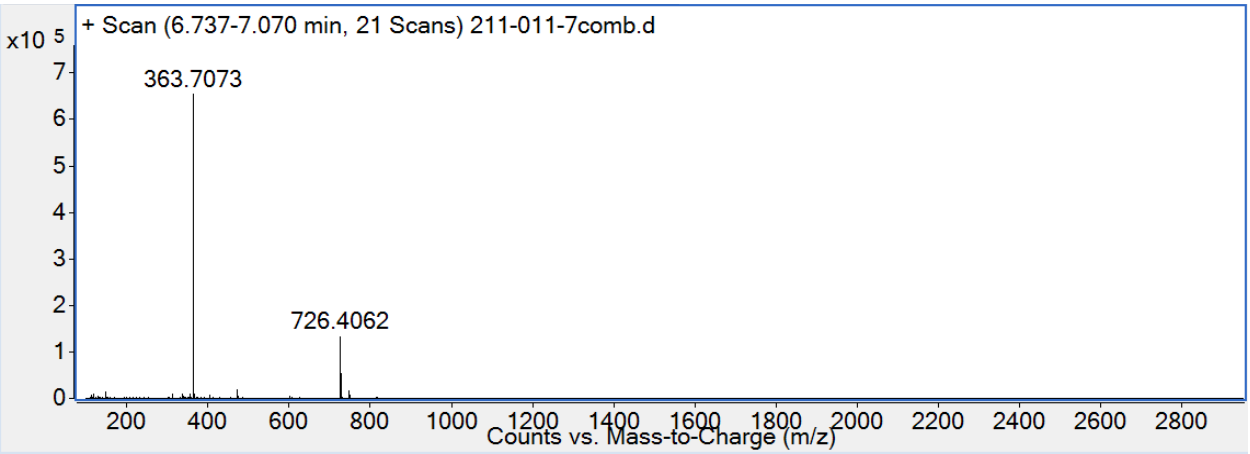
3.6 Author contributions:

C.K.S., E.L.V., S.M., R.G.B., and B.L.P conceptualized the research. E.L.V. and C.K.S. prepared and tested PMO-peptides. C.K.S. prepared and tested PNA and protein conjugates. S.M. and R.G.B. developed the machine learning model with input from C.K.S. and B.L.P. K.B., C.L.W., and J.A.W, performed in vivo study with input from A.M. C.K.S., E.L.V., A.L. and B.L.P wrote the manuscript with input from all authors.

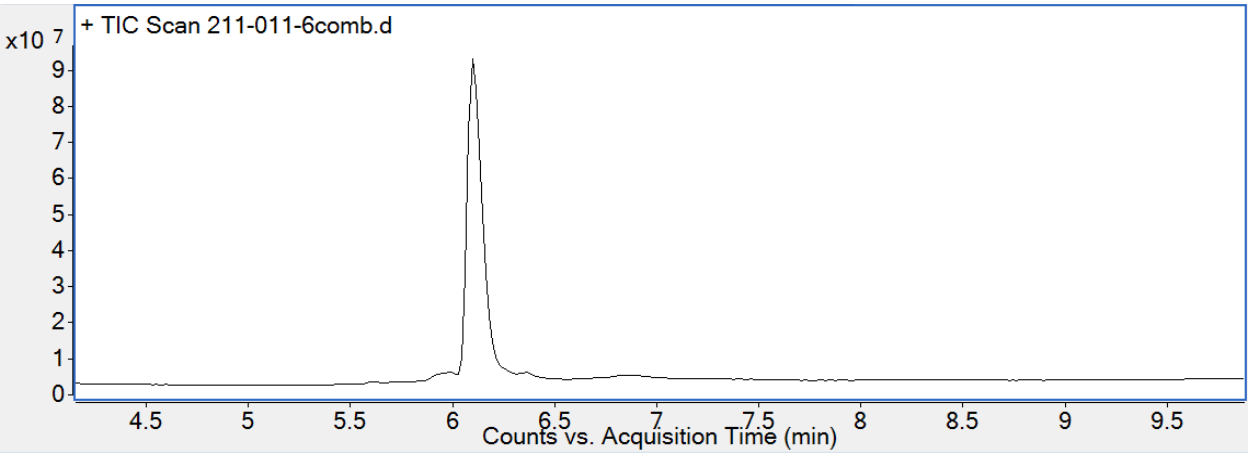
3.7 Appendix I: LC-MS Characterization

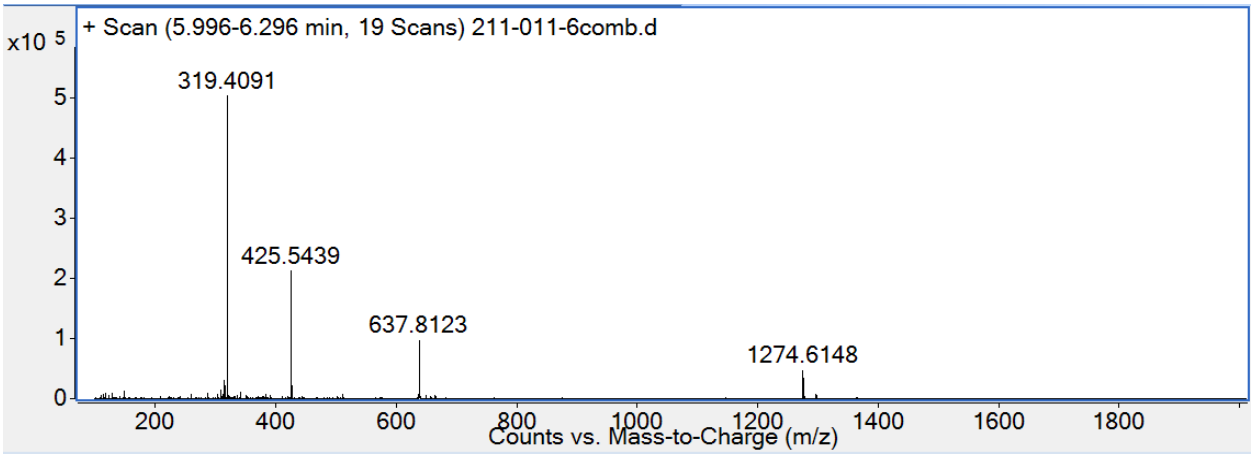
azide-P1
Mass expected: 725.5 Da
Mass observed: 725.4 Da
Peptide sequence: KXKXT



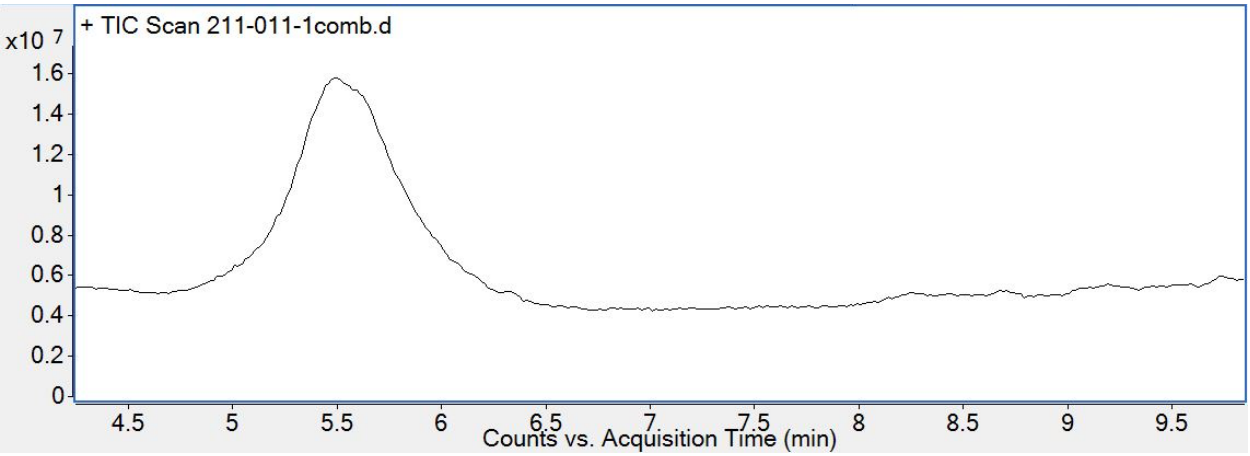


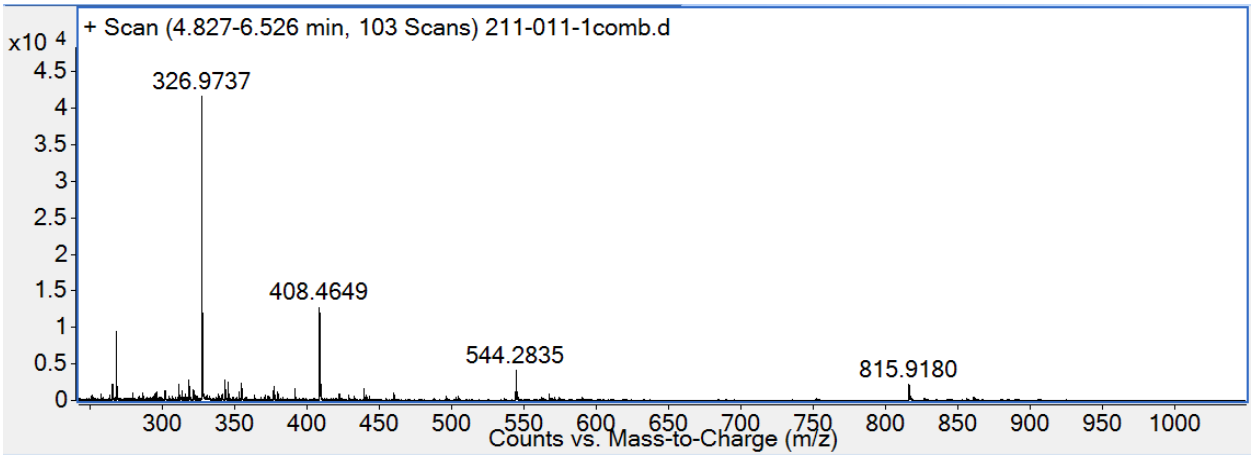
azide-P2
Mass expected: 1273.8 Da
Mass observed: 1273.6 Da
Peptide sequence: KXKHQQQXK



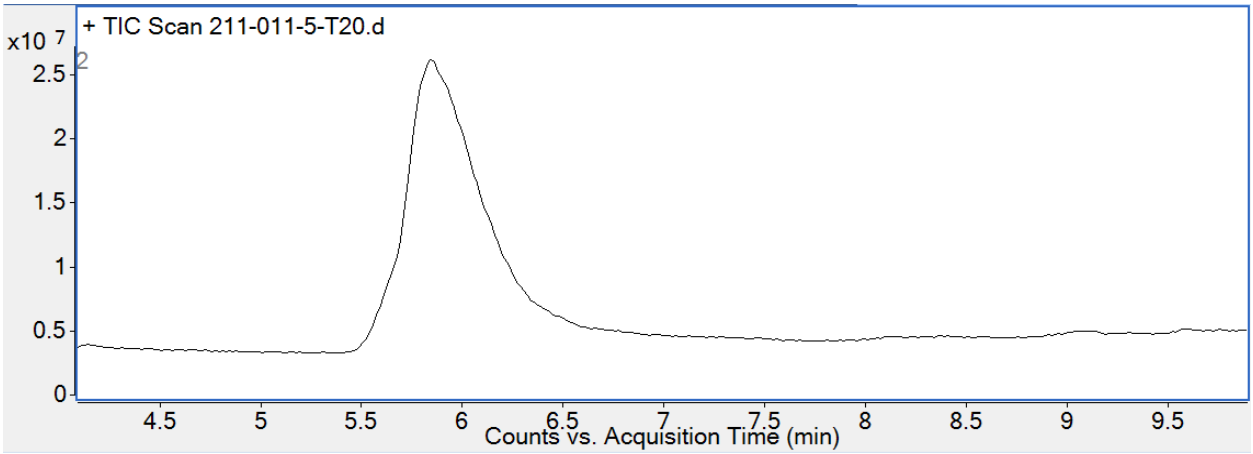


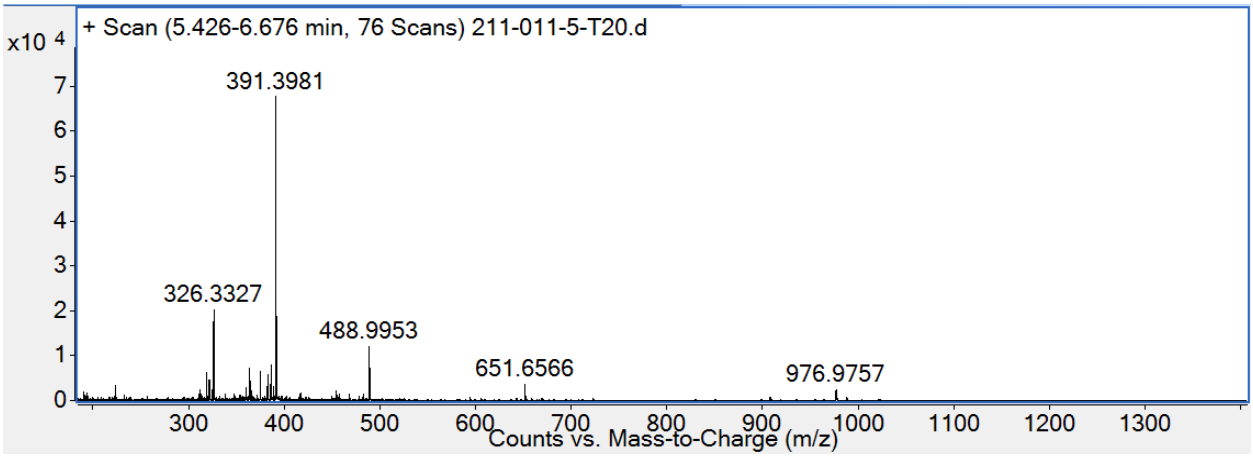
azide-P3
Mass expected: 1630.0 Da
Mass observed: 1629.9 Da
Peptide sequence: KXKKQQGKKKHR



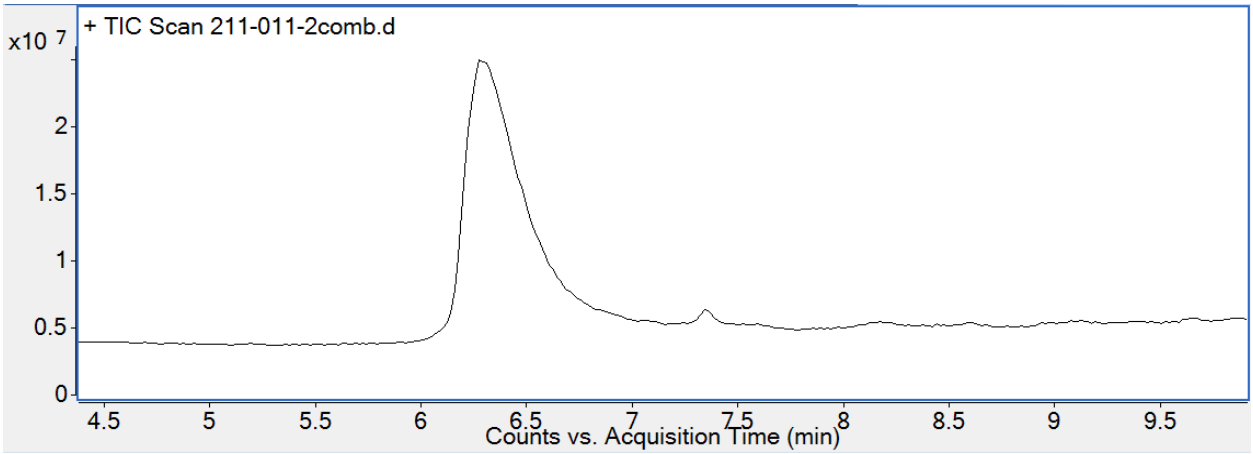


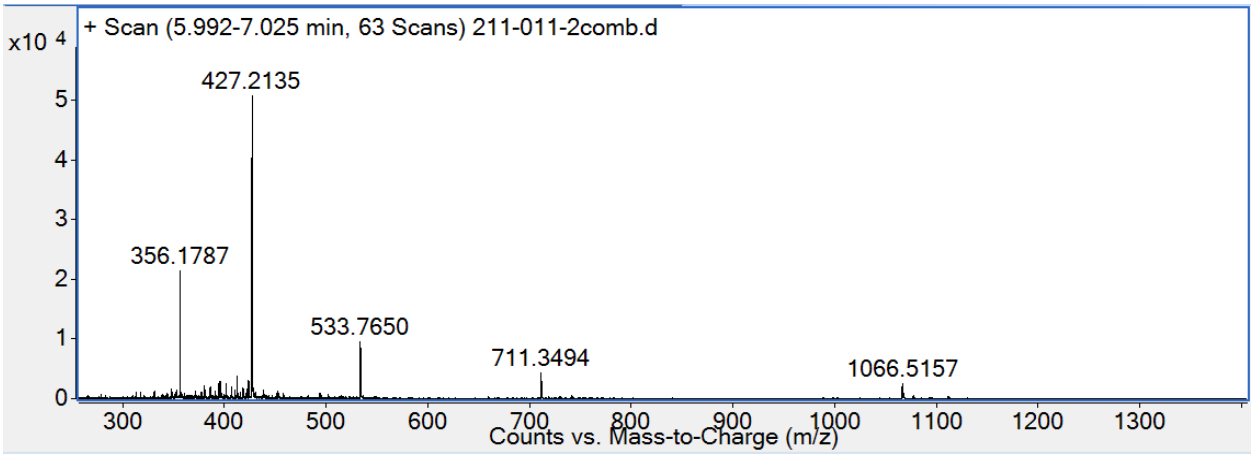
azide-P4
Mass expected: 1951.2 Da
Mass observed: 1952.0 Da
Peptide sequence: HKKKKQBKKKHRWP



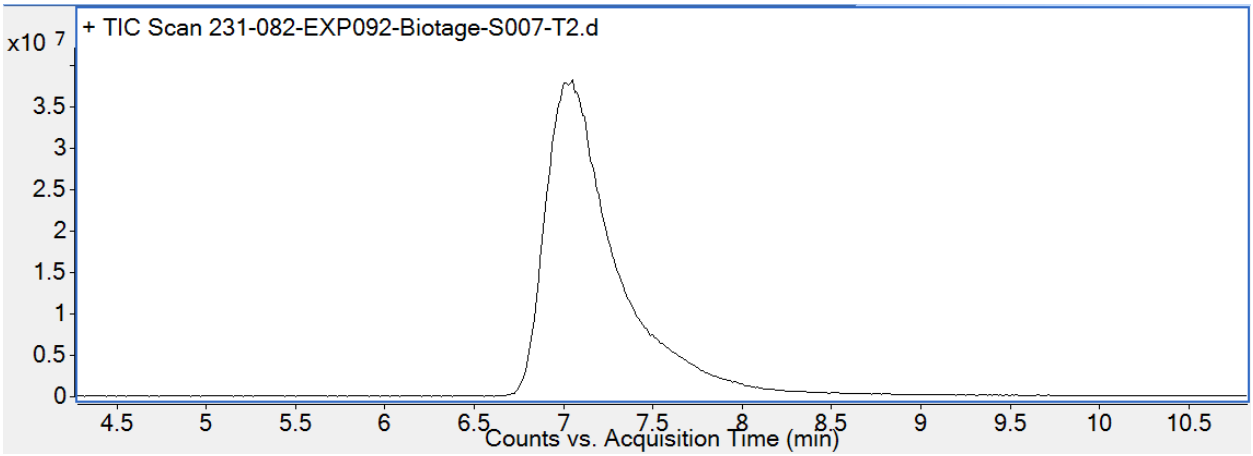


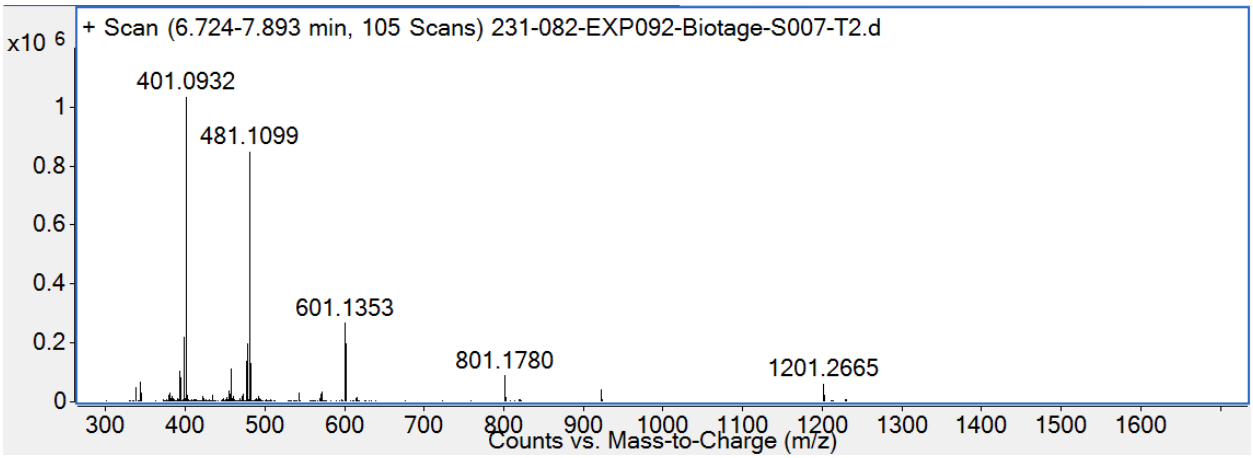
azide-P5
Mass expected: 2131.3 Da
Mass observed: 2131.1 Da
Peptide sequence: KKKKKQBKKKHRWPMG



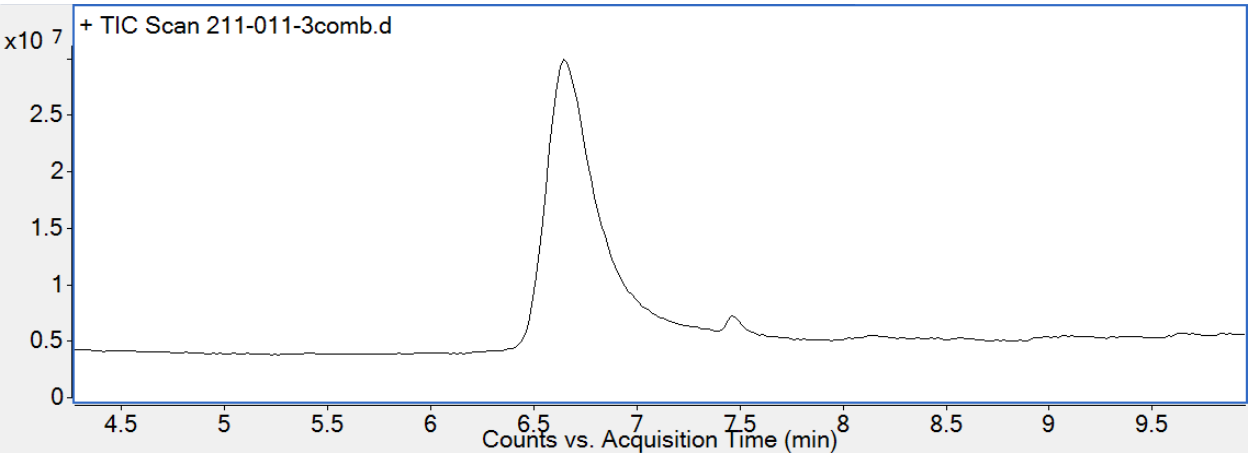


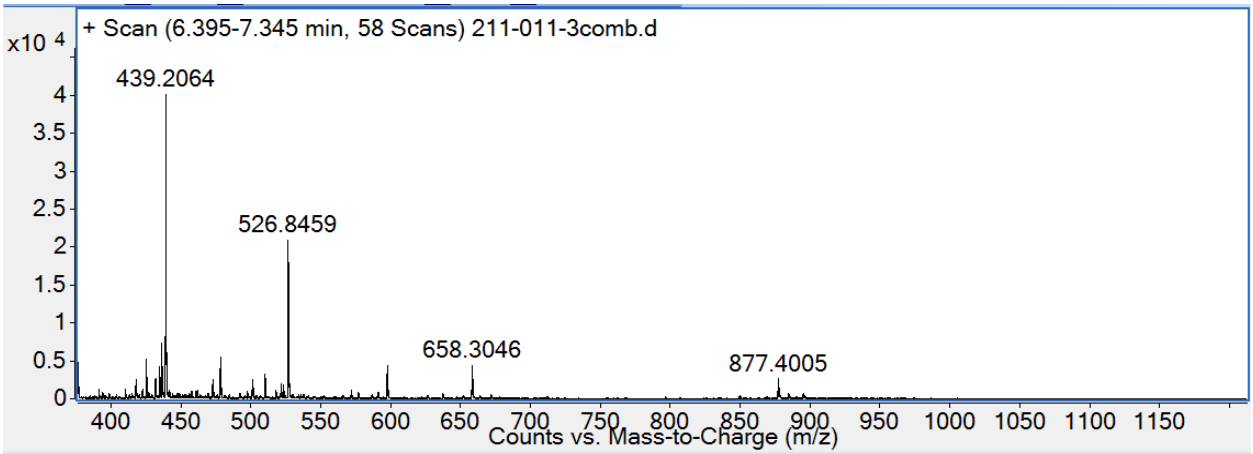
azide-P6
Mass expected: 2400.5 Da
Mass observed: 2400.2 Da
Peptide sequence: KKKKKQBKKKHRWPKXXC



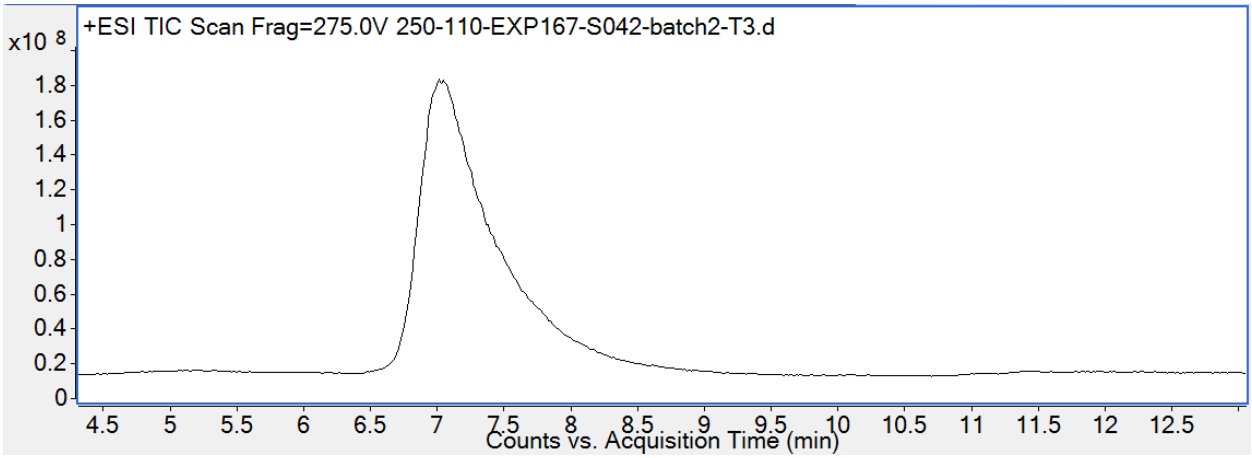


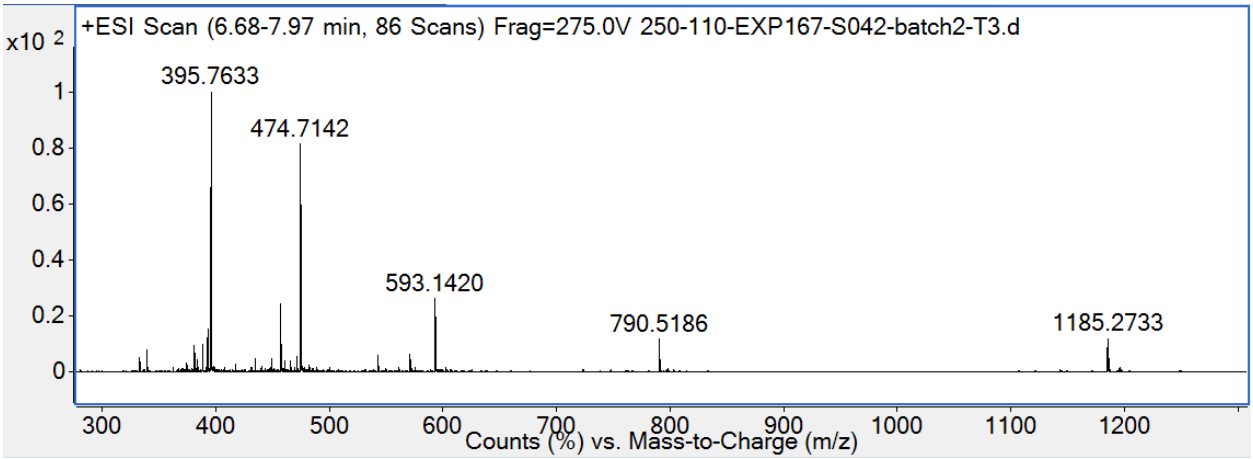
azide-P7
Mass expected: 2629.5 Da
Mass observed: 2629.2 Da
Peptide sequence: KKKKNQBKKKHRWPMKXCPQ



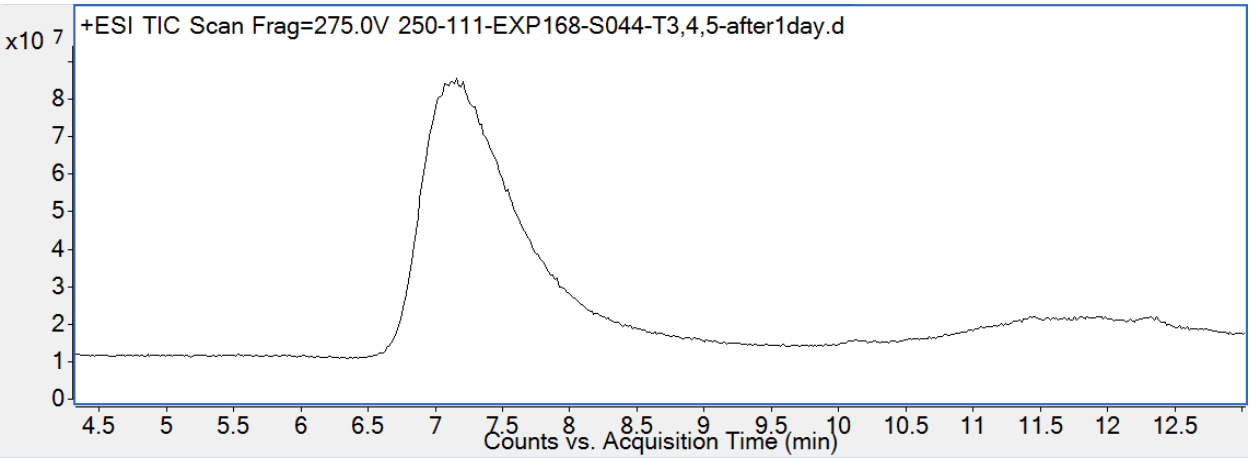


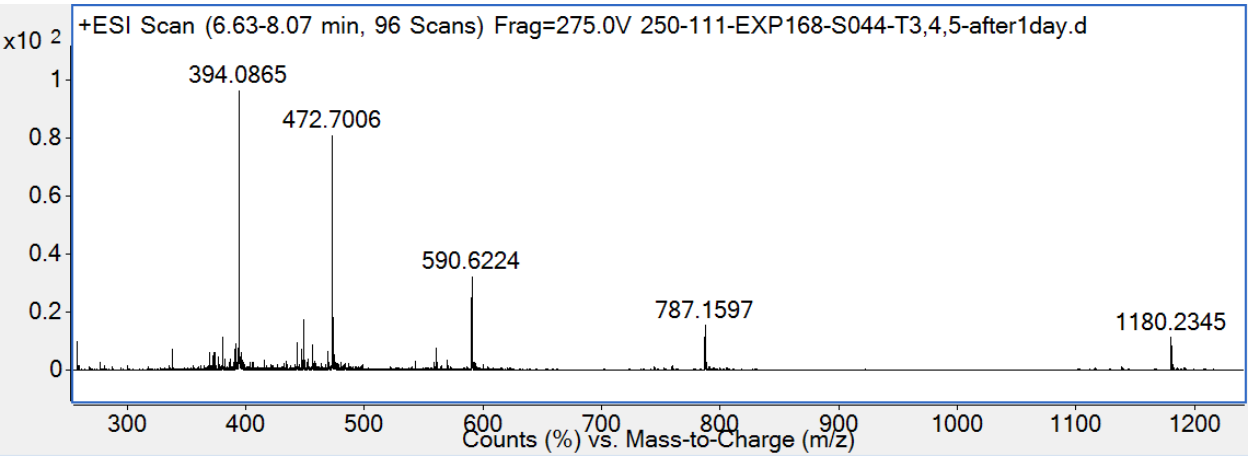
azide-P8
Mass expected: 2367.5 Da
Mass observed: 2368.6 Da
Peptide sequence: KKKKKQBKKKHRWPKXXA



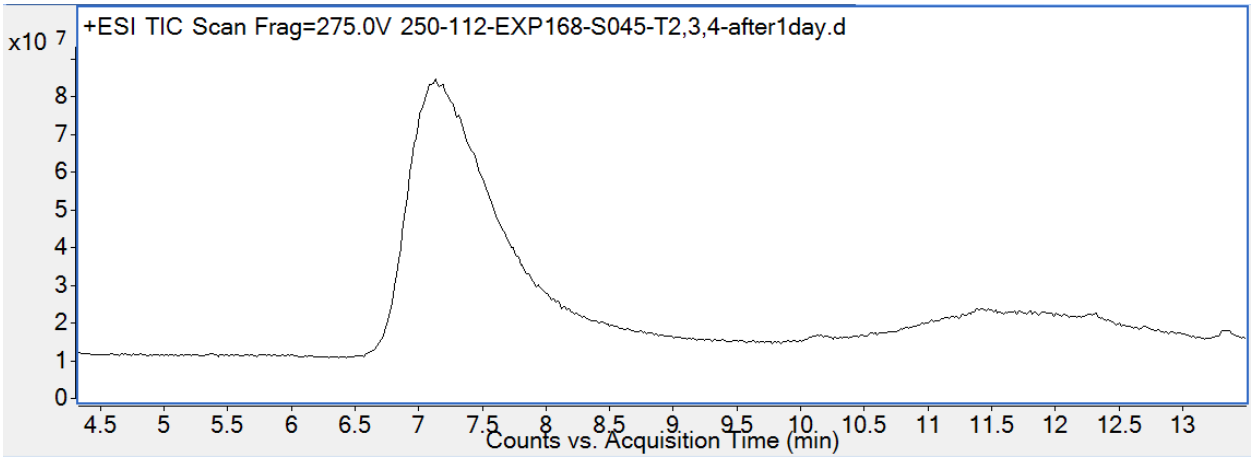


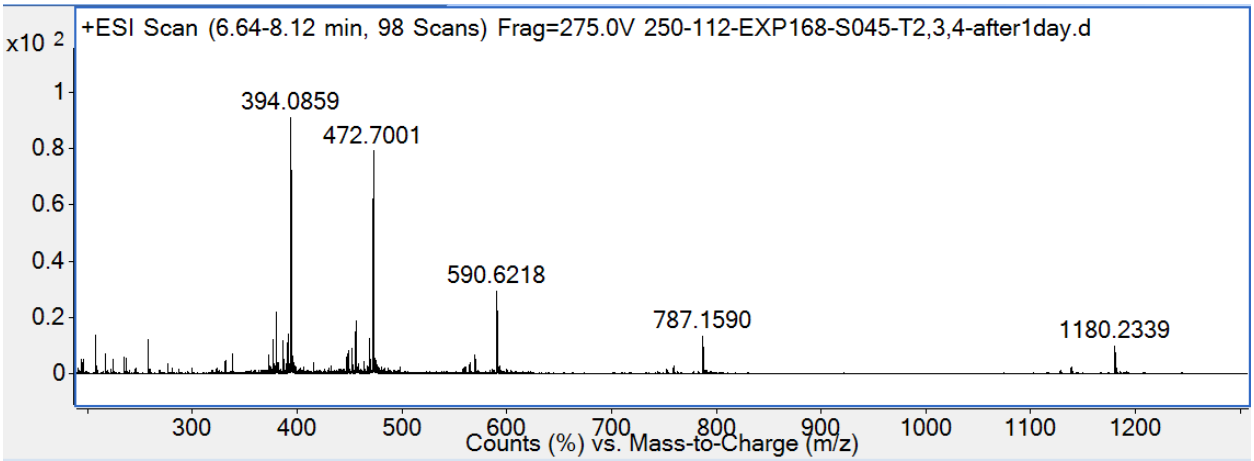
azide-P9
Mass expected: 2357.5 Da
Mass observed: 2358.5 Da
Peptide sequence: KKKKKQBKKKHRWPKXAC



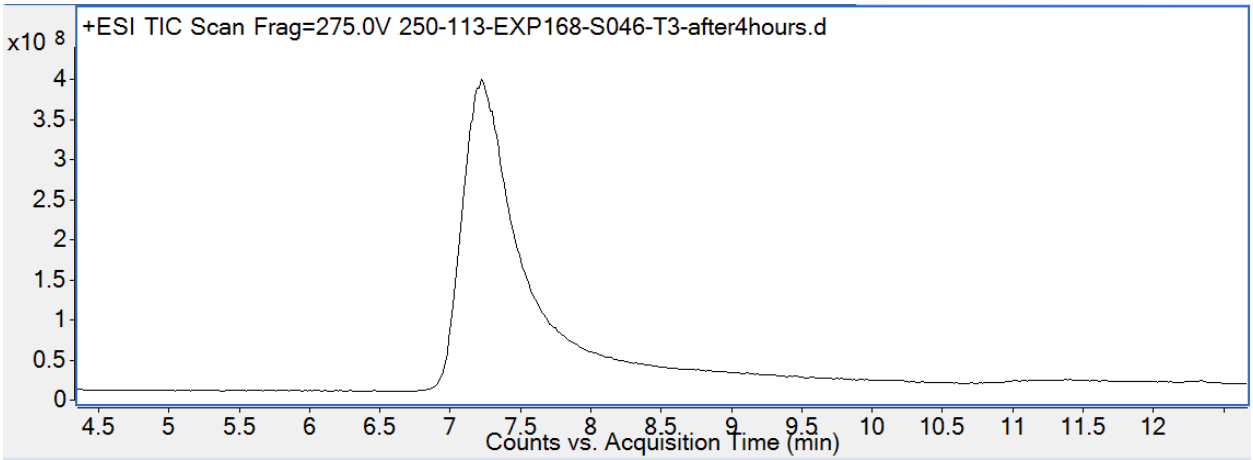


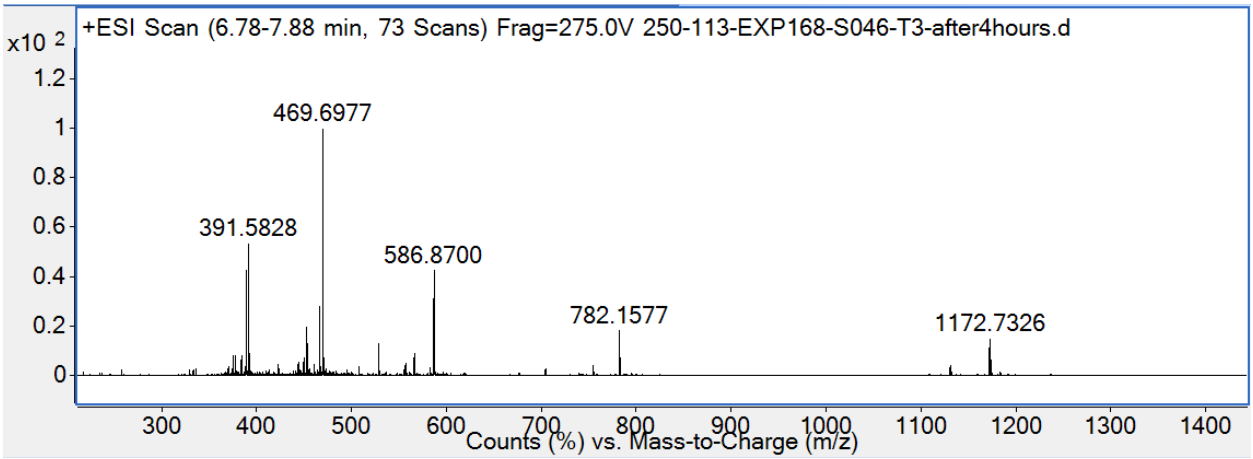
azide-P10
Mass expected: 2357.5 Da
Mass observed: 2358.5 Da
Peptide sequence: KKKKKQBKKKHRWPKAXC



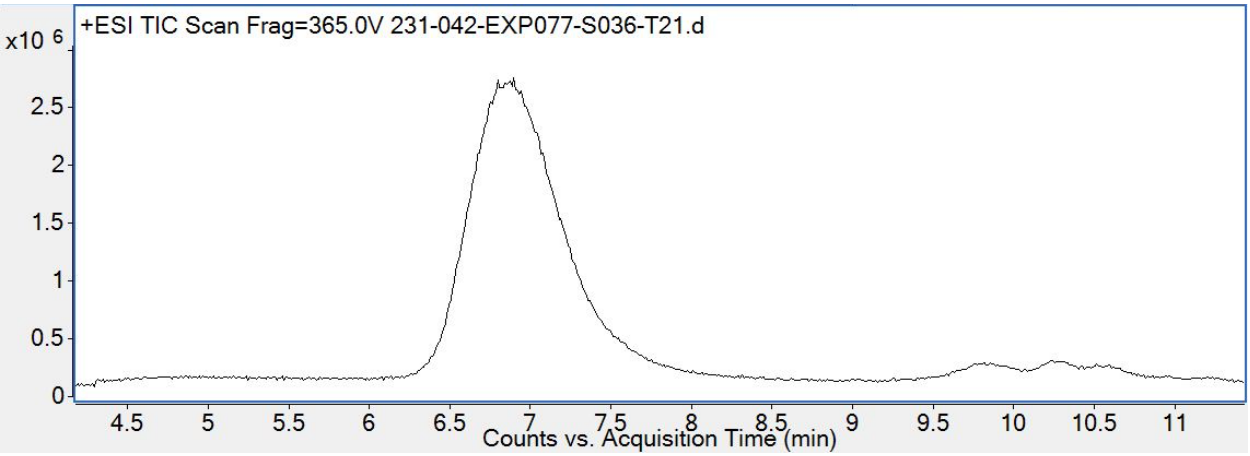


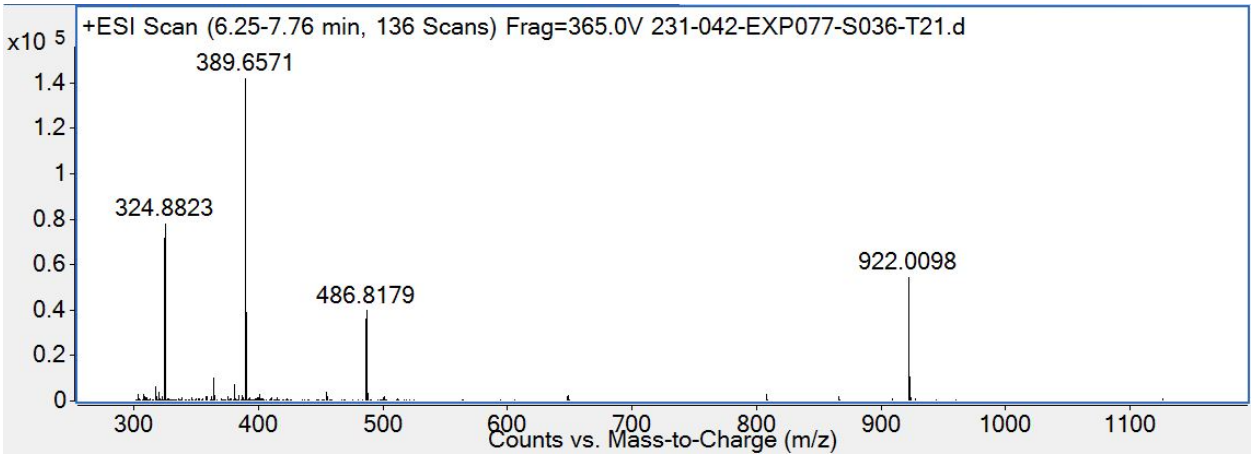
azide-P11
Mass expected: 2342.5 Da
Mass observed: 2343.5 Da
Peptide sequence: KKKKKQBKKKHRWPAXXC



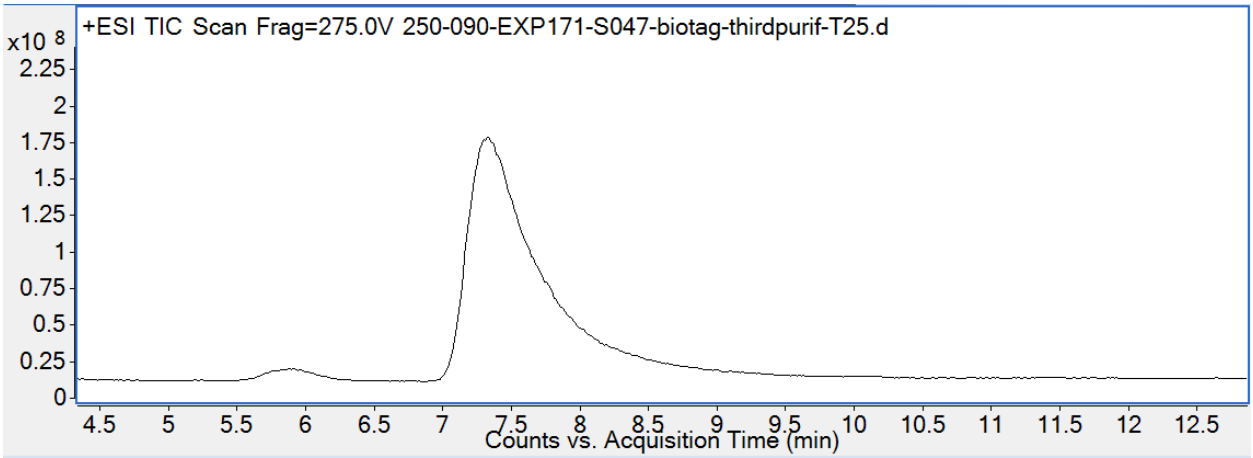


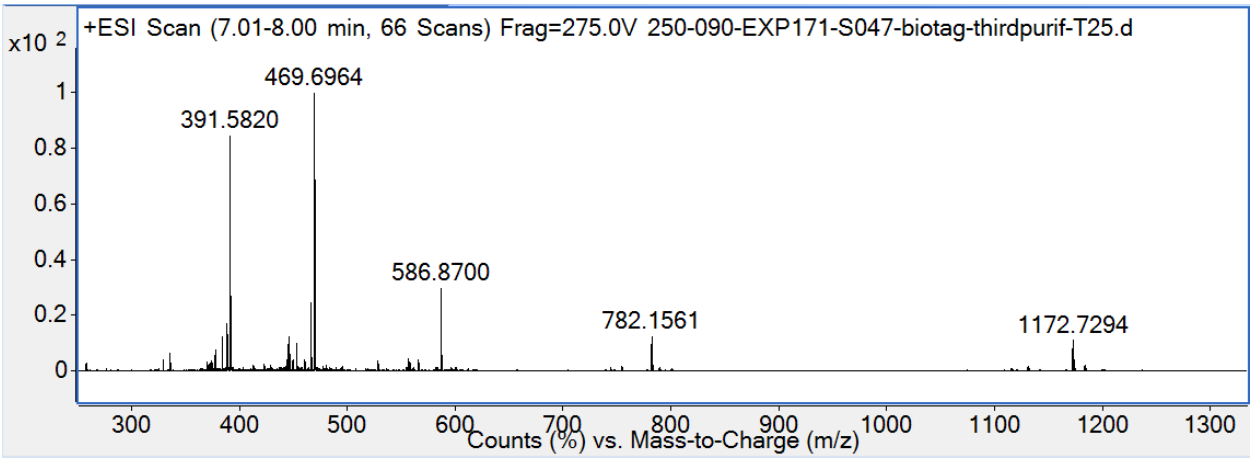
azide-P12
Mass expected: 1943.2 Da
Mass observed: 1943.3 Da
Peptide sequence: KKKKKQBKKKHRWP



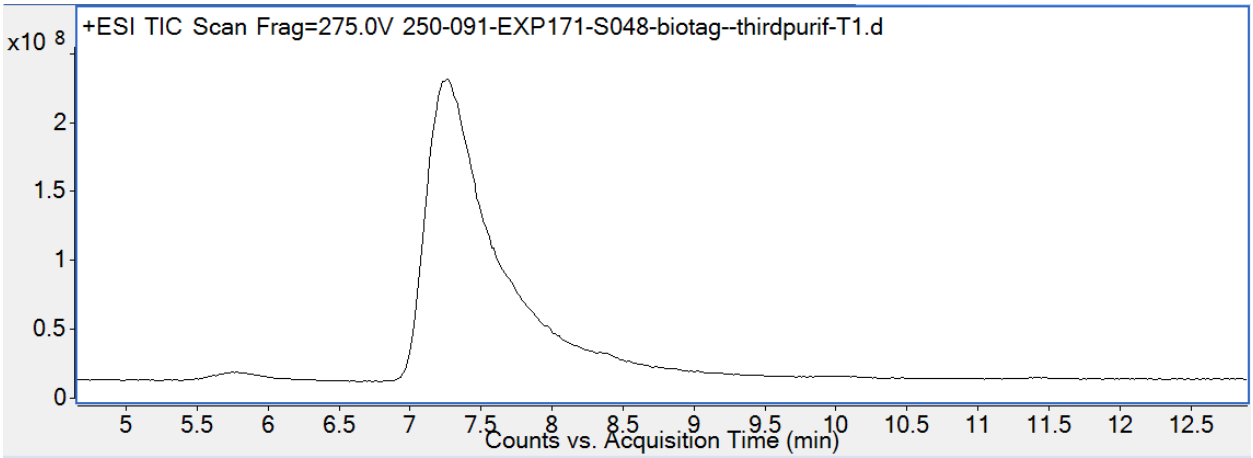


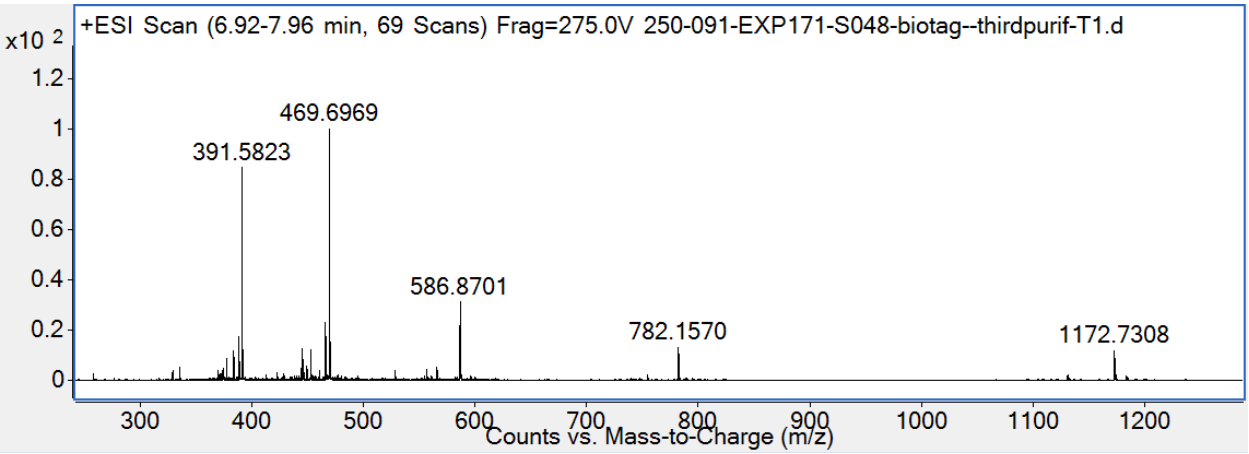
azide-P13
Mass expected: 2342.5 Da
Mass observed: 2343.5 Da
Peptide sequence: KAKKKQBKKKHRWPKXXC



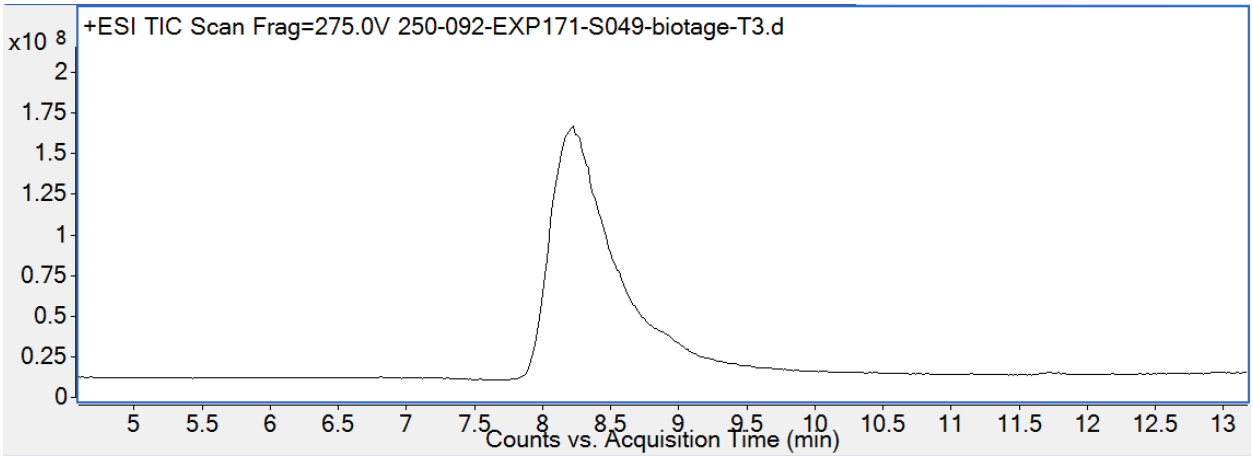


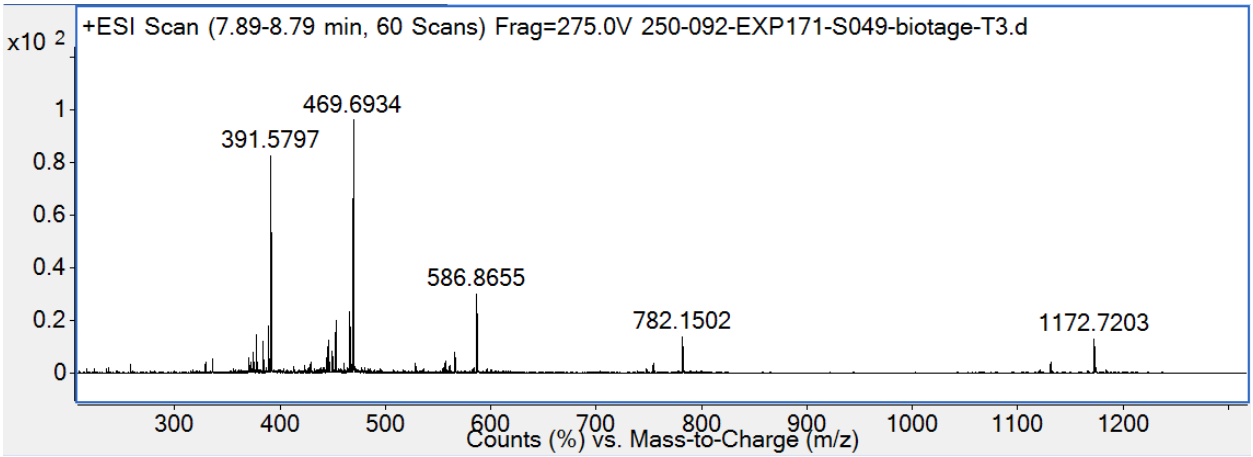
azide-P14
Mass expected: 2342.5 Da
Mass observed: 2343.5 Da
Peptide sequence: KKKKKQBKAKHRWPKXXC



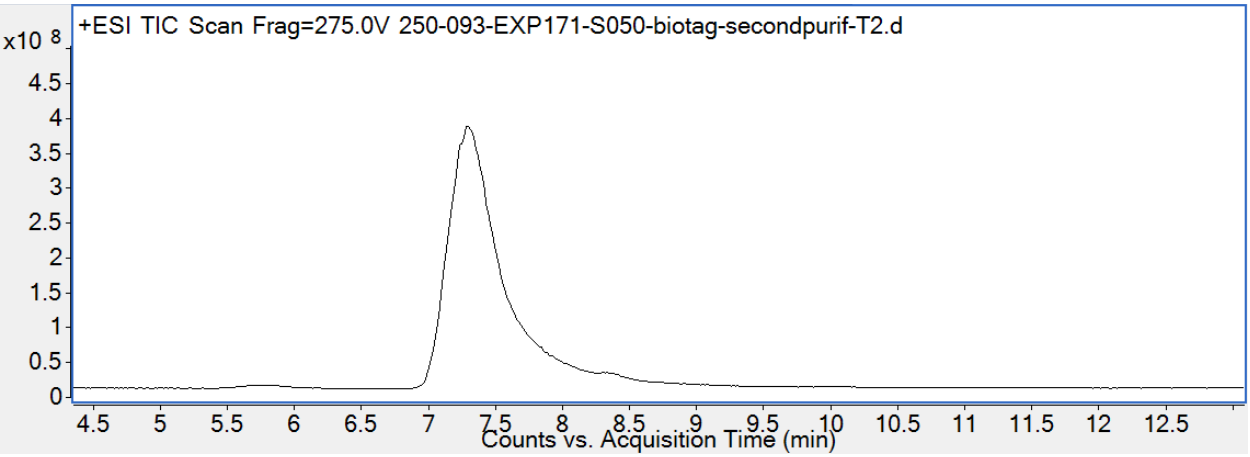


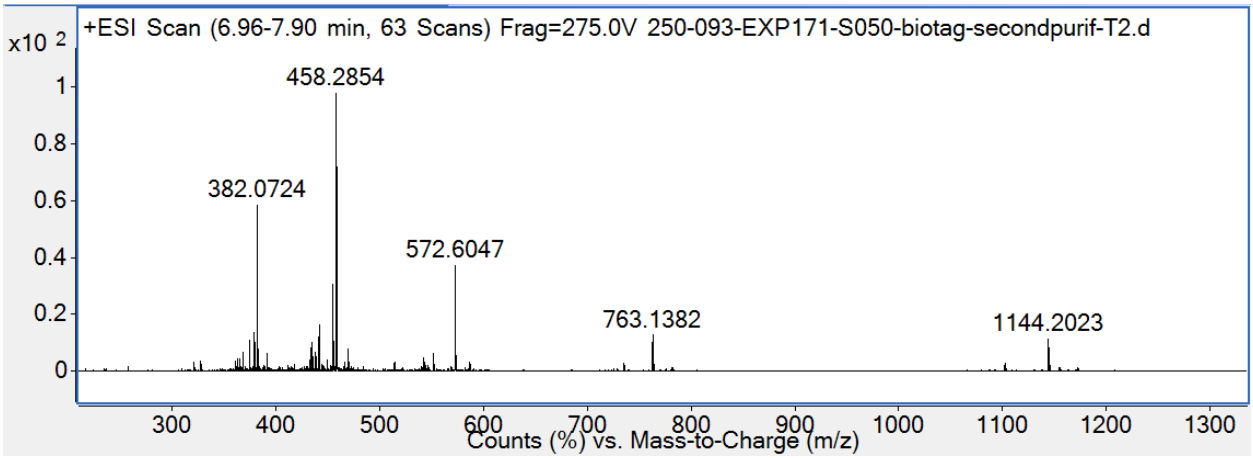
azide-P15
Mass expected: 2342.5 Da
Mass observed: 2343.5 Da
Peptide sequence: KKKAKQBKKKHRWPKXXC



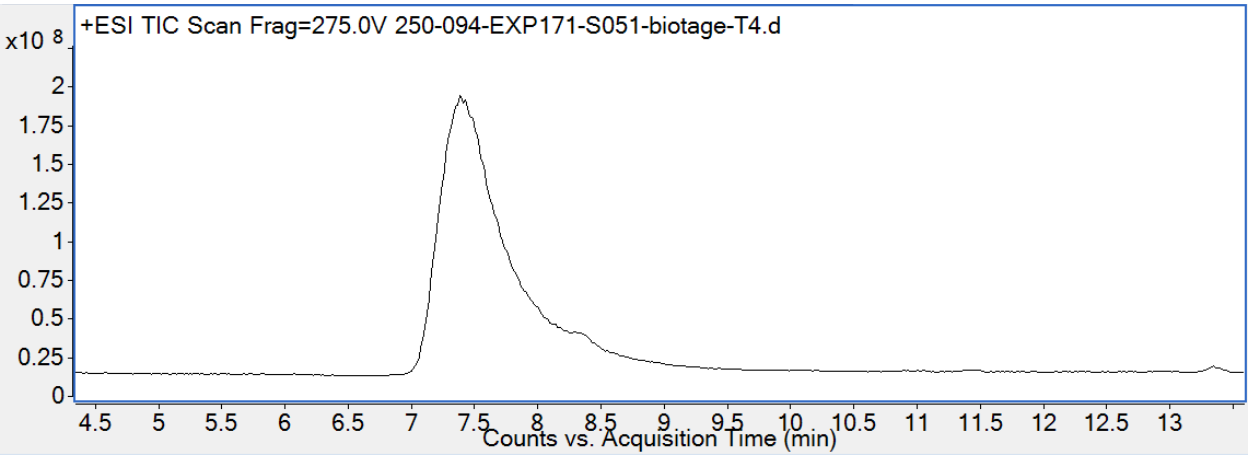


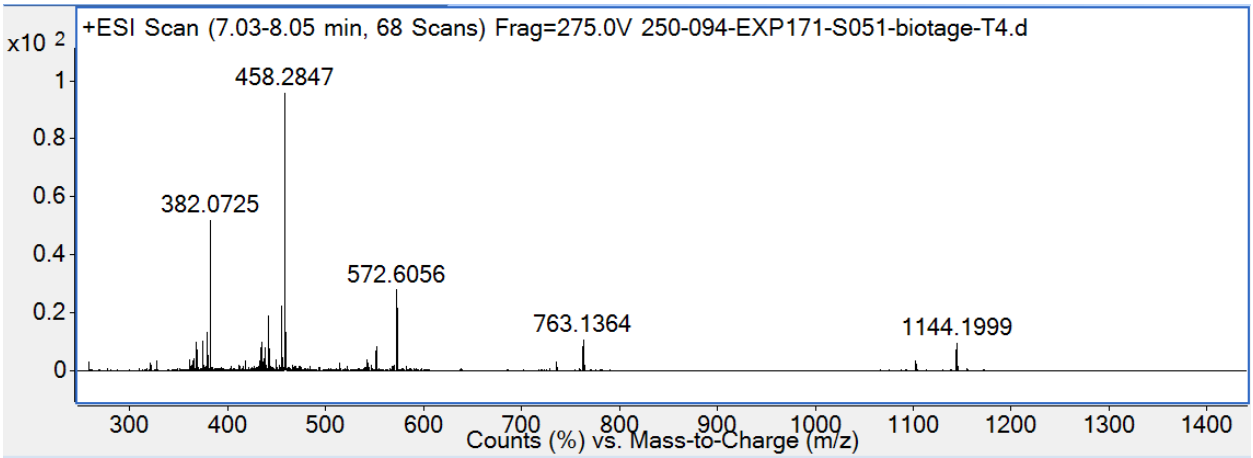
azide-P16
Mass expected: 2285.4 Da
Mass observed: 2286.4 Da
Peptide sequence: KAKKKQBKAKHRWPKXXC



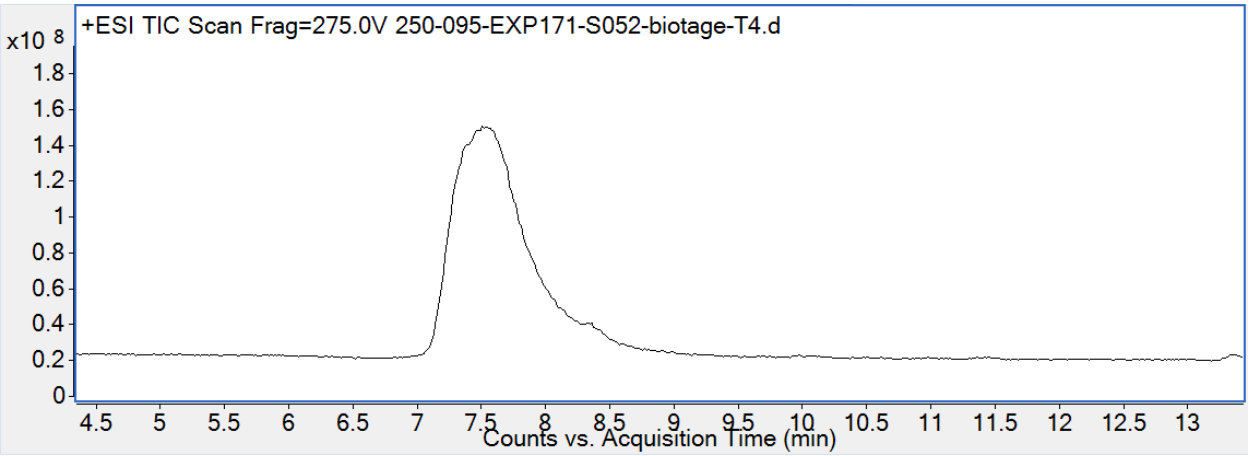


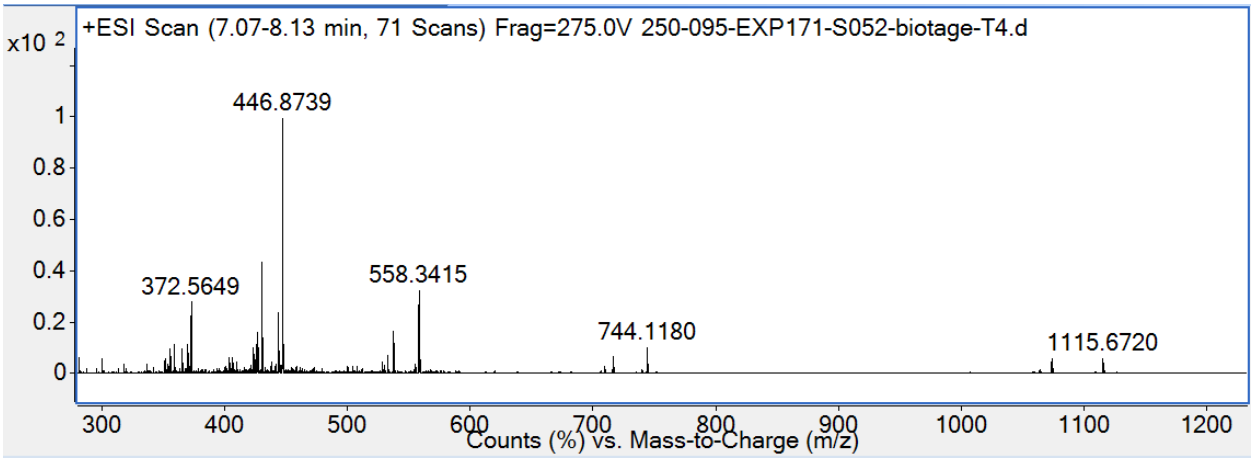
azide-P17
Mass expected: 2285.4 Da
Mass observed: 2286.4 Da
Peptide sequence: KAKAKQBKKKHRWPKXXC





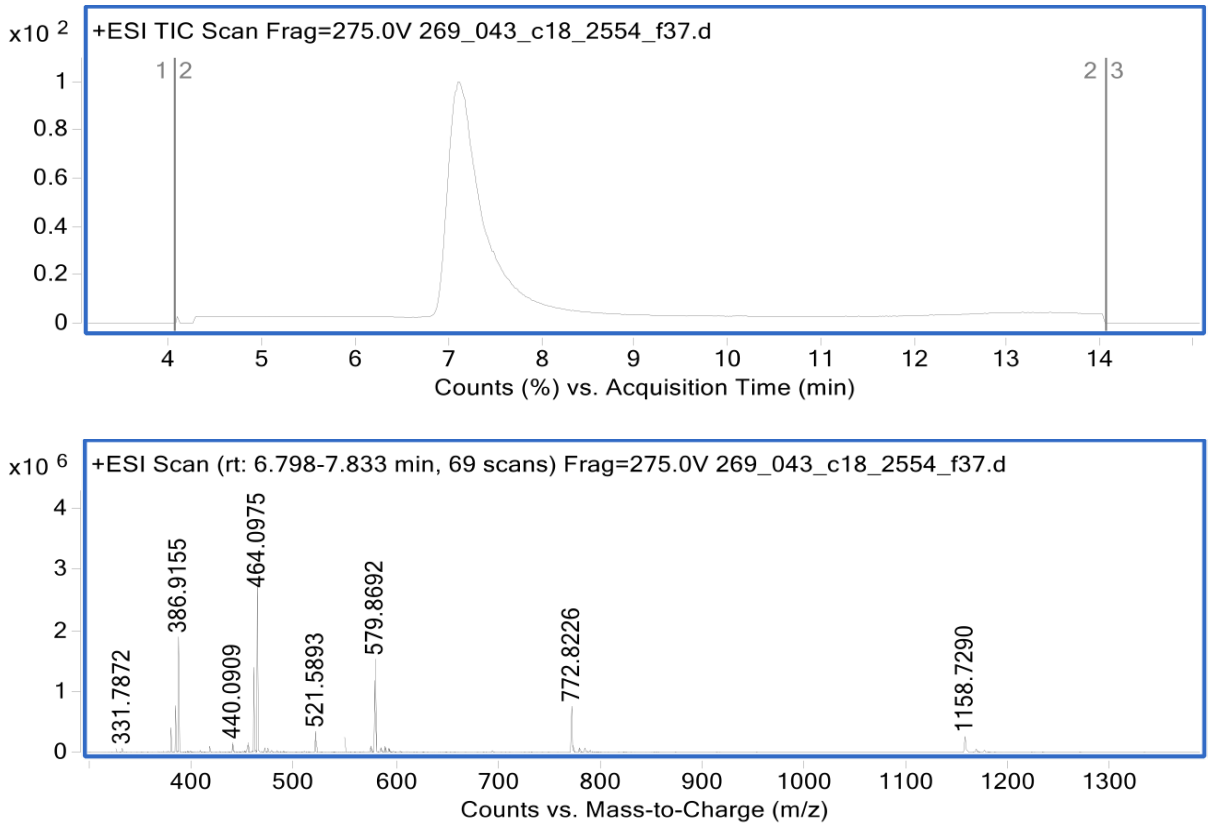
azide-P18
Mass expected: 2228.3 Da
Mass observed: 2229.4 Da
Peptide sequence: KAKAKQBKAKHRWPKXXC



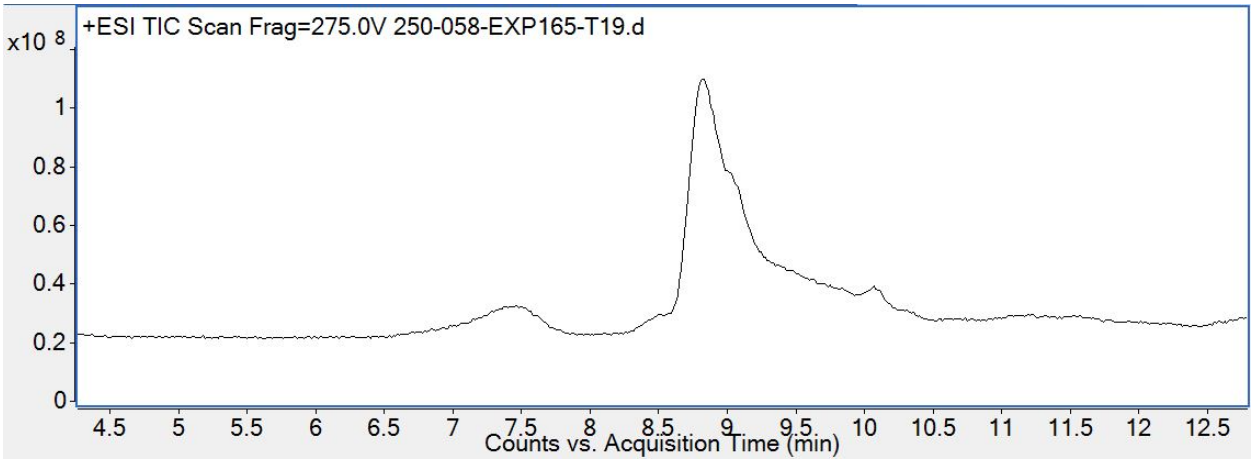


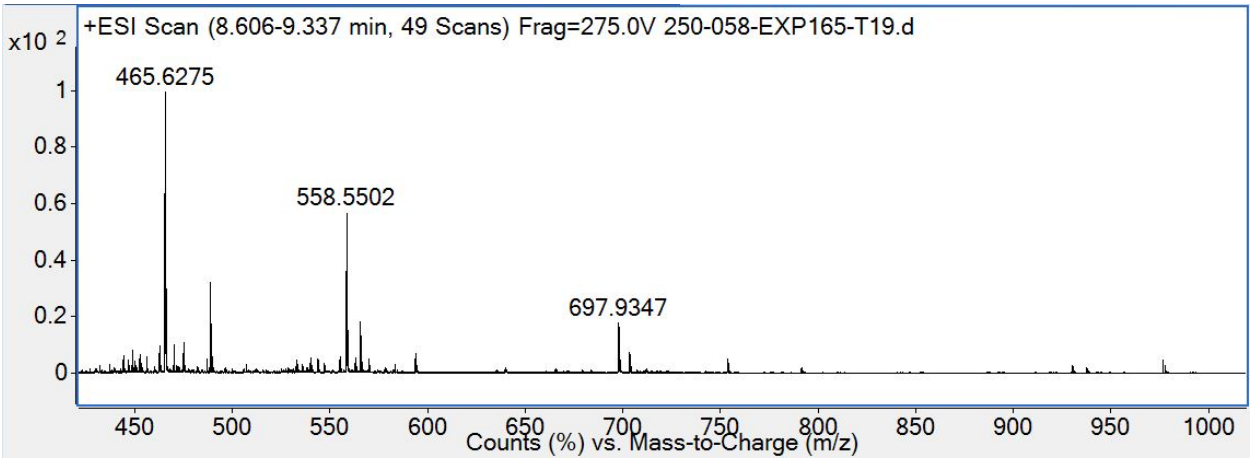
azide-P19

Mass expected: 2314.4 Da
Mass observed: 2314.5 Da
Peptide sequence: KKKKKQBKKKHAWPKXXC

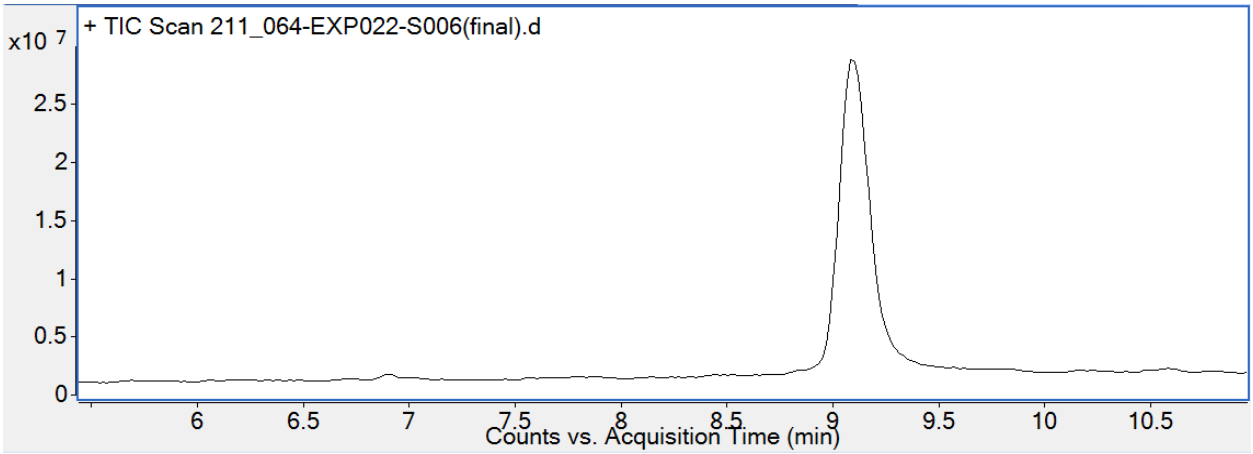


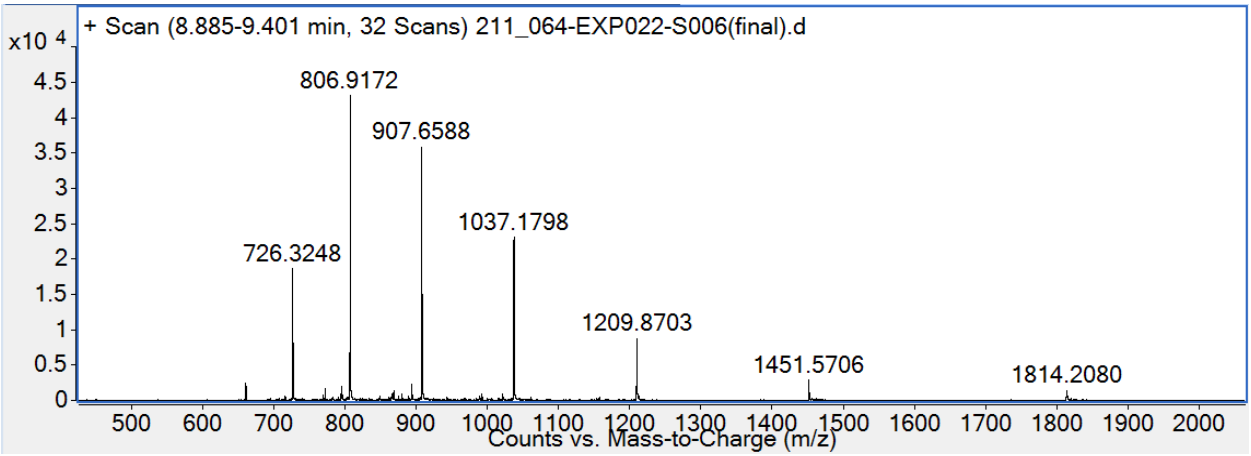
P6-LPSTGG
Mass expected: 2788.5 Da
Mass observed: 2787.8 Da
Peptide sequence: KKKKKQBKKKHRWPKXXCLPSTGG



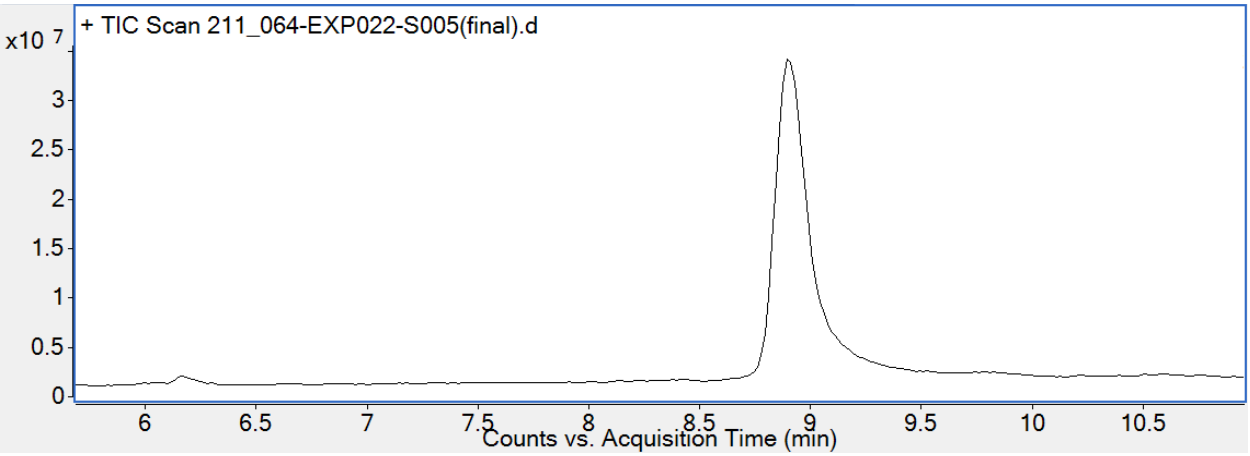


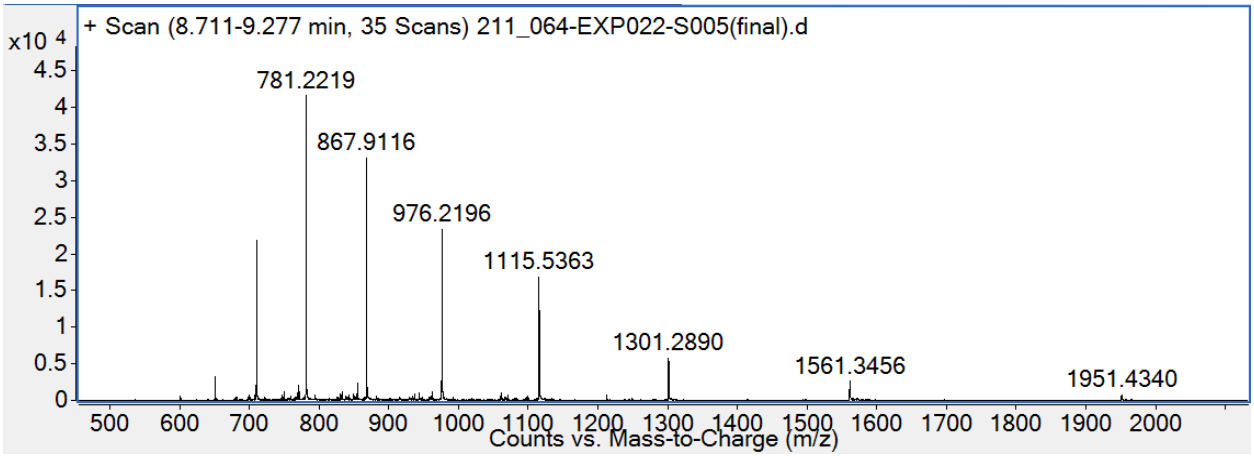
PMO-P1
Mass expected: 7253.9 Da
Mass observed: 7254.0 Da
Peptide sequence: KXKXT



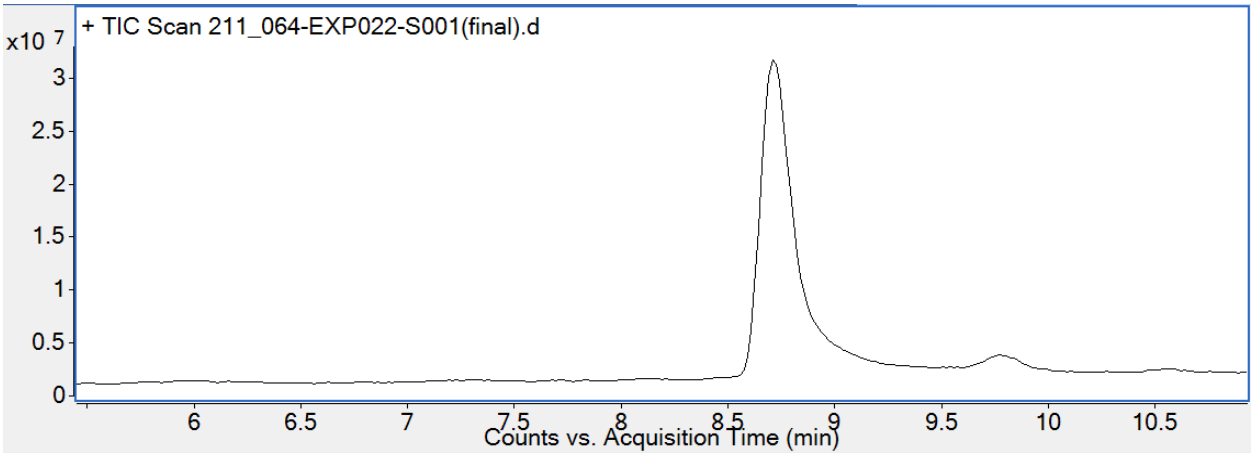


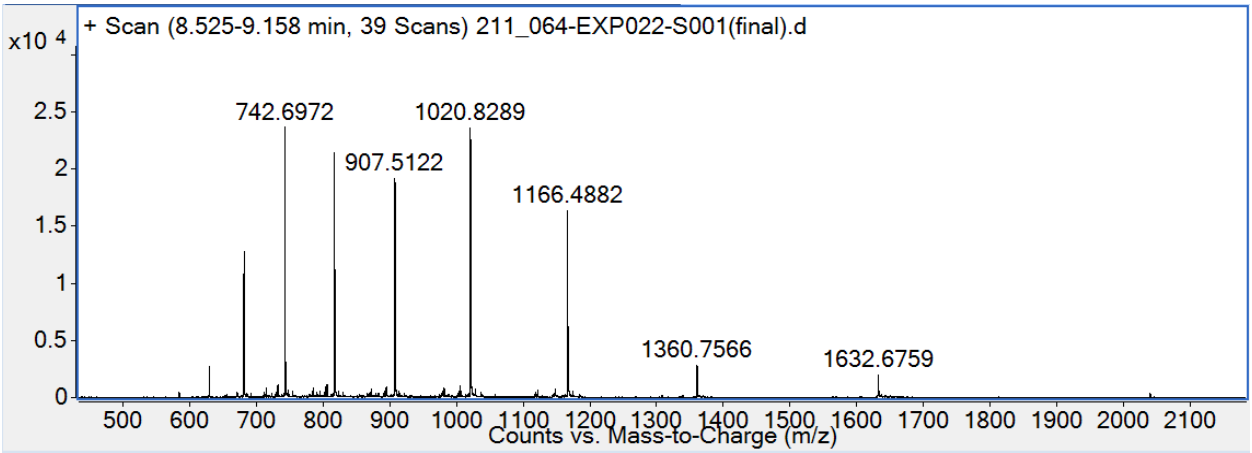
PMO-P2
Mass expected: 7802.5 Da
Mass observed: 7802.7 Da
Peptide sequence: KXKHQQQXK



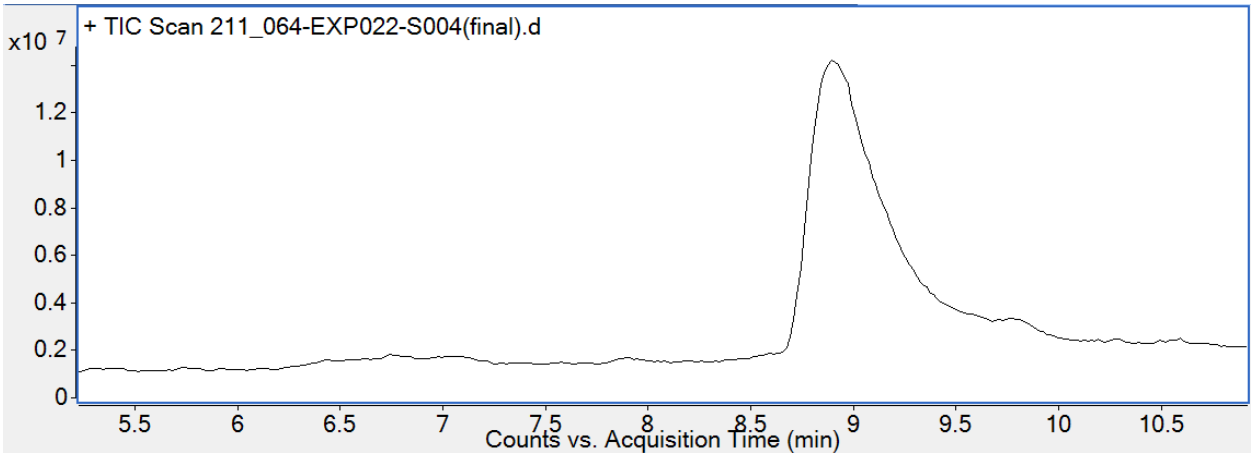


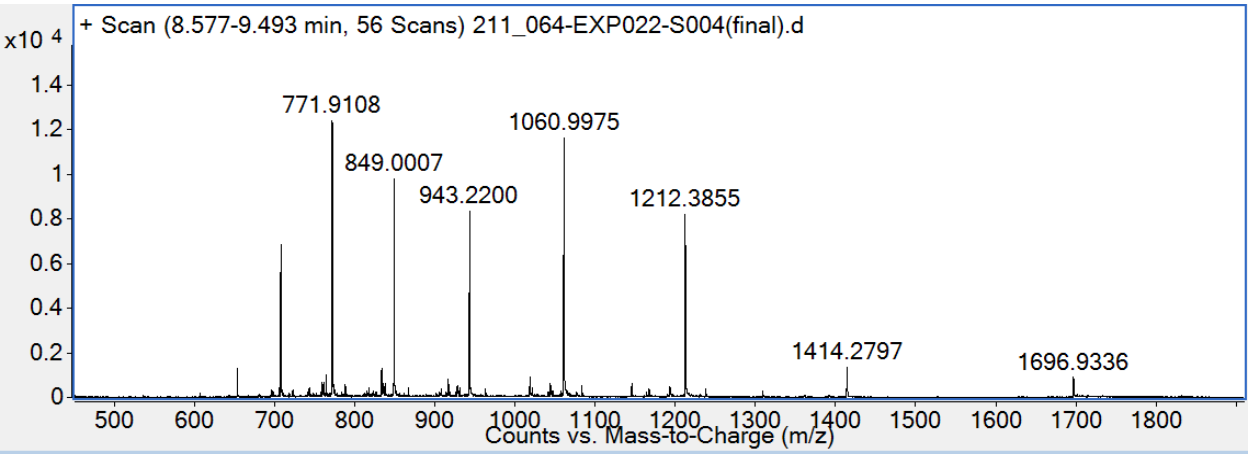
PMO-P3
Mass expected: 8159.0 Da
Mass observed: 8159.1 Da
Peptide sequence: KXKKQQGKKKHR



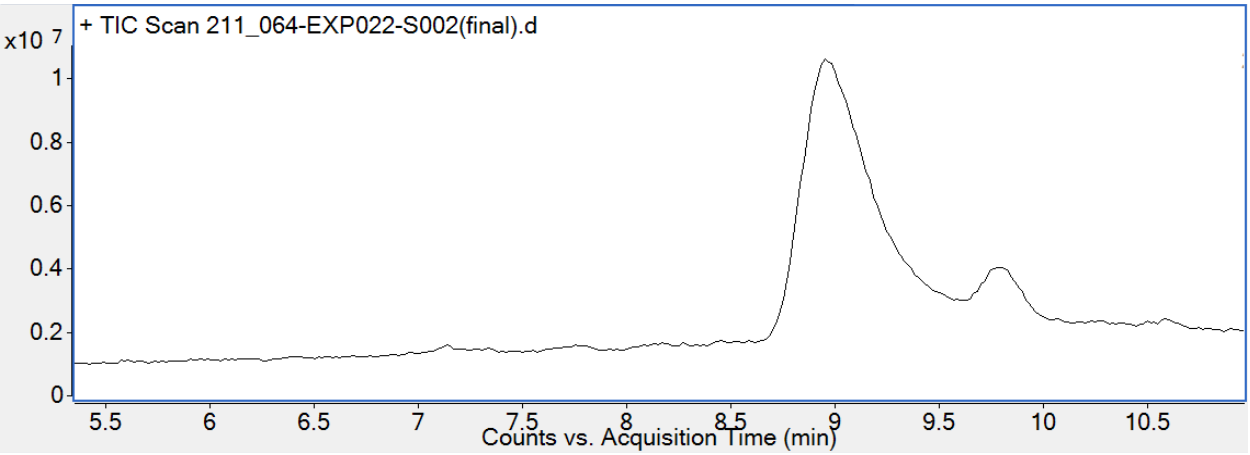


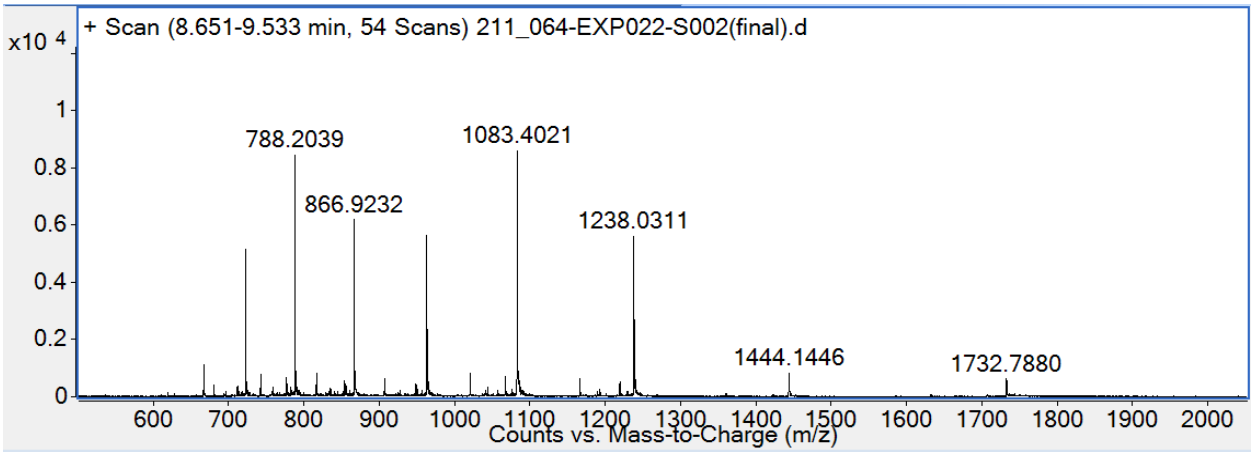
PMO-P4
Mass expected: 8480.4 Da
Mass observed: 8480.4 Da
Peptide sequence: HKKKKQBKKKHRWP



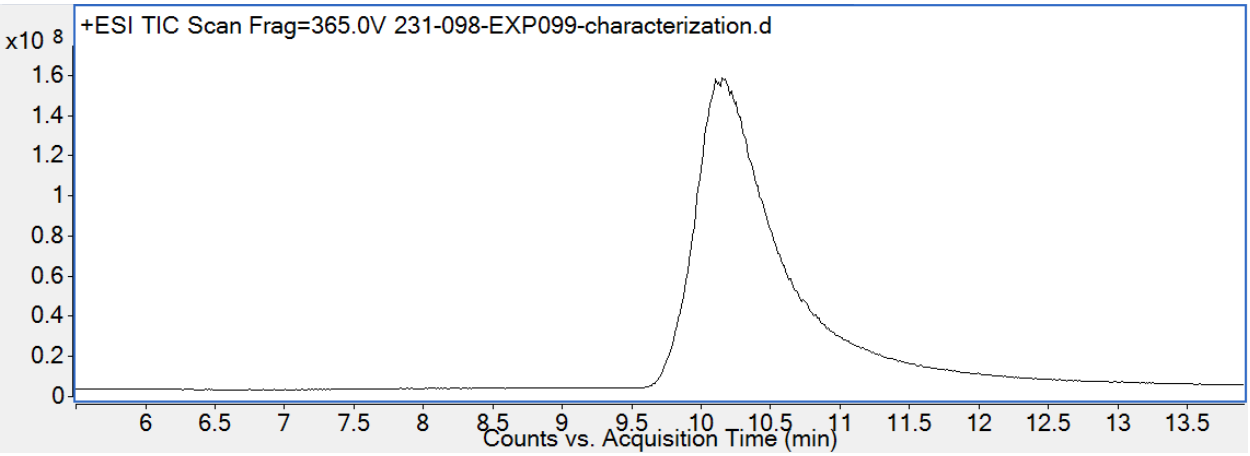


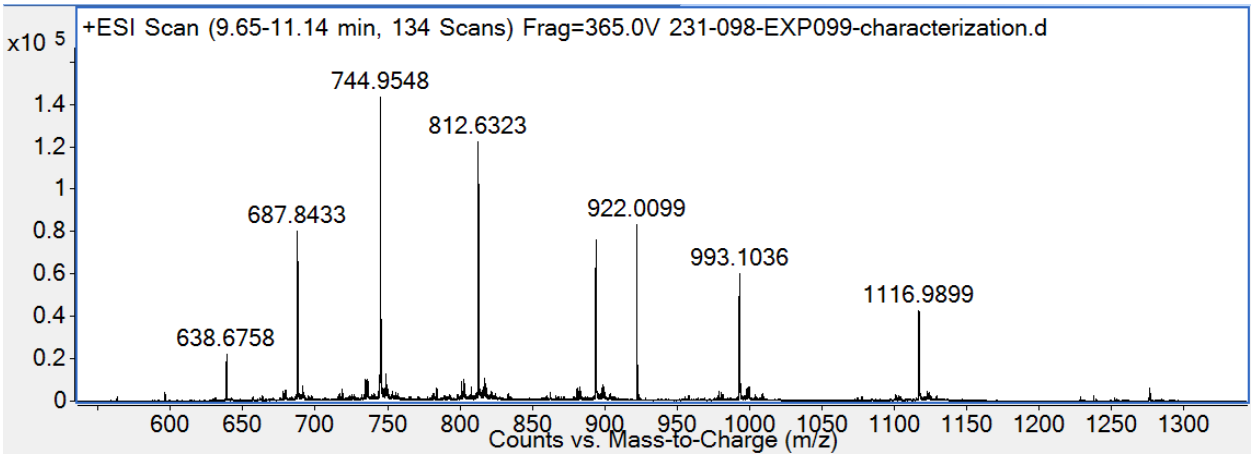
PMO-P5
Mass expected: 8659.7 Da
Mass observed: 8659.6 Da
Peptide sequence: KKKKKQBKKKHRWPMG



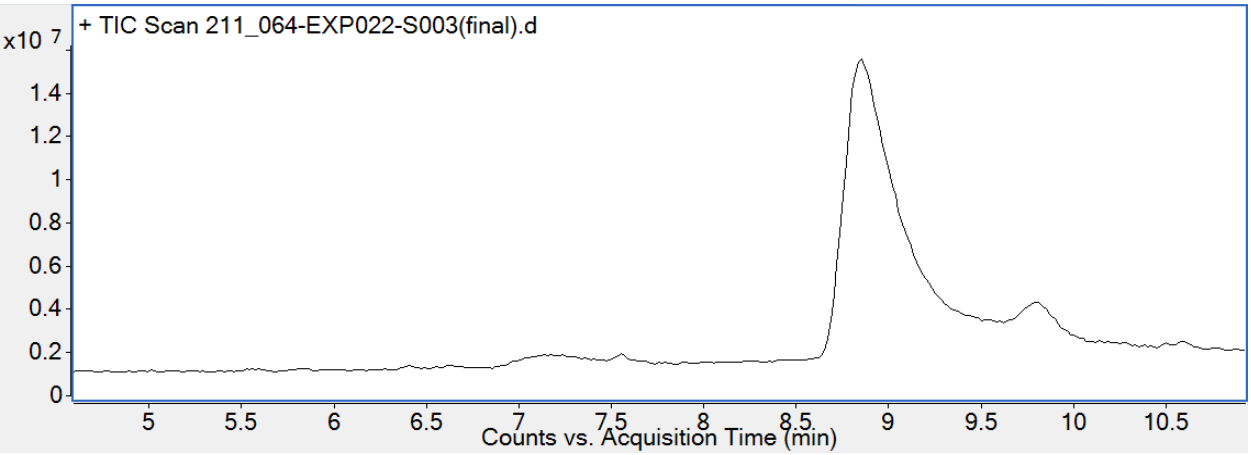


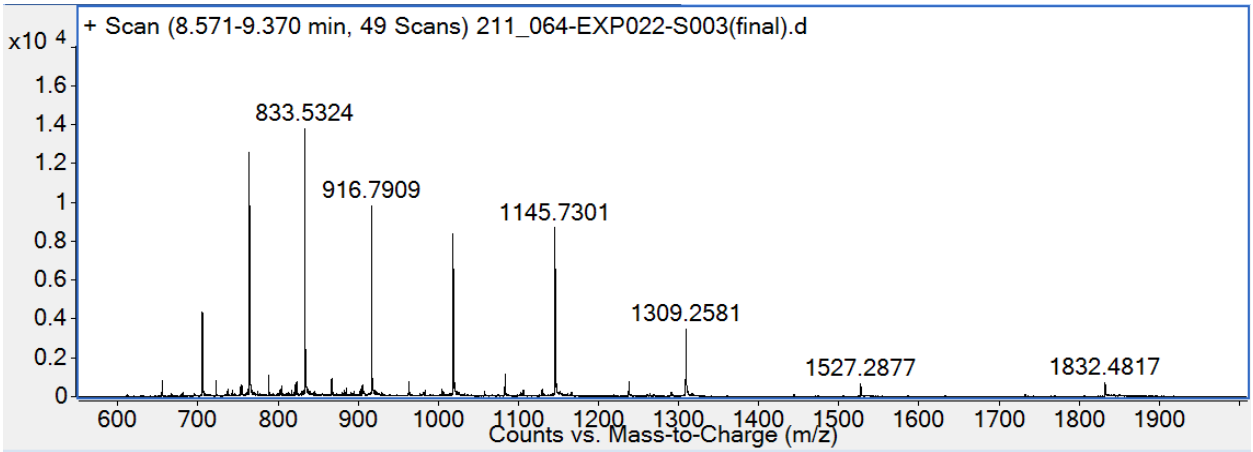
PMO-P6
Mass expected: 8929.1 Da
Mass observed: 8929.1 Da
Peptide sequence: KKKKKQBKKKHRWPKXXC



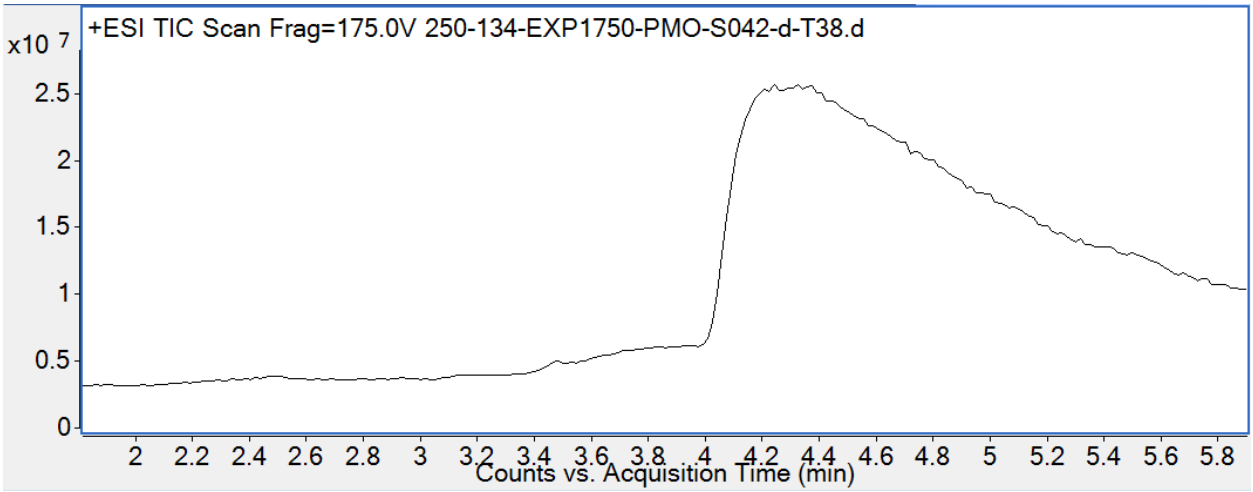


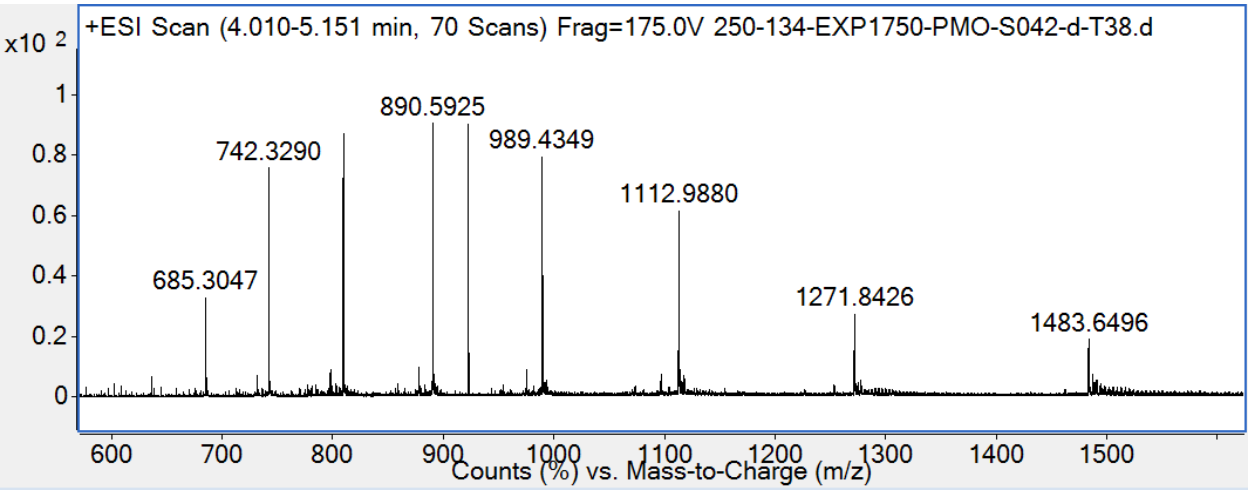
PMO-P7
Mass expected: 9158.3 Da
Mass observed: 9158.3 Da
Peptide sequence: KKKKNQBKKKHRWPMKXCPQ



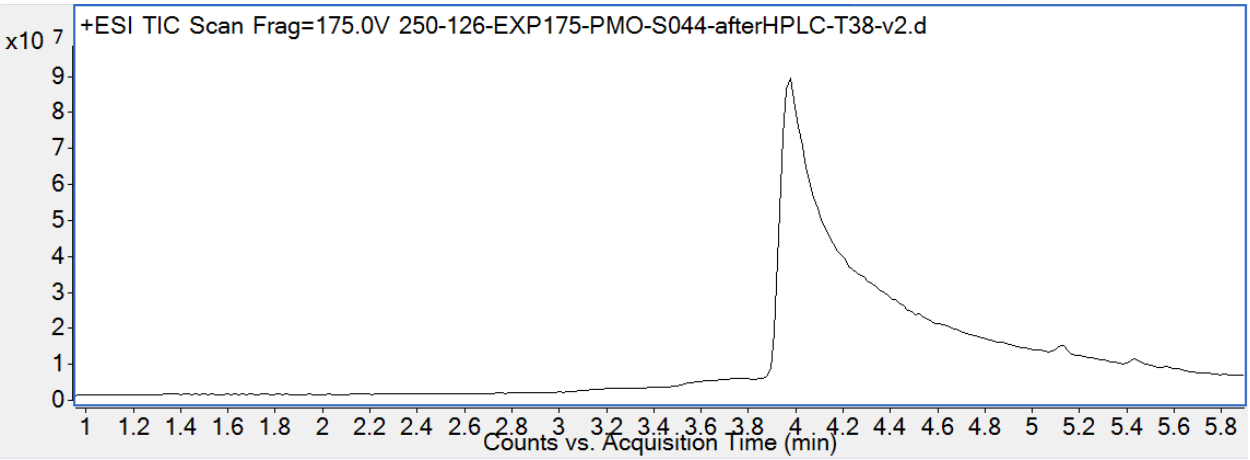


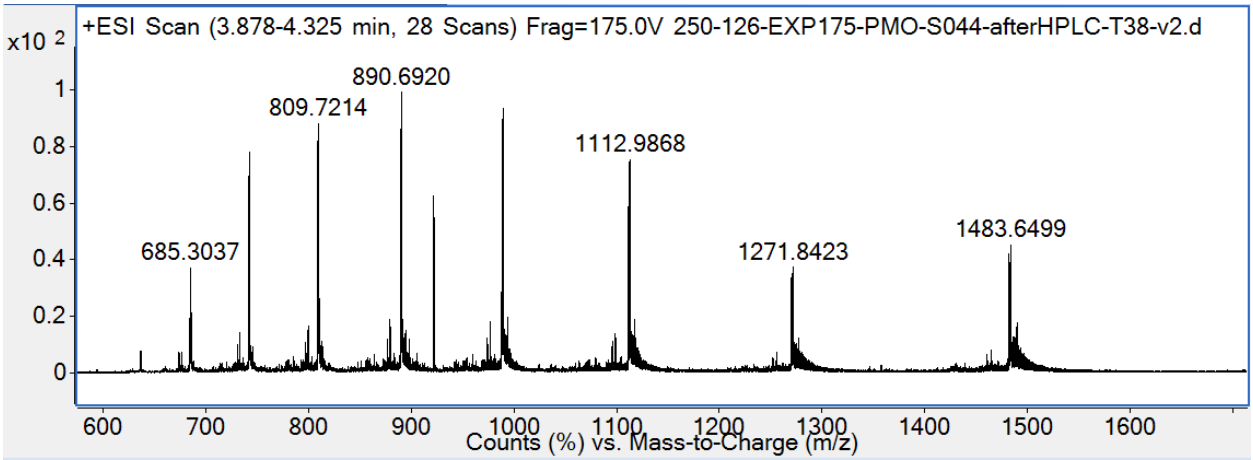
PMO-P8
Mass expected: 8895.5 Da
Mass observed: 8897.1 Da
Peptide sequence: KKKKKQBKKKHRWPKXXA



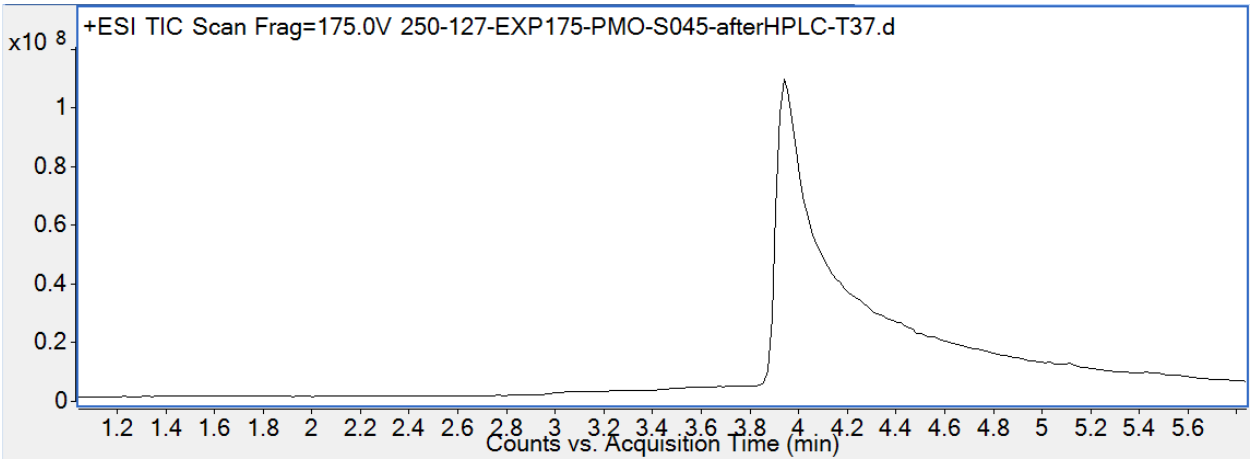


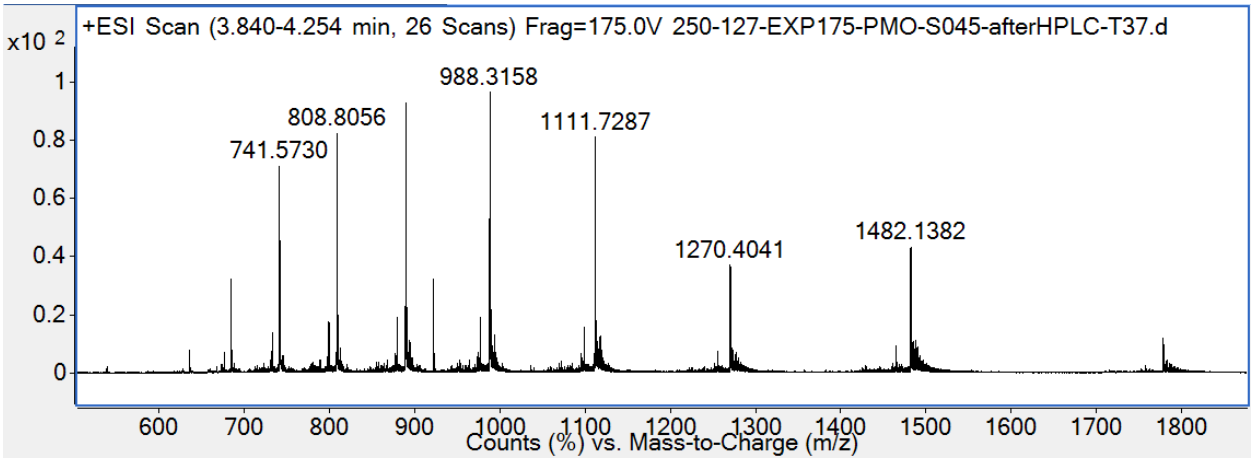
PMO-P9
Mass expected: 8885.5 Da
Mass observed: 8887.1 Da
Peptide sequence: KKKKKQBKKKHRWPKXAC



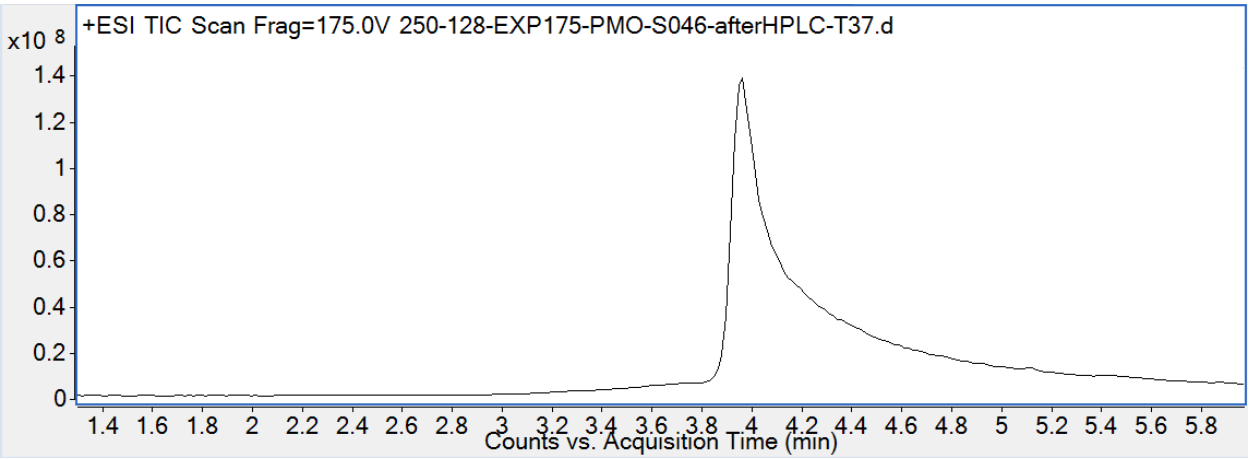


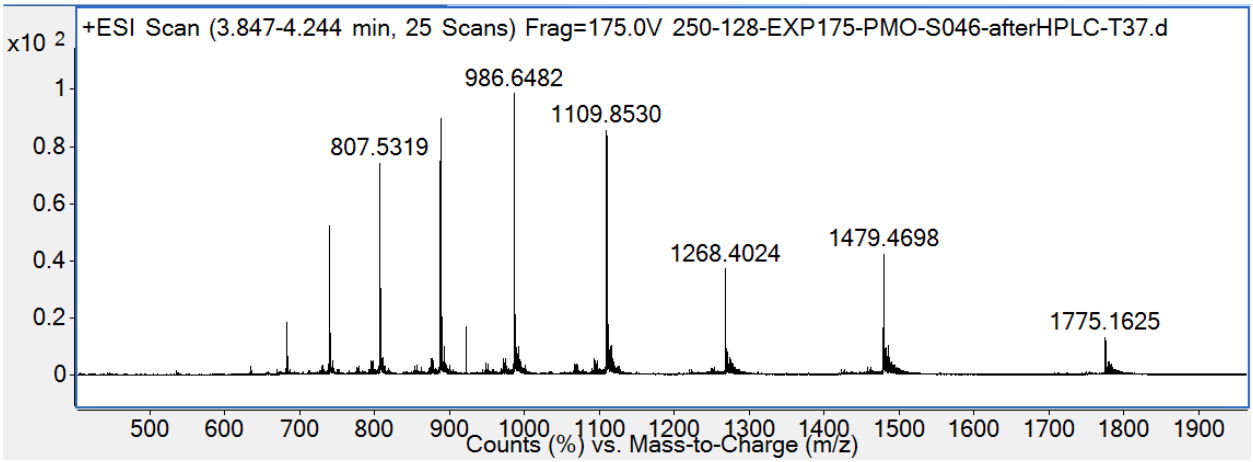
PMO-P10
Mass expected: 8885.5 Da
Mass observed: 8887.1 Da
Peptide sequence: KKKKKQBKKKHRWPKAXC



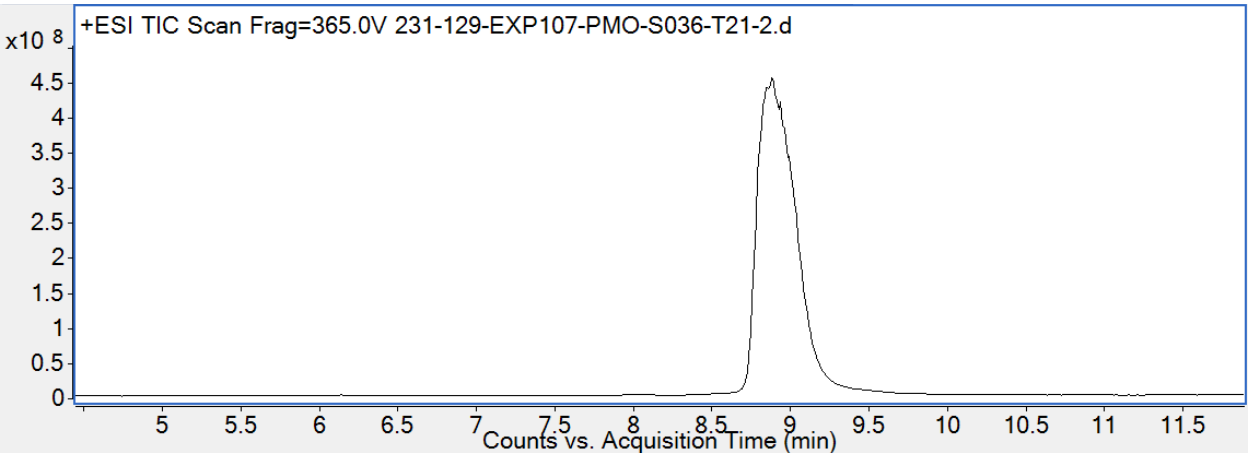


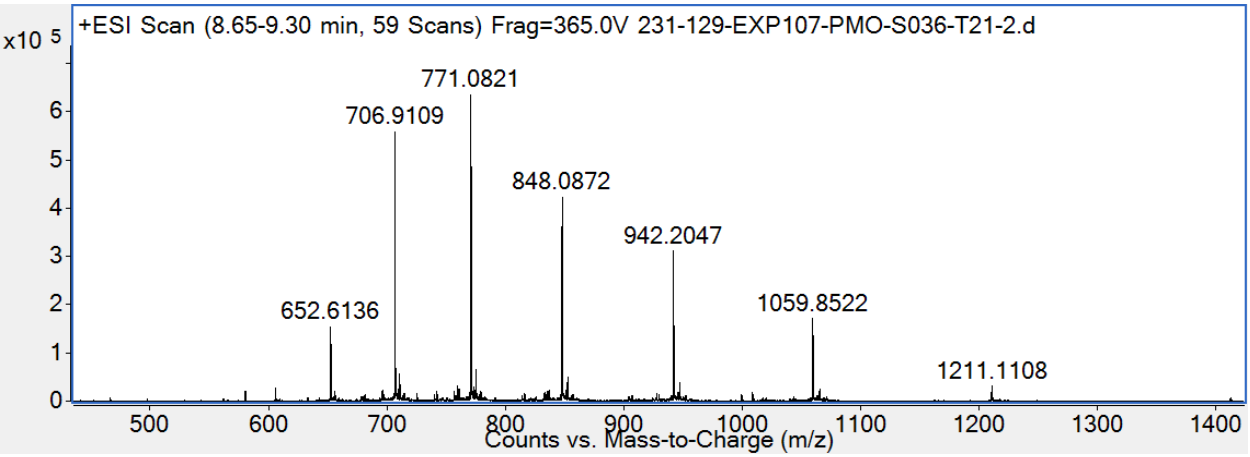
PMO-P11
Mass expected: 8870.5 Da
Mass observed: 8872.2 Da
Peptide sequence: KKKKKQBKKKHRWPAXXC



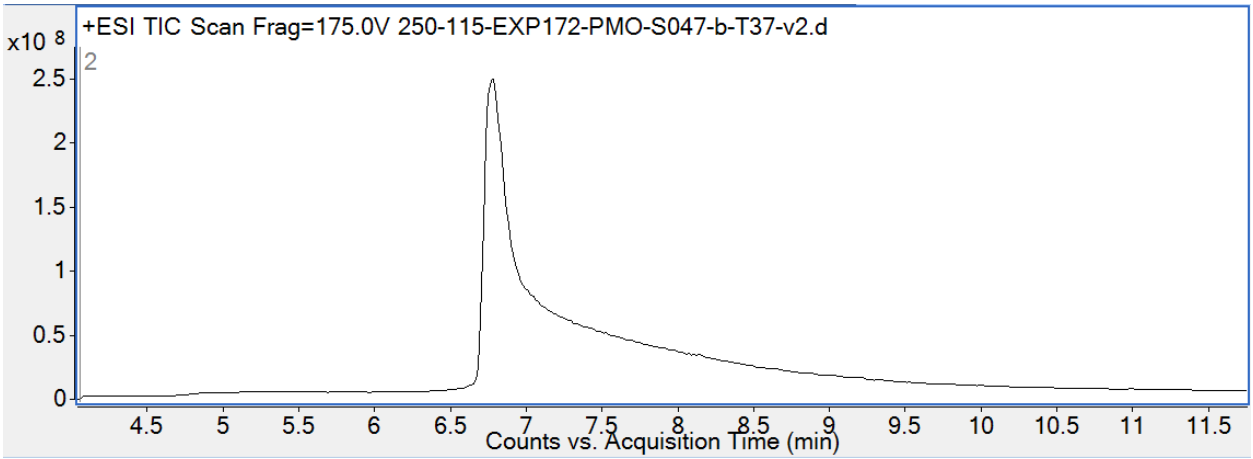


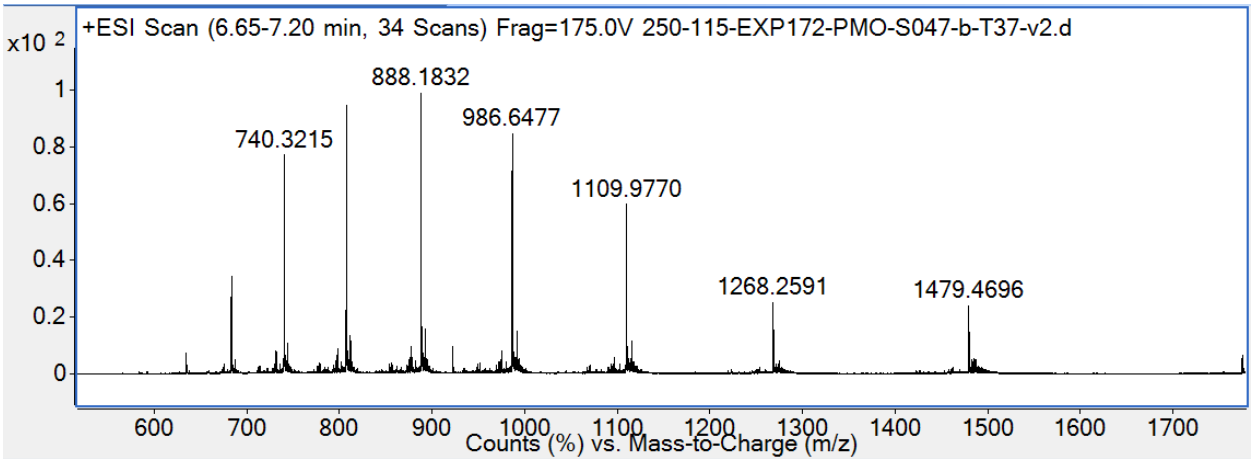
PMO-P12
Mass expected: 8471.4 Da
Mass observed: 8472.1 Da
Peptide sequence: KKKKKQBKKKHRWP



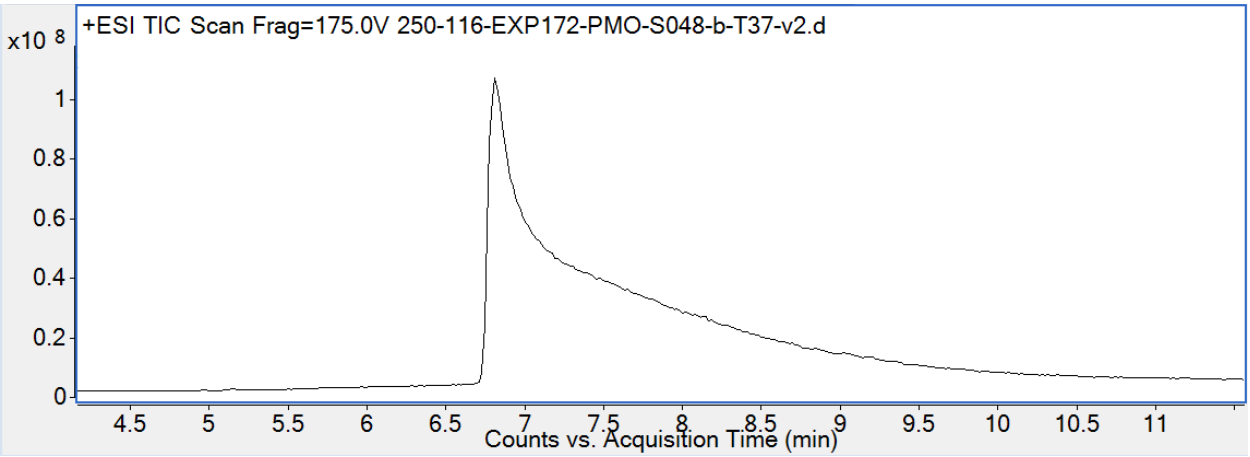


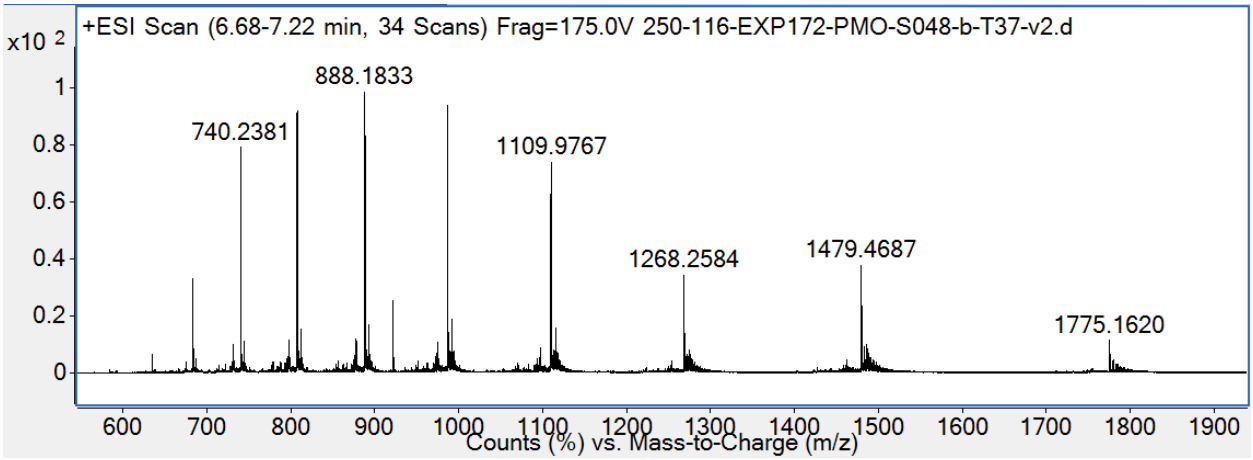
PMO-P13
Mass expected: 8871.5 Da
Mass observed: 8872.2 Da
Peptide sequence: KAKKKQBKKKHRWPKXXC



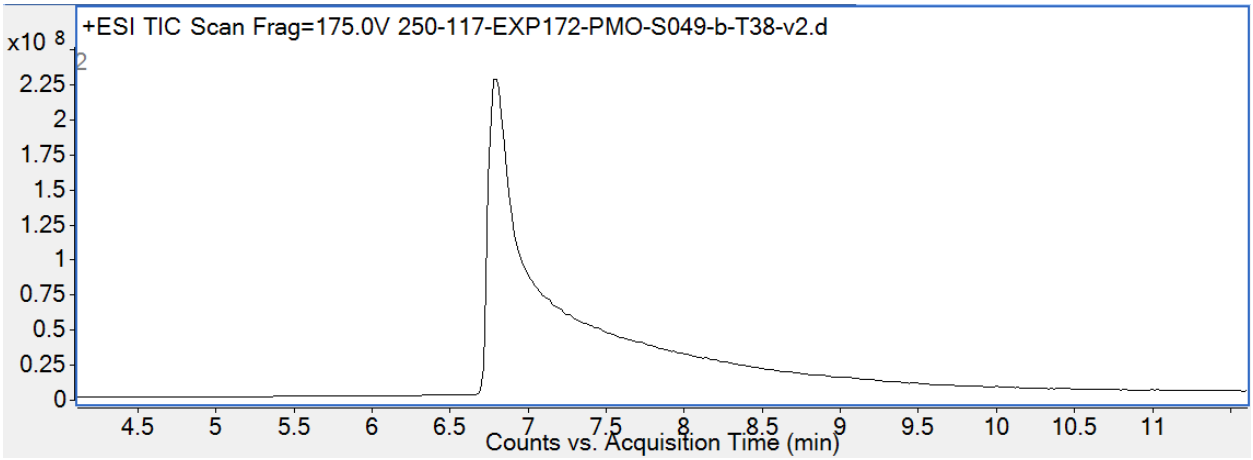


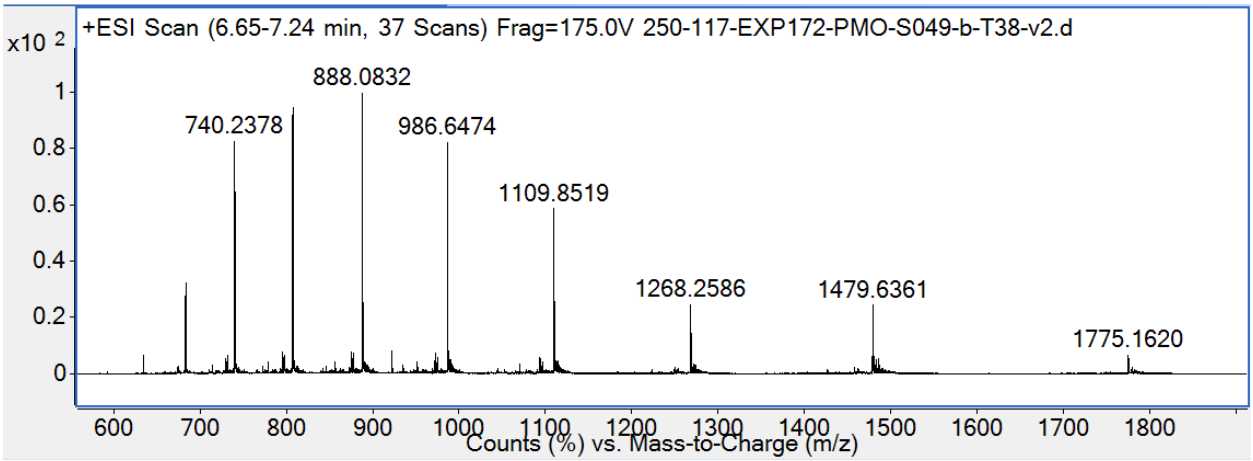
PMO-P14
Mass expected: 8871.5 Da
Mass observed: 8872.2 Da
Peptide sequence: KKKKKQBKAKHRWPKXXC



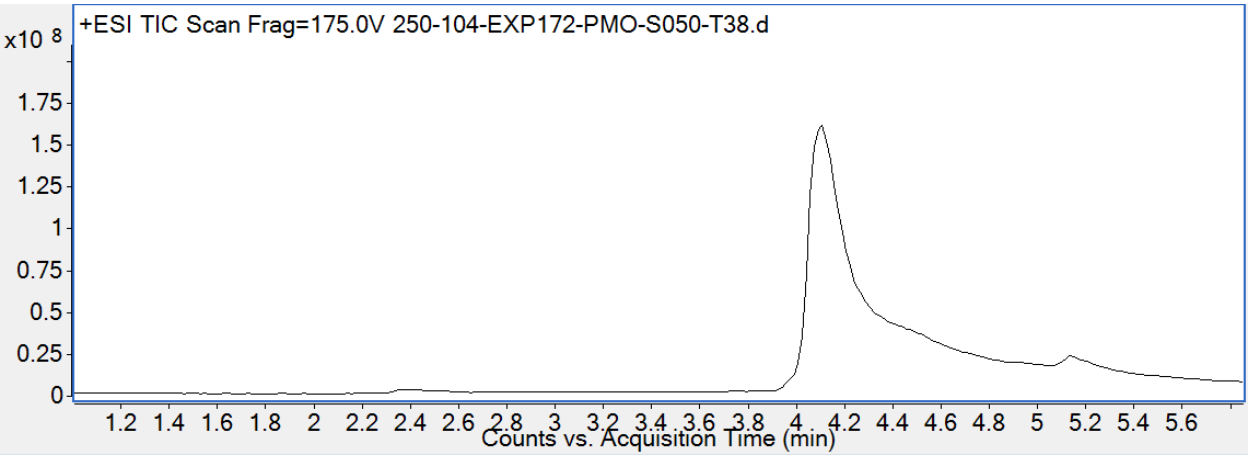


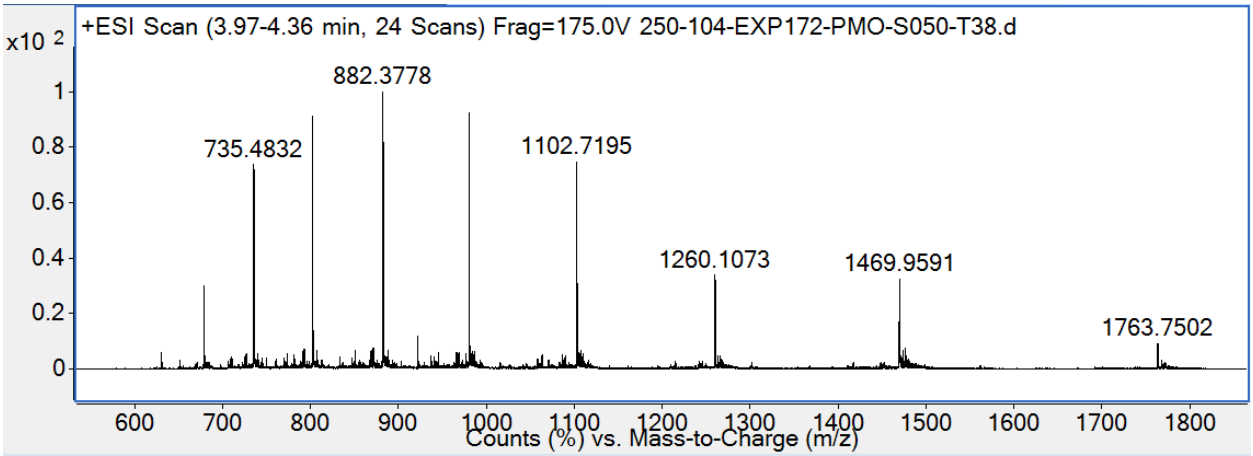
PMO-P15
Mass expected: 8871.5 Da
Mass observed: 8872.2 Da
Peptide sequence: KKKAKQBKKKHRWPKXXC



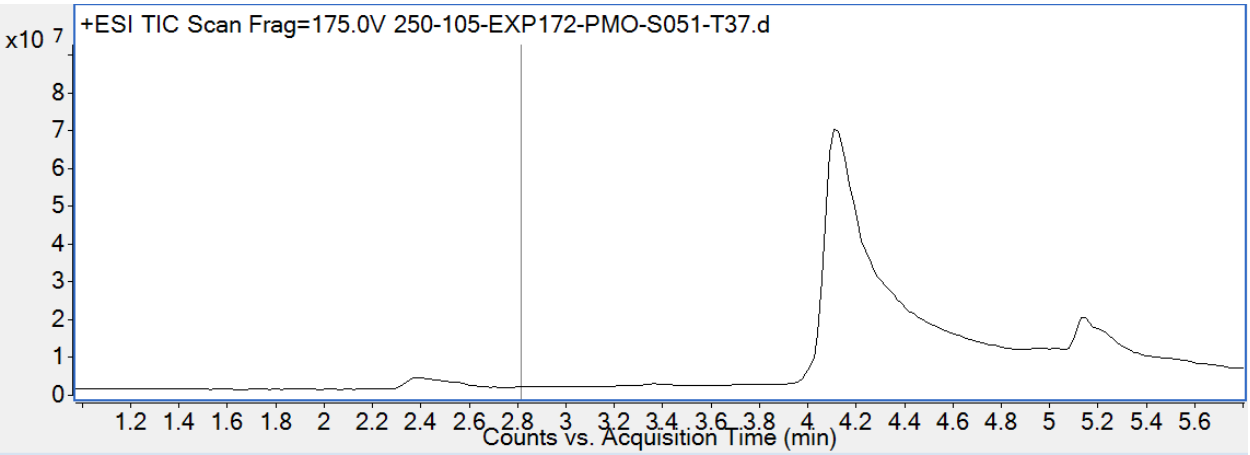


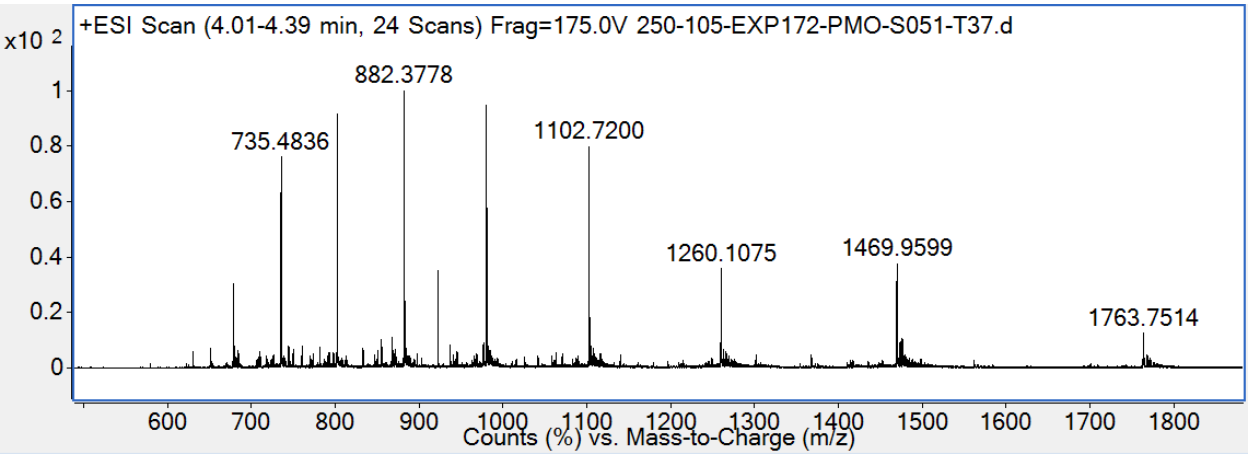
PMO-P16
Mass expected: 8814.4 Da
Mass observed: 8815.2 Da
Peptide sequence: KAKKKQBKAKHRWPKXXC



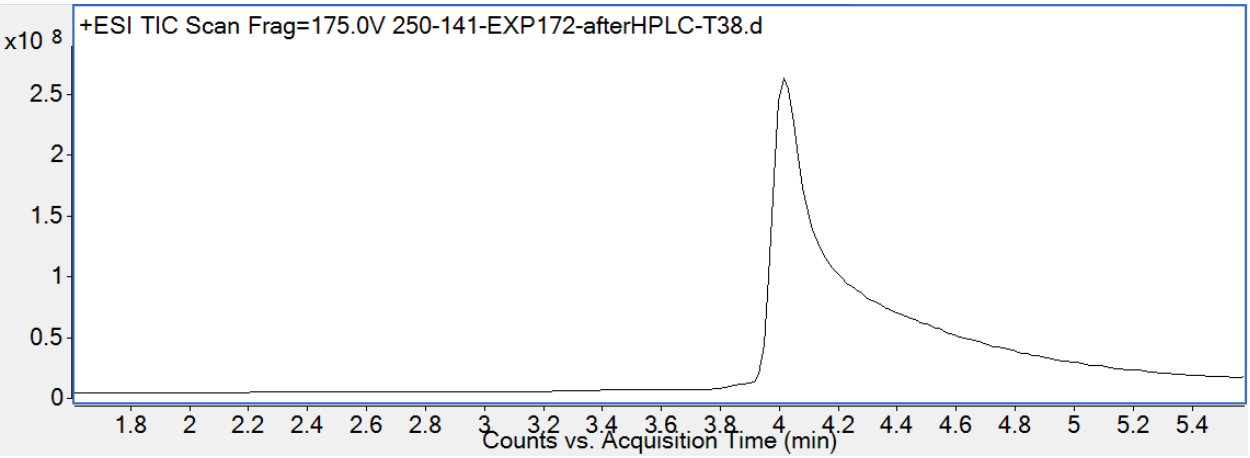


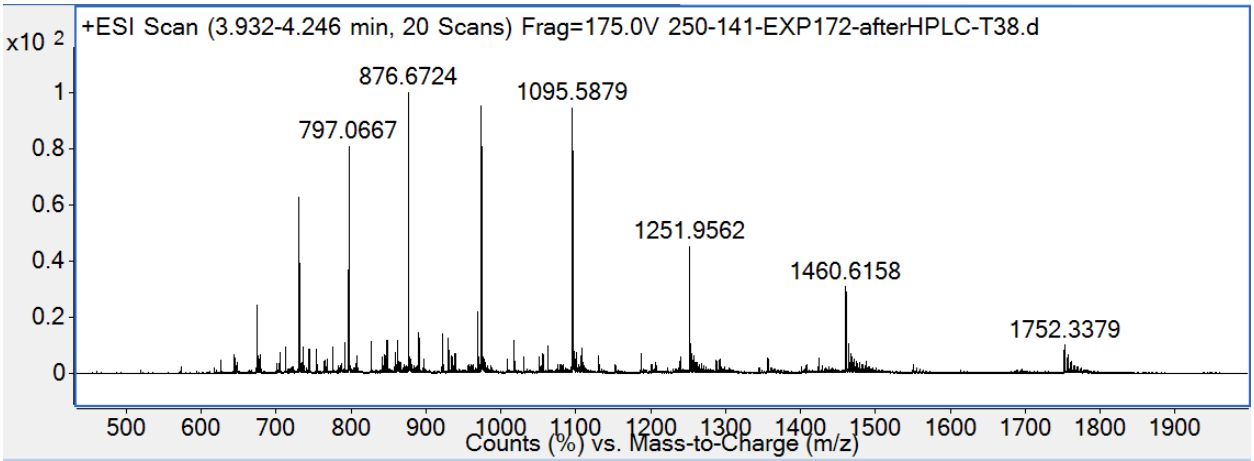
PMO-P17
Mass expected: 8814.4 Da
Mass observed: 8815.1 Da
Peptide sequence: KAKAKQBKKKHRWPKXXC



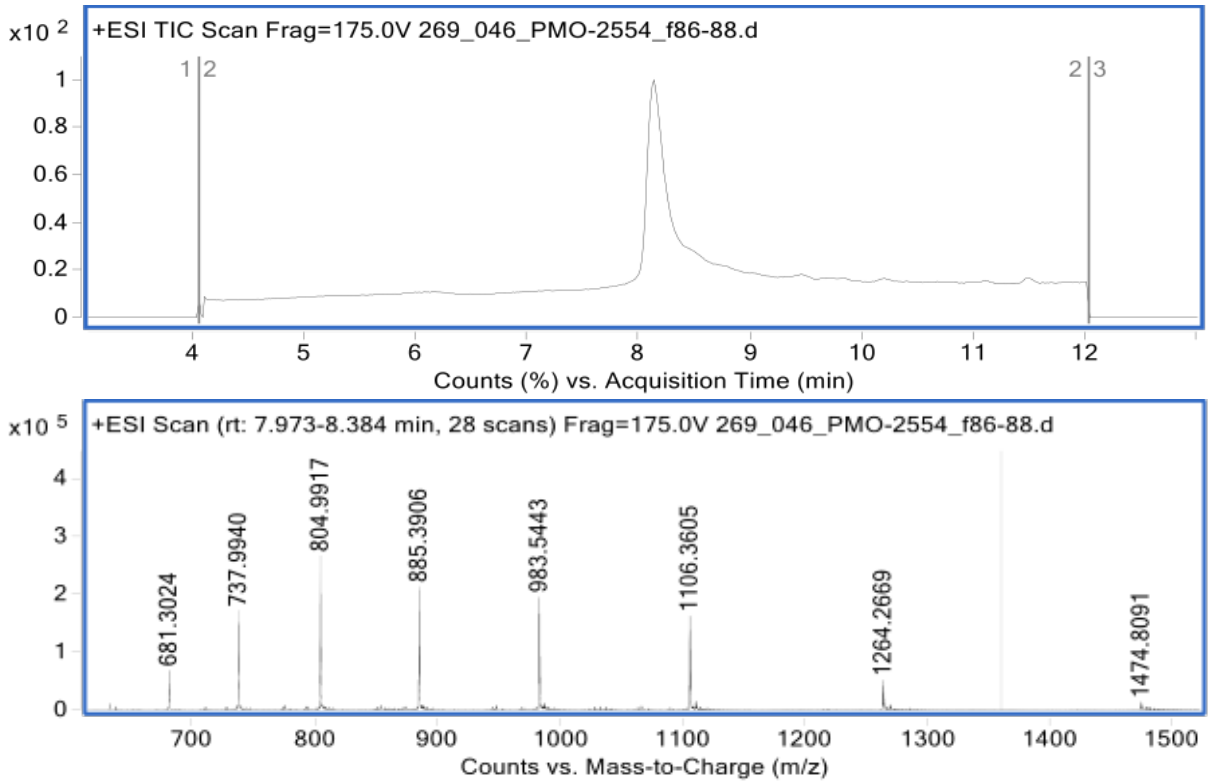


PMO-P18
Mass expected: 8757.4 Da
Mass observed: 8756.7 Da
Peptide sequence: KAKAKQBKAKHRWPKXXC

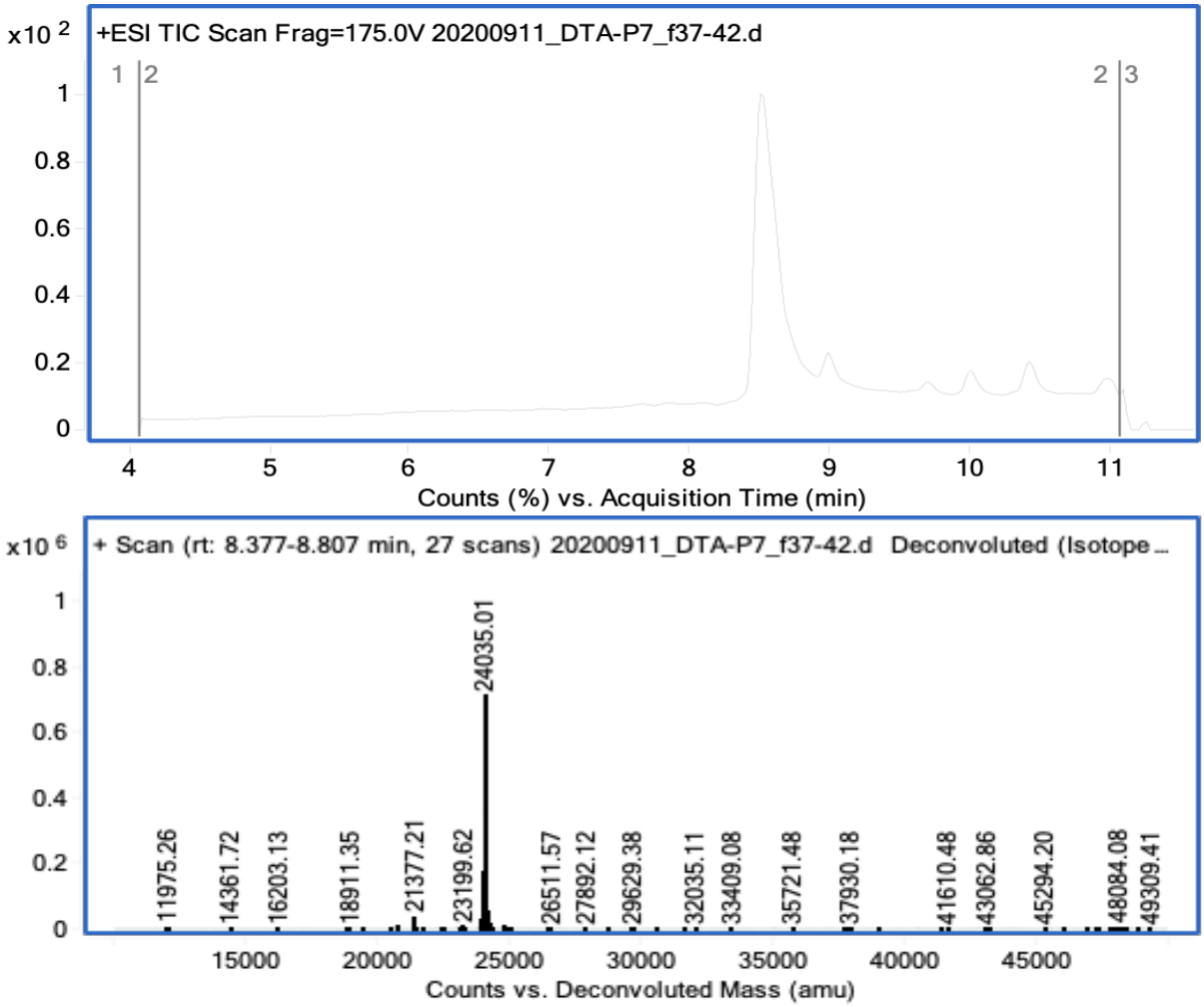




PMO-P19
Mass expected: 8844.1 Da
Mass observed: 8844.3 Da
Peptide sequence: KKKKKQBKKKHAWPKXXC



P6-DTA
Mass expected: 24034.0 Da
Mass observed: 24035.0 Da
Peptide sequence: KKKKKQBKKKHRWPKXXC



3.8 References

(1) Lim, K. R. Q.; Maruyama, R.; Yokota, T. Eteplirsen in the Treatment of Duchenne Muscular Dystrophy. *Drug Des. Devel. Ther.* **2017**, *11*, 533–545. https://doi.org/10.2147/DDDT.S97635.

(2) Juliano, R. L. The Delivery of Therapeutic Oligonucleotides. *Nucleic Acids Res.* **2016**, *44* (14), 6518–6548. https://doi.org/10.1093/nar/gkw236.

(3) Moulton, H. M.; Moulton, J. D. Morpholinos and Their Peptide Conjugates: Therapeutic Promise and Challenge for Duchenne Muscular Dystrophy. *Biochim. Biophys. Acta BBA - Biomembr.* **2010**, *1798* (12), 2296–2303. https://doi.org/10.1016/j.bbamem.2010.02.012.

(4) Malam, Y.; Loizidou, M.; Seifalian, A. M. Liposomes and Nanoparticles: Nanosized Vehicles for Drug Delivery in Cancer. *Trends Pharmacol. Sci.* **2009**, *30* (11), 592–599. https://doi.org/10.1016/j.tips.2009.08.004.

(5) Angelov, B.; Garamus, V. M.; Drechsler, M.; Angelova, A. Structural Analysis of Nanoparticulate Carriers for Encapsulation of Macromolecular Drugs. *J. Mol. Liq.* **2017**, *235*, 83–89. https://doi.org/10.1016/j.molliq.2016.11.064.

(6) Feger, G.; Angelov, B.; Angelova, A. Prediction of Amphiphilic Cell-Penetrating Peptide Building Blocks from Protein-Derived Amino Acid Sequences for Engineering of Drug Delivery Nanoassemblies. *J. Phys. Chem. B* **2020**, *124* (20), 4069–4078. https://doi.org/10.1021/acs.jpcb.0c01618.

(7) Fu, A.; Tang, R.; Hardie, J.; Farkas, M. E.; Rotello, V. M. Promises and Pitfalls of Intracellular Delivery of Proteins. *Bioconjug. Chem.* **2014**, *25* (9), 1602–1608. https://doi.org/10.1021/bc500320j.

(8) Inc, S. T. Sarepta Therapeutics Announces Positive Clinical Results from MOMENTUM, a Phase 2 Clinical Trial of SRP-5051 in Patients with Duchenne Muscular Dystrophy Amenable to Skipping Exon 51 http://www.globenewswire.com/news-release/2020/12/07/2140613/0/en/Sarepta-Therapeutics-Announces-Positive-Clinical-Results-from-MOMENTUM-a-Phase-2-Clinical-Trial-of-SRP-5051-in-Patients-with-Duchenne-Muscular-Dystrophy-Amenable-to-Skipping-Exon-5.html (accessed 2020 -12 -07).

(9) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS Cent. Sci.* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.

(10) Fadzen, C. M.; Holden, R. L.; Wolfe, J. M.; Choo, Z.-N.; Schissel, C. K.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. *Biochemistry* **2019**. https://doi.org/10.1021/acs.biochem.9b00413.

(11) Bechara, C.; Sagan, S. Cell-Penetrating Peptides: 20years Later, Where Do We Stand? *FEBS Lett.* **2013**, *587* (12), 1693–1702. https://doi.org/10.1016/j.febslet.2013.04.031.

(12) Schissel, C. K.; Mohapatra, S.; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep Learning to Design Nuclear-Targeting Abiotic Miniproteins. *Nat. Chem.* **2021**, 1–9. https://doi.org/10.1038/s41557-021-00766-3.

(13) McClorey, G.; Banerjee, S. Cell-Penetrating Peptides to Enhance Delivery of Oligonucleotide-Based Therapeutics. *Biomedicines* **2018**, *6* (2), 51. https://doi.org/10.3390/biomedicines6020051.
(14) Brock, R. The Uptake of Arginine-Rich Cell-Penetrating Peptides: Putting the Puzzle Together. *Bioconjug. Chem.* **2014**, *25* (5), 863–868. https://doi.org/10.1021/bc500017t.
(15) Åmand, H. L.; Rydberg, H. A.; Fornander, L. H.; Lincoln, P.; Nordén, B.; Esbjörner, E. K. Cell Surface Binding and Uptake of Arginine- and Lysine-Rich Penetratin Peptides in Absence and Presence of Proteoglycans. *Biochim. Biophys. Acta BBA - Biomembr.* **2012**, *1818* (11), 2669–2678. https://doi.org/10.1016/j.bbamem.2012.06.006.
(16) Chuard, N.; Fujisawa, K.; Morelli, P.; Saarbach, J.; Winssinger, N.; Metrangolo, P.; Resnati, G.; Sakai, N.; Matile, S. Activation of Cell-Penetrating Peptides with Ionpair−π Interactions and Fluorophiles. *J. Am. Chem. Soc.* **2016**, *138* (35), 11264–11271. https://doi.org/10.1021/jacs.6b06253.
(17) Erazo-Oliveras, A.; Muthukrishnan, N.; Baker, R.; Wang, T.-Y.; Pellois, J.-P. Improving the Endosomal Escape of Cell-Penetrating Peptides and Their Cargos: Strategies and Challenges. *Pharmaceuticals* **2012**, *5* (11), 1177–1209. https://doi.org/10.3390/ph5111177.
(18) Jearawiriyapaisarn, N.; Moulton, H. M.; Buckley, B.; Roberts, J.; Sazani, P.; Fucharoen, S.; Iversen, P. L.; Kole, R. Sustained Dystrophin Expression Induced by Peptide-Conjugated Morpholino Oligomers in the Muscles of Mdx Mice. *Mol. Ther.* **2008**, *16* (9), 1624–1629. https://doi.org/10.1038/mt.2008.120.
(19) Li, Q.; Xu, M.; Cui, Y.; Huang, C.; Sun, M. Arginine-rich Membrane-permeable Peptides Are Seriously Toxic. *Pharmacol. Res. Perspect.* **2017**, *5* (5). https://doi.org/10.1002/prp2.334.
(20) Amantana, A.; Moulton, H. M.; Cate, M. L.; Reddy, M. T.; Whitehead, T.; Hassinger, J. N.; Youngblood, D. S.; Iversen, P. L. Pharmacokinetics, Biodistribution, Stability and Toxicity of a Cell-Penetrating Peptide−Morpholino Oligomer Conjugate. *Bioconjug. Chem.* **2007**, *18* (4), 1325–1331. https://doi.org/10.1021/bc070060v.
(21) Grogg, M.; Hilvert, D.; Ebert, M.-O.; Beck, A. K.; Seebach, D.; Kurth, F.; Dittrich, P. S.; Sparr, C.; Wittlin, S.; Rottmann, M.; Mäser, P. Cell Penetration, Herbicidal Activity, and *in-Vivo* -Toxicity of Oligo-Arginine Derivatives and of Novel Guanidinium-Rich Compounds Derived from the Biopolymer Cyanophycin. *Helv. Chim. Acta* **2018**, *101* (10), e1800112. https://doi.org/10.1002/hlca.201800112.
(22) Aguilera, T. A.; Timmers, M. M.; Olson, E. S.; Jiang, T.; Tsien, R. Y. Systemic in Vivo Distribution of Activatable Cell Penetrating Peptides Is Superior to Cell Penetrating Peptides. *Integr. Biol. Quant. Biosci. Nano Macro* **2009**, *1* (5–6), 371–381. https://doi.org/10.1039/b904878b.
(23) Albright, C. F.; Graciani, N.; Han, W.; Yue, E.; Stein, R.; Lai, Z.; Diamond, M.; Dowling, R.; Grimminger, L.; Zhang, S.-Y.; Behrens, D.; Musselman, A.; Bruckner, R.; Zhang, M.; Jiang, X.; Hu, D.; Higley, A.; Dimeo, S.; Rafalski, M.; Mandlekar, S.; Car, B.; Yeleswaram, S.; Stern, A.; Copeland, R. A.; Combs, A.; Seitz, S. P.; Trainor, G. L.; Taub, R.; Huang, P.; Oliff, A. Matrix Metalloproteinase-Activated Doxorubicin Prodrugs Inhibit HT1080 Xenograft Growth Better than Doxorubicin with Less Toxicity. *Mol. Cancer Ther.* **2005**, *4* (5), 751–760. https://doi.org/10.1158/1535-7163.MCT-05-0006.

(24) Spänig, S.; Heider, D. Encodings and Models for Antimicrobial Peptide Classification for Multi-Resistant Pathogens. *BioData Min.* **2019**, *12*, 7. https://doi.org/10.1186/s13040-019-0196-x.
(25) Witten, J.; Witten, Z. Deep Learning Regression Model for Antimicrobial Peptide Design. *bioRxiv* **2019**, 692681. https://doi.org/10.1101/692681.
(26) Su, R.; Hu, J.; Zou, Q.; Manavalan, B.; Wei, L. Empirical Comparison and Analysis of Web-Based Cell-Penetrating Peptide Prediction Tools. *Brief. Bioinform.* **2020**, *21* (2), 408–420. https://doi.org/10.1093/bib/bby124.
(27) Sanders, W. S.; Johnston, C. I.; Bridges, S. M.; Burgess, S. C.; Willeford, K. O. Prediction of Cell Penetrating Peptides by Support Vector Machines. *PLOS Comput. Biol.* **2011**, *7* (7), e1002101. https://doi.org/10.1371/journal.pcbi.1002101.
(28) Manavalan, B.; Subramaniyam, S.; Shin, T. H.; Kim, M. O.; Lee, G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *J. Proteome Res.* **2018**, *17* (8), 2715–2726. https://doi.org/10.1021/acs.jproteome.8b00148.
(29) Ellert-Miklaszewska, A.; Poleszak, K.; Kaminska, B. Short Peptides Interfering with Signaling Pathways as New Therapeutic Tools for Cancer Treatment. *Future Med. Chem.* **2017**, *9* (2), 199–221. https://doi.org/10.4155/fmc-2016-0189.
(30) Wolfe, J. M. Peptide Conjugation to Enhance Oligonucleotide Delivery. *PhD Thesis Mass. Inst. Technol.* **2018**.
(31) Cardozo, A. K.; Buchillier, V.; Mathieu, M.; Chen, J.; Ortis, F.; Ladrière, L.; Allaman-Pillet, N.; Poirot, O.; Kellenberger, S.; Beckmann, J. S.; Eizirik, D. L.; Bonny, C.; Maurer, F. Cell-Permeable Peptides Induce Dose- and Length-Dependent Cytotoxic Effects. *Biochim. Biophys. Acta BBA - Biomembr.* **2007**, *1768* (9), 2222–2234. https://doi.org/10.1016/j.bbamem.2007.06.003.
(32) Sajjad, M.; Khan, S.; Muhammad, K.; Wu, W.; Ullah, A.; Baik, S. W. Multi-Grade Brain Tumor Classification Using Deep CNN with Extensive Data Augmentation. *J. Comput. Sci.* **2019**, *30*, 174–182. https://doi.org/10.1016/j.jocs.2018.12.003.
(33) Shorten, C.; Khoshgoftaar, T. M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6* (1), 60. https://doi.org/10.1186/s40537-019-0197-0.
(34) Sazani, P.; Gemignani, F.; Kang, S.-H.; Maier, M. A.; Manoharan, M.; Persmark, M.; Bortner, D.; Kole, R. Systemically Delivered Antisense Oligomers Upregulate Gene Expression in Mouse Tissues. *Nat. Biotechnol.* **2002**, *20* (12), 1228–1233.
(35) Wolfe, J. M.; Fadzen, C. M.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Perfluoroaryl Bicyclic Cell-Penetrating Peptides for Delivery of Antisense Oligonucleotides. *Angew. Chem. Int. Ed.* **2018**, *57* (17), 4756–4759. https://doi.org/10.1002/anie.201801167.
(36) Mahlapuu, M.; Håkansson, J.; Ringstad, L.; Björn, C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. *Front. Cell. Infect. Microbiol.* **2016**, *6*. https://doi.org/10.3389/fcimb.2016.00194.
(37) Dutta, D.; Donaldson, J. G. Search for Inhibitors of Endocytosis: Intended Specificity and Unintended Consequences. *Cell. Logist.* **2012**, *2* (4), 203–208. https://doi.org/10.4161/cl.23967.
(38) Choe, S.; Bennett, M. J.; Fujii, G.; Curmi, P. M. G.; Kantardjieff, K. A.; Collier, R. J.; Eisenberg, D. The Crystal Structure of Diphtheria Toxin. *Nature* **1992**, *357* (6375), 216–222. https://doi.org/10.1038/357216a0.

(39) Schneider, A. F. L.; Kithil, M.; Cardoso, M. C.; Lehmann, M.; Hackenberger, C. P. R. Cellular Uptake of Large Biomolecules Enabled by Cell-Surface-Reactive Cell-Penetrating Peptide Additives. *Nat. Chem.* **2021**, 1–10. https://doi.org/10.1038/s41557-021-00661-x.
(40) Gait, M. J.; Arzumanov, A. A.; McClorey, G.; Godfrey, C.; Betts, C.; Hammond, S.; Wood, M. J. A. Cell-Penetrating Peptide Conjugates of Steric Blocking Oligonucleotides as Therapeutics for Neuromuscular Diseases from a Historical Perspective to Current Prospects of Treatment. *Nucleic Acid Ther.* **2019**, *29* (1), 1–12. https://doi.org/10.1089/nat.2018.0747.
(41) Laurent, Q.; Martinent, R.; Lim, B.; Pham, A.-T.; Kato, T.; López-Andarias, J.; Sakai, N.; Matile, S. Thiol-Mediated Uptake. *JACS Au* **2021**, *1* (6), 710–728. https://doi.org/10.1021/jacsau.1c00128.
(42) Kalafatovic, D.; Giralt, E. Cell-Penetrating Peptides: Design Strategies beyond Primary Structure and Amphipathicity. *Mol. Basel Switz.* **2017**, *22* (11). https://doi.org/10.3390/molecules22111929.
(43) Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Liu, S.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by Automated Flow Chemistry. *Science* **2020**, *368* (6494), 980–987. https://doi.org/10.1126/science.abb2491.
(44) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014**, *15* (5), 713–720. https://doi.org/10.1002/cbic.201300796.

4 Cell-Penetrating D-Peptides Retain Antisense Morpholino Delivery Activity

The work presented in this chapter has been reproduced from the following publication with permission from the American Chemical Society:

Schissel, C.K.; Farquhar, C. E.; Malmberg, A.B.; Loas, A.; Pentelute, B.L. Cell-Penetrating D-Peptides Retain Antisense Morpholino Oligomer Delivery Activity. *ACS Bio & Med Chem Au,* **2022**

4.1 Introduction

Cell-penetrating peptides (CPPs) can help treat disease by enhancing the delivery of cell-impermeable cargo. CPPs are a class of peptides that are capable of directly entering the cell cytosol.[1–3] These sequences can deliver covalently bound cargo, offering therapeutic potential to macromolecules otherwise restricted to extracellular targets. Although CPPs have been widely studied since their discovery, the field lacks robust methodology to quantify cell entry and penetration efficacy. This dearth of knowledge is due to the complicated mechanisms of CPP cell entry and the many variables that affect CPP efficacy in any given assay—such as peptide concentration, cell type, temperature, treatment time, and cargo.[4] For example, for the well-studied CPP penetratin (RQIKIWFQNRRMKWKK), the reported ratio between intracellular and extracellular concentration ranges from 0.6:1.0 to 95.0:1.0.[5,6] In addition, it is challenging to determine subcellular localization once a peptide is internalized, despite advances in fluorescence, immunoblot, and mass spectrometry detection.[7] The choice of CPP cargo adds an additional confounding factor, as studies in our laboratory have demonstrated that the cell-penetrating ability of more than ten common CPPs differs when bound to a cyanine dye versus a macromolecular drug, with no discernable trend.[8] Therefore, effective development of CPPs requires new methodology for understanding CPP cell entry and subcellular localization that can be carried out on the CPP-cargo conjugate.

A therapeutic macromolecule that would benefit from enhanced delivery is phosphorodiamidate morpholino oligomer (PMO), which has recently reached the market as an antisense "exon skipping" therapy for Duchenne muscular dystrophy (DMD). One of the drugs, eteplirsen, is a 10 kDa synthetic antisense oligomer that must reach the nucleus and bind pre-mRNA for its therapeutic effect. However, studies have shown that two-thirds of eteplirsen is cleared renally within 24 h of administration.[9,10] Several CPPs have been shown to increase PMO uptake, and recent clinical trial results have shown that once-monthly dosing of SRP-5051 resulted in higher muscle concentration, increased exon-skipping and dystrophin production at 12 weeks as compared to once weekly dosing of eteplirsen after 24 weeks in a different study.[11] Although a wide variety of CPPs have been tested for PMO delivery, they have been limited to the native L-form and studied predominantly with an activity-based assay, forgoing quantitative information on the amount of material inside the cell.[12]

Because PMO is proteolytically stable and invisible to the immune system[10], attachment to an L-peptide may disrupt the desirable characteristics of PMO; however, conjugates using mirror image peptides may retain these characteristics.[13–15] D-peptides have been explored as CPPs and have been found to display at times greater activities to their L-counterparts, despite suggested chiral binding interactions between CPPs and the cell membrane.[16,17] A study on TAT endosomalytic peptide analogs found that the full D-form decreased uptake, but enhanced endosomal escape and proteolytic stability compared to the native form.[15,18] Another reported that a cyclic D-peptide, when co-administered with insulin, enhanced its oral bioavailability and therapeutic effect in the gut.[19] Some reports are contentious as to whether mutations to D-amino acids are detrimental to CPP activity, and certainly CPPs that depend on a chiral interaction or highly ordered structure to enter the membrane would lose efficacy from D-amino acid substitutions.[20,21] Previous studies have suggested that efficient CPPs for PMO delivery lack secondary structure and can enter the cell through clathrin-mediated endocytosis.[22,23] While it has been found that PMO-D-CPPs have enhanced proteolytic stability over their L-counterparts[24], their PMO delivery activity has not been investigated. In addition to their underexplored potential, we are interested in studying D-CPPs as potential therapeutic carrier moieties because their fully noncanonical sequence is resistant to proteolysis and may go unrecognized by the host immune system.[15,20] Despite this potential, mirror image peptides have not yet been fully explored for the delivery of PMO.

The proteolytic stability of mirror image peptides would simplify their characterization after uptake into cells, providing orthogonal information to activity-based assays. Currently, the main method used to characterize PMO-CPP internalization is an in vitro assay in which successful delivery of the active oligomer to the nucleus results in green fluorescence.[25] While this is an excellent assay to measure PMO-CPP activity, this assay does not give information on the quantity of material inside the cell. Especially for conjugates with a known endocytic mechanism, understanding endosomal escape is crucial. High concentrations of peptides trapped in the endosome would not be apparent by the activity assay alone, but this loss of active peptide could be of great therapeutic detriment.[26] Therefore, an additional assay that reveals relative quantities of material in different parts of the cell would provide a valuable metric of CPP delivery, in parallel with our current activity assay. This could be achieved by extracting the cytosolic fraction of treated cells using a mild detergent, such as digitonin, and comparison to the whole cell fraction.[27]

Comparing activity and relative quantity would be a valuable metric for efficiency, where a high efficiency peptide is one with a high ratio of antisense activity to internal concentration.

Existing methods to quantify uptake into cells include fluorescence, immunoblot, and mass spectrometry, but it is still challenging to distinguish between endosomal and cytosolic localization. Several assays can differentiate between endosomal and cytosolic localization using indirect quantification via a readout generated by a delivered cargo, including the chloroalkane penetration assay (CAPA),[28] GFP complementation assays,[29] and more recently the NanoClick[30] assay and SLEEQ[31] assay. Direct quantification of the cytosolic concentration of a fluorescently-labeled protein construct is possible using fluorescence correlation spectroscopy.[32]

Mass spectrometry is a direct quantification tool that would give information about the concentration of peptides recovered from biological mixtures with limited labeling required. Past studies have illustrated how MALDI-TOF mass spectrometry is a practical tool for absolute and relative quantification of peptides and proteins. For example, using an internal standard of a similar molecular weight is sufficient for generation of a calibration curve.[33] Quantitation of total uptake of L-CPPs was achieved using heavy atom-labeled internal standards.[34,35] While this assay provided information regarding whole cell uptake of CPPs and CPP-peptide conjugates, it is limited by the need for heavy-atom labeling and the rapid degradation of L-peptides.[36] A method for circumnavigating the need for spike-in of heavy atom-labeled standards was developed for the relative quantification of phosphopeptides.[37] However, the proteolytic stability of D-peptides would facilitate their recovery and analysis as a mixture from inside cells and animals, allowing for the use of a new metric of antisense delivery efficiency.

Here we report that compared to the native L-forms, the mirror image forms of several sequences were equally able to deliver antisense molecules to the nucleus, but their increased proteolytic stability simplified mass spectrometry-based characterization following cytosolic delivery. Cytosolic delivery can be quantified based on the recovery of intact constructs from inside the cell. We profiled the uptake of biotinylated CPPs and PMO-CPPs to determine their relative concentrations in the whole cell and cytosol using careful extraction with digitonin and direct detection via MALDI-TOF. By comparing PMO delivery activity to relative internal concentration, we can derive a new metric for cargo delivery efficiency that may be useful for future development of CPPs for PMO delivery.

4.2 Results

4.2.1 Mirror image peptides have same PMO delivery activity as native forms

We first established that several mirror image peptides could deliver a model PMO molecule to the nucleus of cells with similar efficacy to their native L-forms. We selected commonly studied, cationic peptides with precedented PMO delivery activity, but not known for their dependence on secondary structure or receptor interaction, and synthesized them in their L and D-form. The first iteration of these constructs contained a biotin linked through a 6-aminohexanoic acid residue and a trypsin-cleavable motif in between the peptide and the cargo. The cargo portion contained an azide for conjugation to PMO, and a biotin for use in affinity capture (Fig. 4.1A). The second generation of these constructs did not include an aminohexanoic acid linker, and did include a single tryptophan residue for quantitation by UV-Vis (Fig. 4.1C). The final constructs included cationic (all with the exception of MPG), oligoarginine (R8, BPEP, Bac7, TAT, TATp), amphipathic (Penetratin, Hell11-7, DPV6, DPV7), and hydrophobic (MPG) sequences.

In total, 14 PMO-peptide sequences were synthesized in their D- and L-forms and tested in an activity-based in vitro assay.[22,23] To synthesize the constructs, L-peptides were synthesized via automated fast-flow peptide synthesis, and D-peptides were synthesized using semi-automated fast flow peptide synthesis. Azido-lysine and biotin moieties were added to the N-terminus of the peptides manually, and the peptides were simultaneously cleaved and deprotected before purification via RP-HPLC. PMO was modified with a dibenzocyclooctyne (DBCO) moiety and purified before attachment to the azido-peptides via strain-promoted azide-alkyne cycloaddition in water. Purified constructs were then tested using an activity-based readout in which nuclear delivery results in fluorescence. Briefly, HeLa cells stably transfected with an EGFP gene interrupted by a mutated intron of β-globin (IVS2-654) produce a non-fluorescent EGFP protein. Successful delivery of PMO IVS2-654 to the nucleus results in corrective splicing and EGFP synthesis. The amount of PMO delivered to the nucleus is therefore correlated with EGFP fluorescence, quantified by flow cytometry. Activity is reported as mean fluorescence intensity (MFI) relative to PMO alone.

Dose-response studies with sequences in the D- and L-form confirmed similar activities between mirror image CPPs. From our initial proof-of-concept experiment, involving eight sequences in L- and D-form tested at a single concentration (5 μM), there was not a significant difference between the activities of the mirror image peptides with the exception of Penetratin and

Hel11-7 (Fig. 4.1B, Fig. 4.2). The second generation constructs were tested at varying concentrations in the EGFP 654 assay, and the results further suggested that mirror image peptides shared nearly identical PMO delivery activities (Fig. 4.1D, 4.3-4.5). Interestingly, while PMO-D-R8 showed a similar concentration-activity curve to PMO-L-R8, there was a difference in activities at 5 μM. This trend has been observed before, albeit without attached cargo, in that L-R8 entered cells more efficiently than D-R8.[20] Though striking similarities were observed for the selected peptides, this similarity is certainly not expected for all CPPs and cargoes, including those that rely on secondary structure or receptor-mediated uptake. These activities correspond with how much PMO is delivered to the nucleus, but do not provide information regarding the total amount of material inside the cell. Relative efficiency of a PMO-CPP could be characterized by comparing activity to internal concentration, as discussed later (Fig. 4.6).

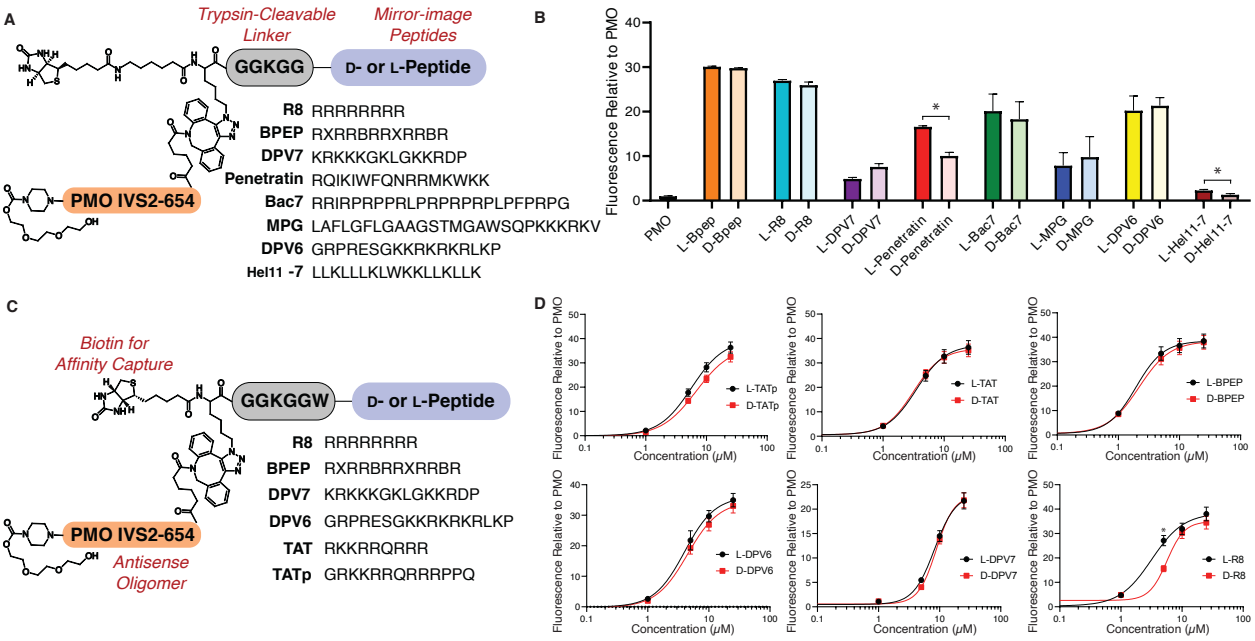


Figure 4.1 Mirror image cell-penetrating peptides have similar PMO delivery activity as their native counterparts. (A) Design of the first generation of conjugates studied, including the eight studied CPP sequences. Macromolecular cargo PMO IVS2-654 is attached to the N-terminus of the peptides, along with a biotin handle for subsequent affinity capture. A trypsin-cleavable linker connects the cargo to the peptides, and a 6-aminohexanoic acid moiety is between the biotin and the peptide. (B) Shown is activity data from the EGFP 654 assay described in Fig. 2 conducted with D- and L-forms of the first generation sequences at 5 μM. (C) Design of the second generation of conjugates used in subsequent experiments. The aminohexanoic acid linker was removed, and each peptide contains a single tryptophan residue for quantitation by UV-Vis. (D) Dose-response curves of D- and L-forms of several sequences at varying concentrations. No membrane toxicity was observed after analyzing the supernatant of these experiments. Bars and data points represent mean ± SD, N = 3 distinct samples from a representative biological replicate. Replicate experiments showed similar results. Statistical significance between the D- and L-samples at each concentration was determined using a two-tailed unpaired t-test and indicated by *p<0.01, with blank indicating not statistically significant.

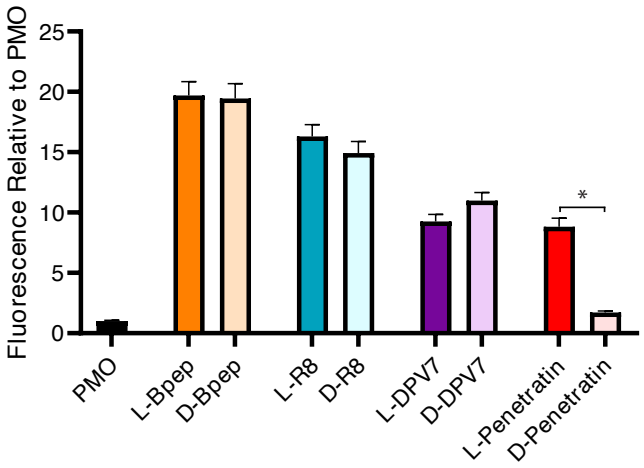


Figure 4.2 Biological replicate 1 of EGFP Assay with several first generation L- and D- PMO-peptides. Performed at 5 µM. Bars represent mean ± SD, N = 3 distinct samples. Statistical significance between the D- and L-samples at each concentration was determined using a two-tailed unpaired t-test and indicated by *p<0.01, all other samples were not significantly different.

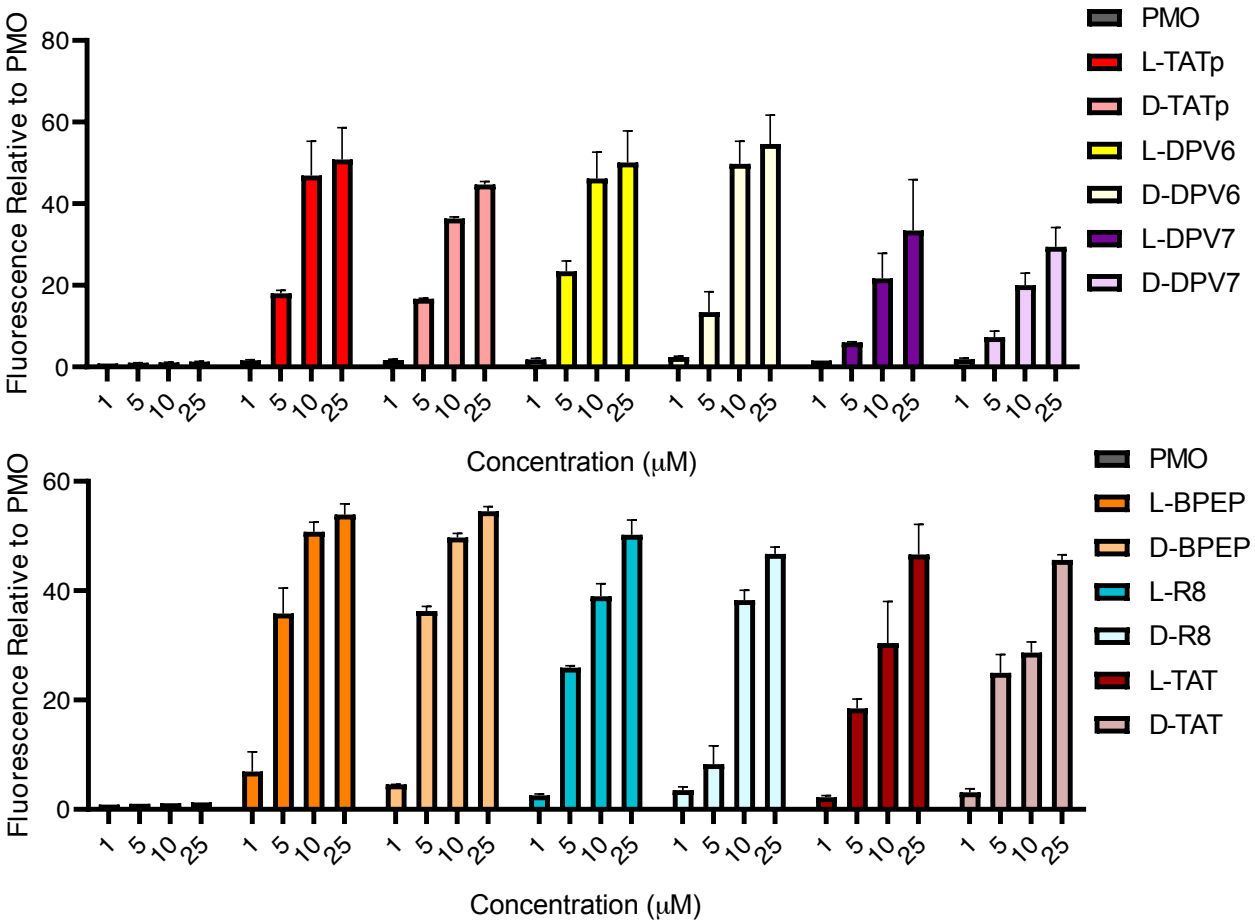


Figure 4.3 Biological replicate 1 of EGFP assay with second generation PMO-D- and L- peptides. Performed at 5 µM. Bars represent mean ± SD, N = 3 distinct samples. Statistical significance between the D- and L-samples at each concentration was determined using a two-tailed unpaired t-test and indicated by *p<0.01, all other samples were not significantly different.

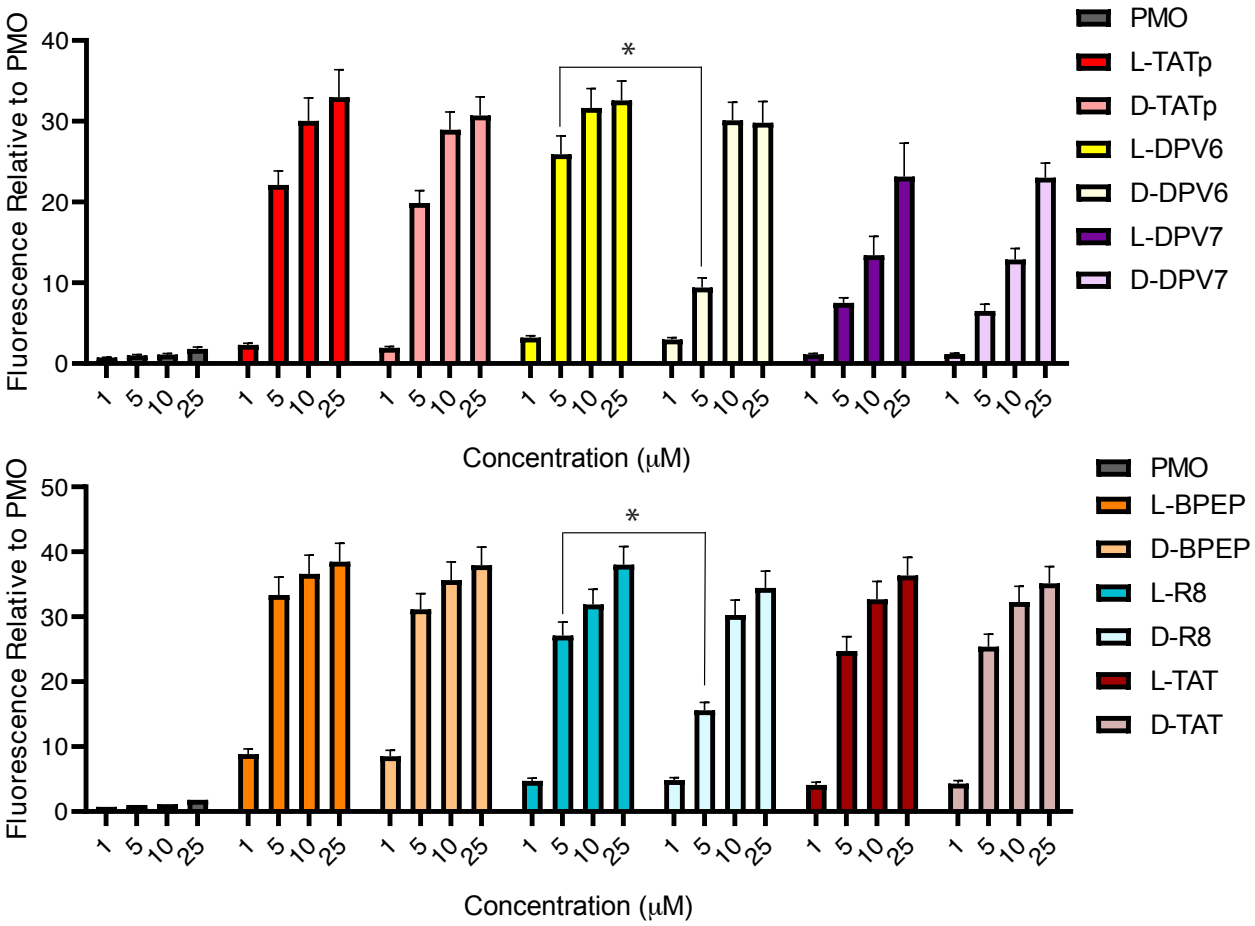


Figure 4.4 Biological replicate 2 of EGFP assay with second generation PMO-D- and L- peptides. Performed at 5 μM. Bars represent mean ± SD, N = 3 distinct samples. Statistical significance between the D- and L-samples at each concentration was determined using a two-tailed unpaired t-test and indicated by *p<0.01, all other concentrations were not significantly different.

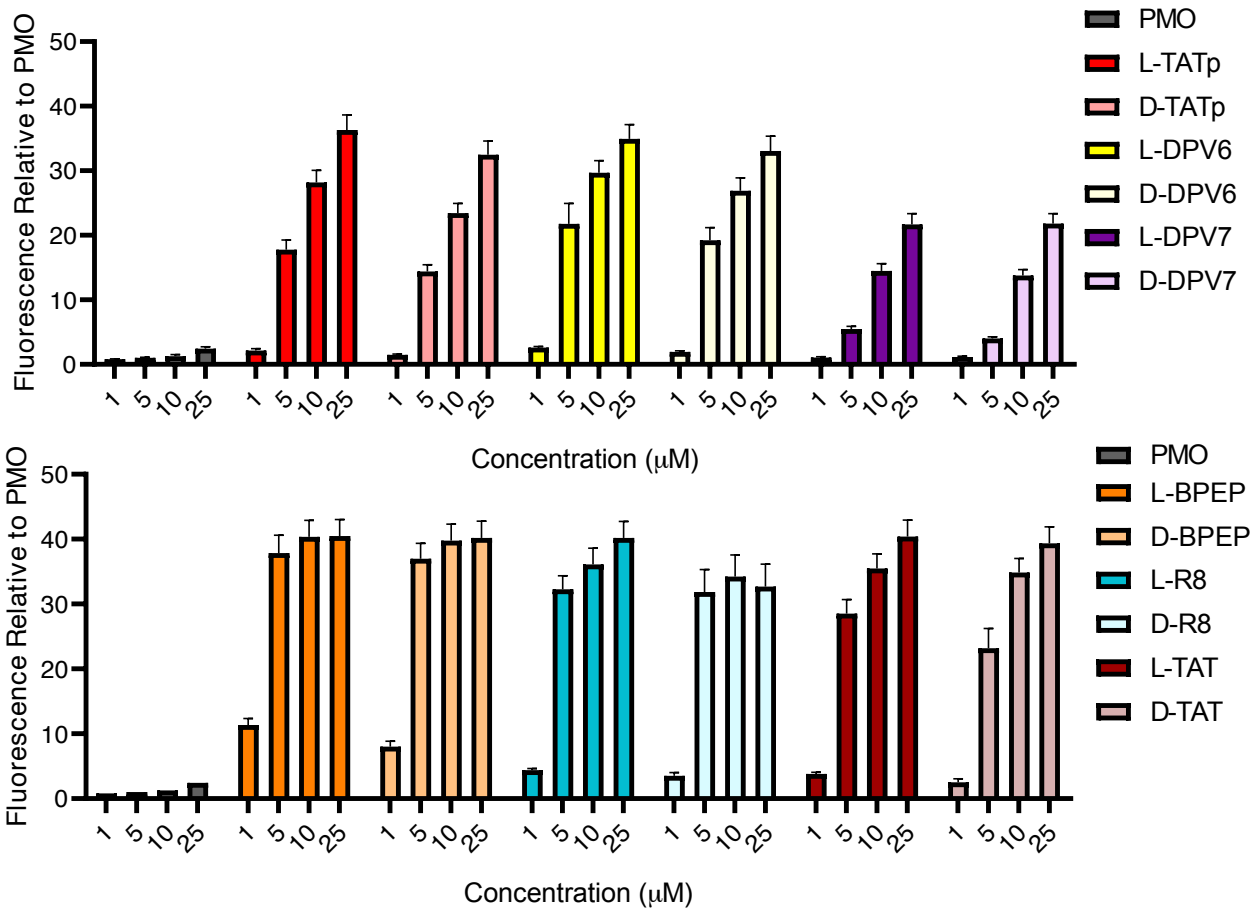


Figure 4.5 Biological replicate 3 of EGFP assay with second generation PMO-D- and L- peptides. Performed at 5 μM. Bars represent mean ± SD, N = 3 distinct samples. Statistical significance between the D- and L-samples at each concentration was determined using a two-tailed unpaired t-test and indicated by *p<0.01, all other concentrations were not significantly different.

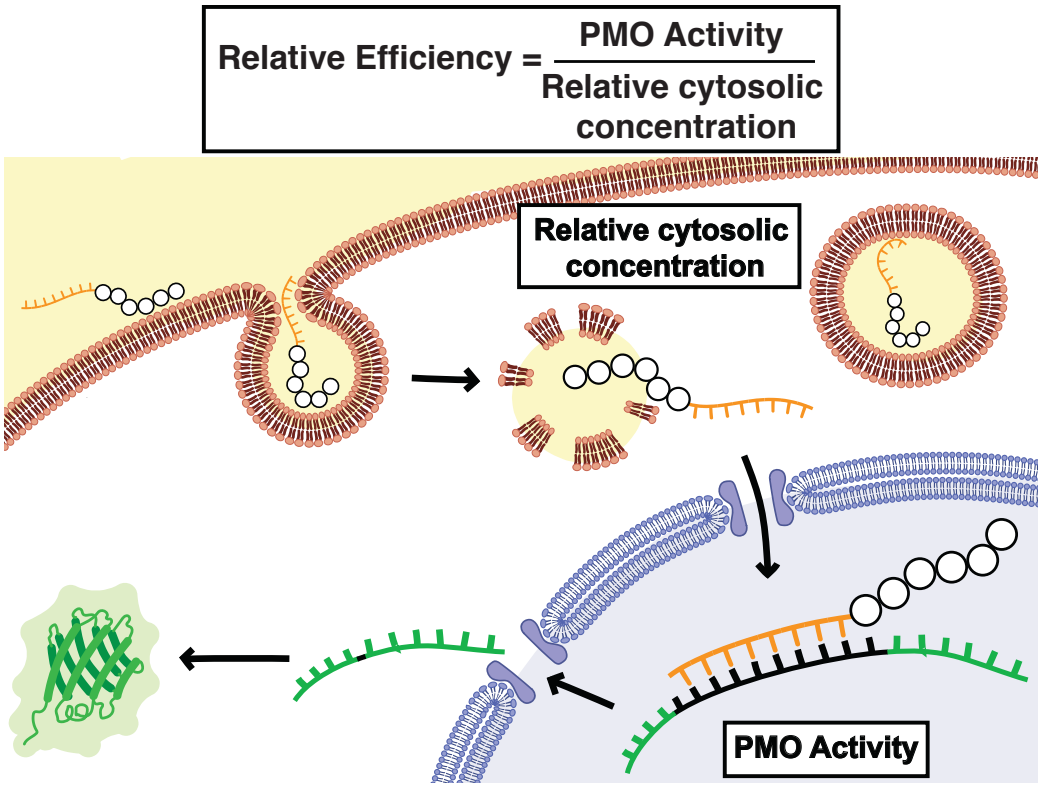


Figure 4.6 Schematic of the proposed cellular internalization pathway of PMO-CPPs and subsequent exon-skipping activity. While activity assays give information of how much active PMO reached its target, a mass spectrometry-based assay may give information on the amount of PMO-CPP located in the cell and cytosol. Comparing these metrics provides a new estimate of CPP efficiency.

Furthermore, the supernatant from these assays was tested for lactate dehydrogenase (LDH) release, indicative of membrane toxicity. PMO-D-penetratin and PMO-D/L-Hel11-7 demonstrated some toxicity and were discontinued from the study (Fig. 4.7). The second-generation constructs did not elicit membrane toxicity at the doses tested (Figs. 4.8-4.9).

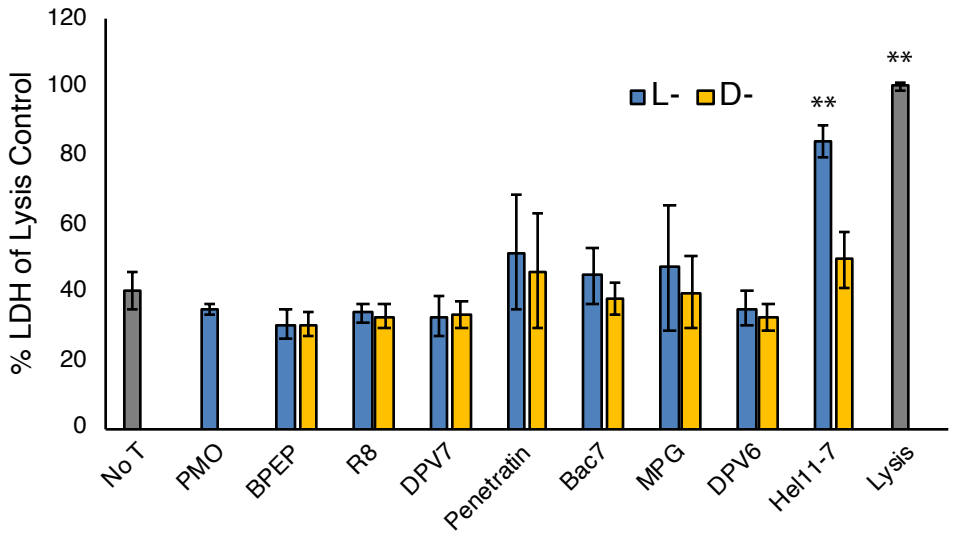


Figure 4.7 LDH Assay of first-generation peptide conjugates. LDH assay conducted on supernatant of cells treated for the EGFP assay shown in Fig 1. Concentration of PMO conjugates was 5 µM. Bars represent mean ± SD, N = 3. Statistical significance of a sample compared to the no treatment condition was determined using a student's paired, two-tailed t-test and indicated as *p<0.01, **p<0.001, all other samples were not significantly different.

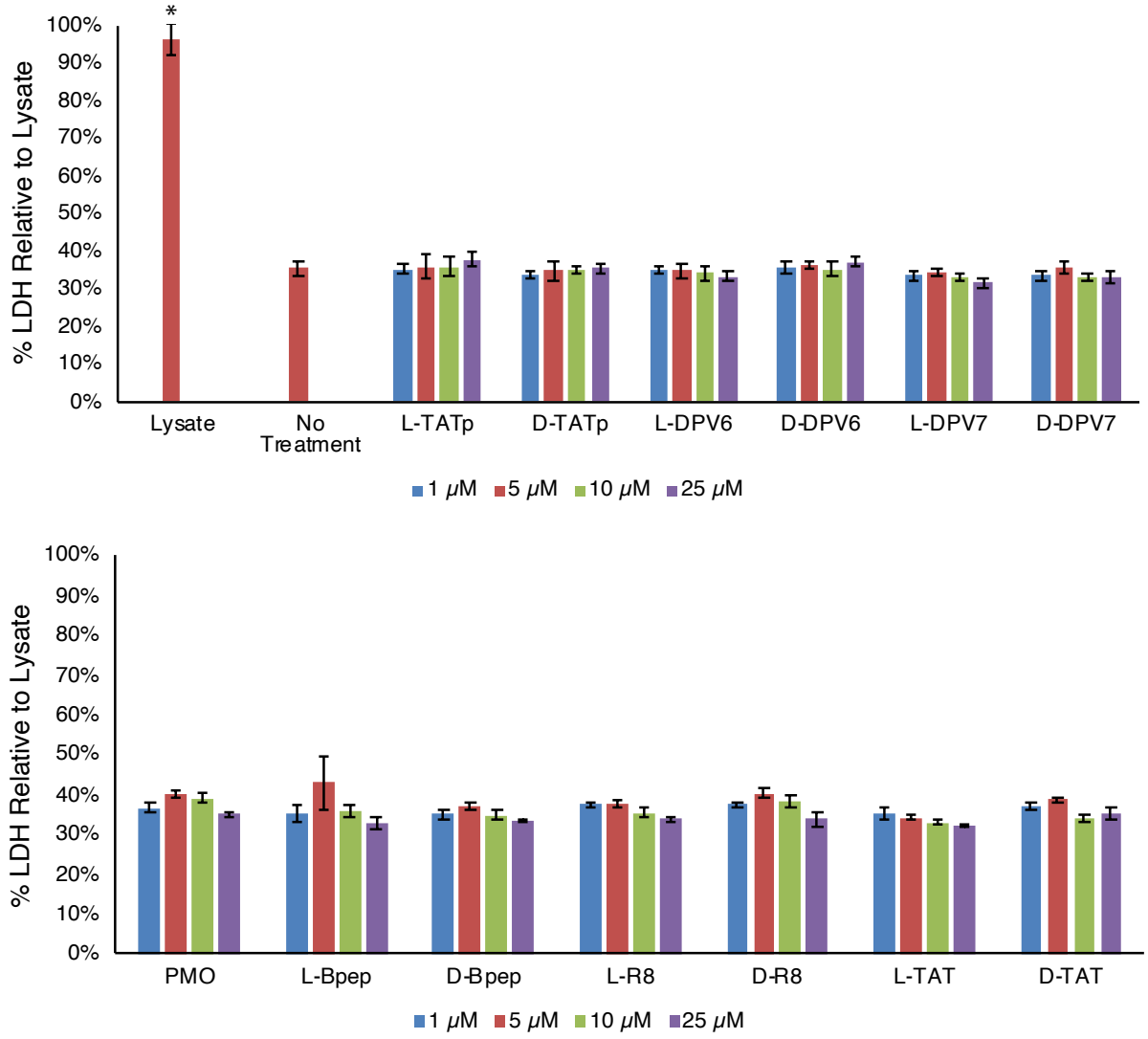


Figure 4.8 LDH Assay of second-generation peptide conjugates conducted on supernatant of cells treated for the EGFP assay shown in Fig 4.3. Bars represent mean ± SD, N = 3. Statistical significance of a sample compared to the no treatment condition was determined using a student's paired, two-tailed t-test and indicated as *p<0.01, **p<0.001, all other samples were not significantly different.

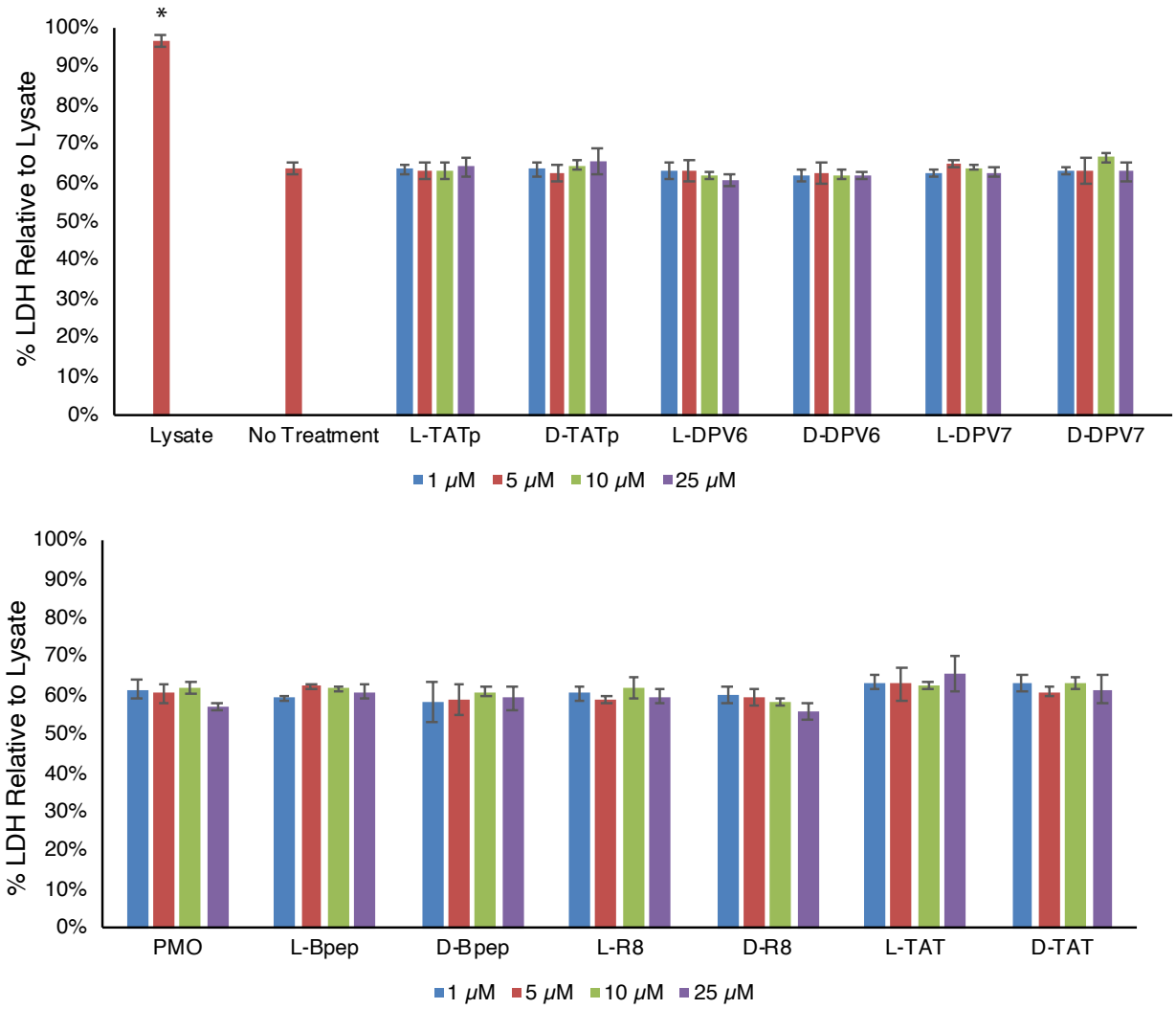


Figure 4.9 Biological replicate LDH Assay of second-generation peptide conjugates conducted on supernatant of cells treated for the EGFP assay shown in Fig 4.4. Bars represent mean ± SD, N = 3. Statistical significance of a sample compared to the no treatment condition was determined using a student's paired, two-tailed t-test and indicated as *p<0.01, **p<0.001, all other samples were not significantly different.

4.2.2 Mirror image peptides are proteolytically stable

Our primary motivation for investigating mirror image peptides for transporting PMO was that the D-form would be stable against proteolysis and thus would match this property of the PMO cargo. D-peptides are indeed stable against degradation, illustrated by a time-course study in which both forms of PMO-CPPs were incubated in 25% human serum. While the studied PMO-D-CPPs remained intact 24 h later, the L-forms rapidly degraded into multiple fragments, leaving the parent construct as a minor product after only 1 h of incubation (Fig. 4.10-4.13). Looking at TATp as an example, the major degradation products correspond to loss of C-terminal RRRPPQ and KKRRQRRRPPQ motifs (parent sequence: GGKGGWGRKKRRQRRRPPQ). Similar observations were noted for DPV6, DPV7, and BPEP. This observation furthers the notion that L-peptides are not suitable for investigation using mass spectrometry after recovery from a biological setting. However, D-peptide conjugates can be recovered from a biological environment such as serum without suffering degradation, simplifying their characterization via mass-spectrometry.

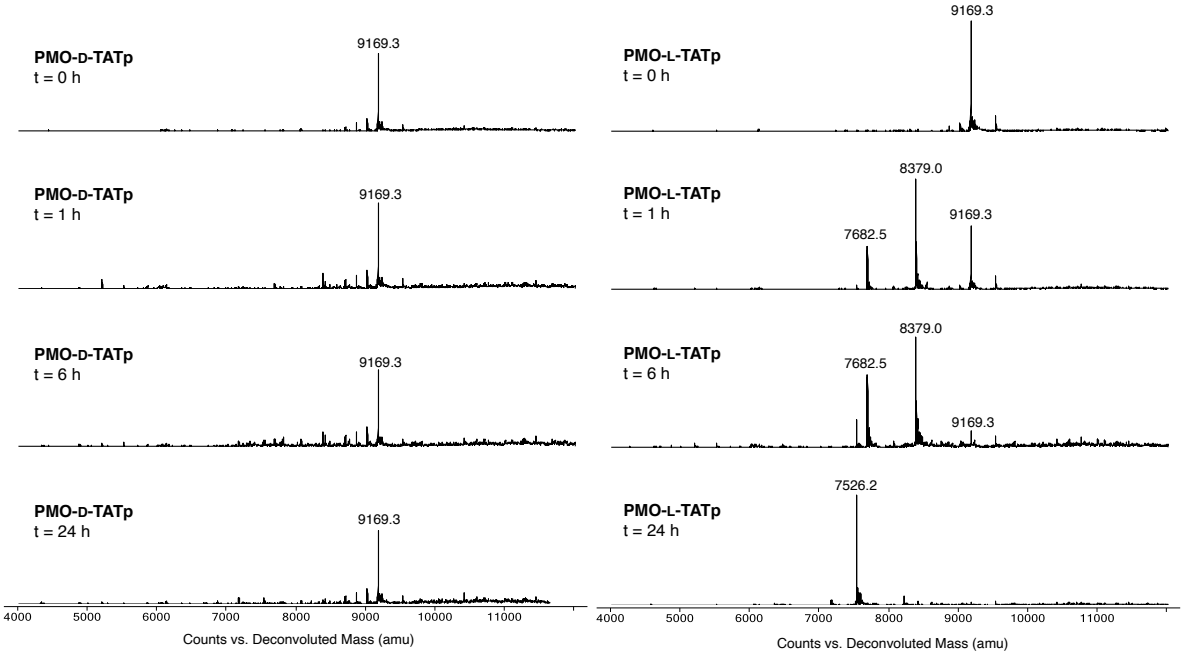


Figure 4.10 Mirror image TATp remains proteolytically stable. Mass spectra of samples containing PMO-D- and L-TATp (50 µM) following incubation with 25% human serum at 37 °C at various time points. The L-variant is degraded within 1 h, while the D-variant is stable after 6 h.

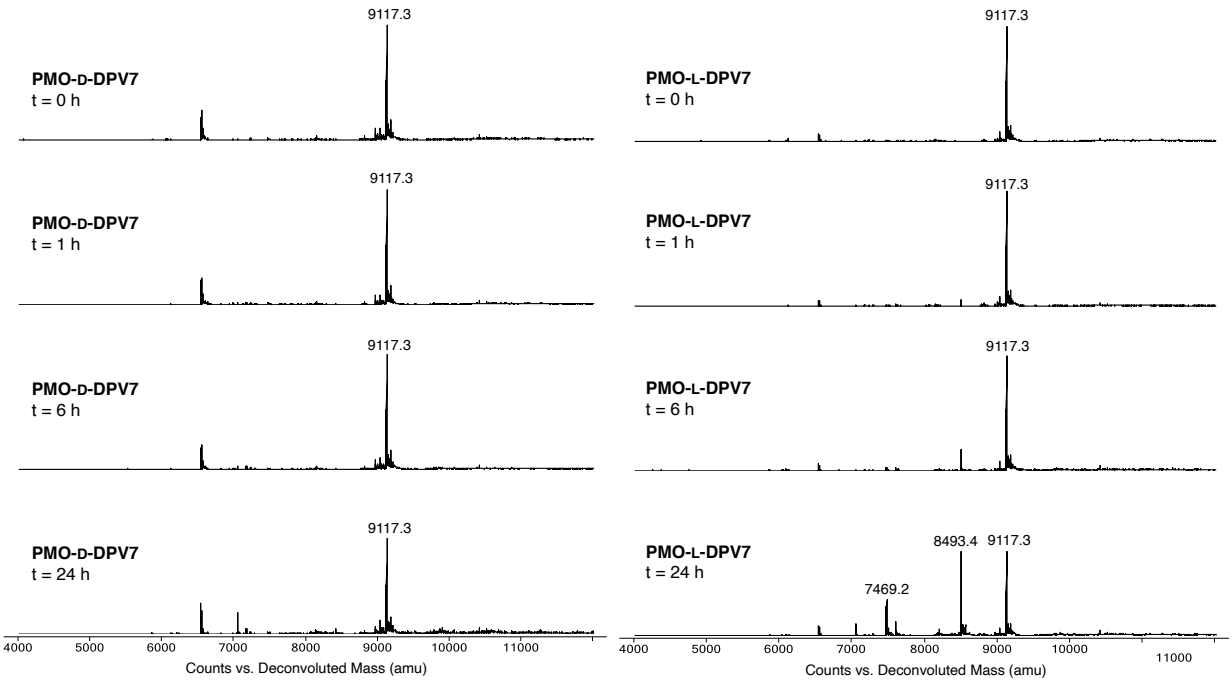


Figure 4.11 Mirror image DPV7 remains proteolytically stable. Mass spectra of samples containing PMO-D- and L-DPV7 (50 µM) following incubation with 25% human serum at 37 °C at various time points. The L-variant is degraded within 1 h, while the D-variant is stable after 6 h.

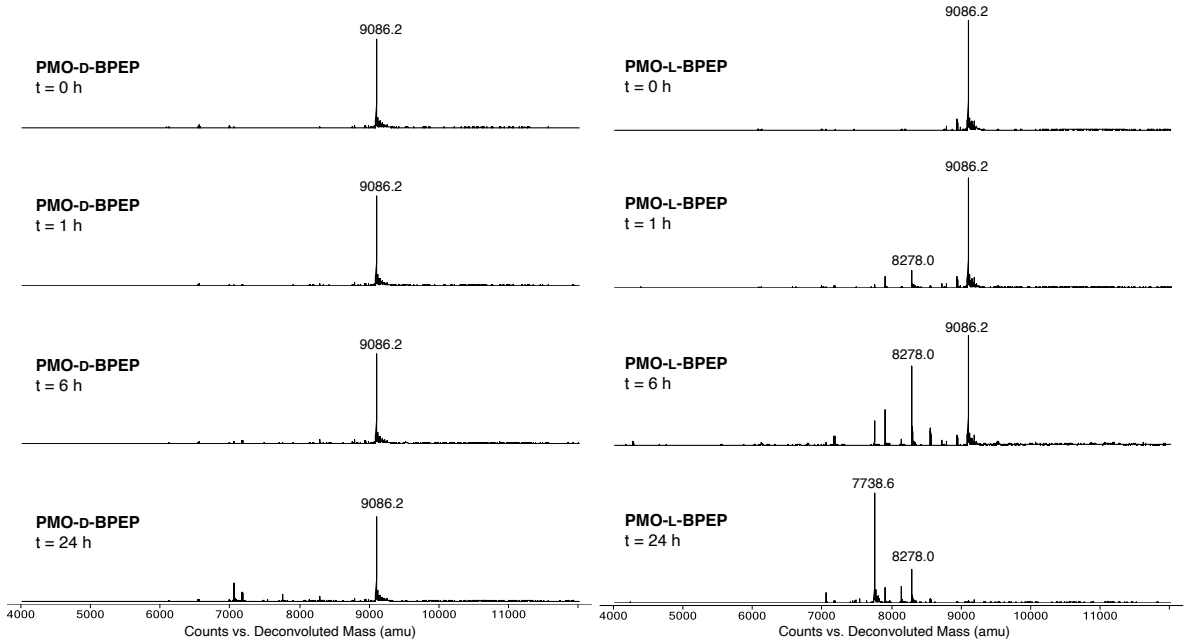


Figure 4.12 Mirror image BPEP remains proteolytically stable. Mass spectra of samples containing PMO-D- and L-BPEP (50 µM) following incubation with 25% human serum at 37 °C at various time points. The L-variant is degraded within 1 h, while the D-variant is stable after 6 h.

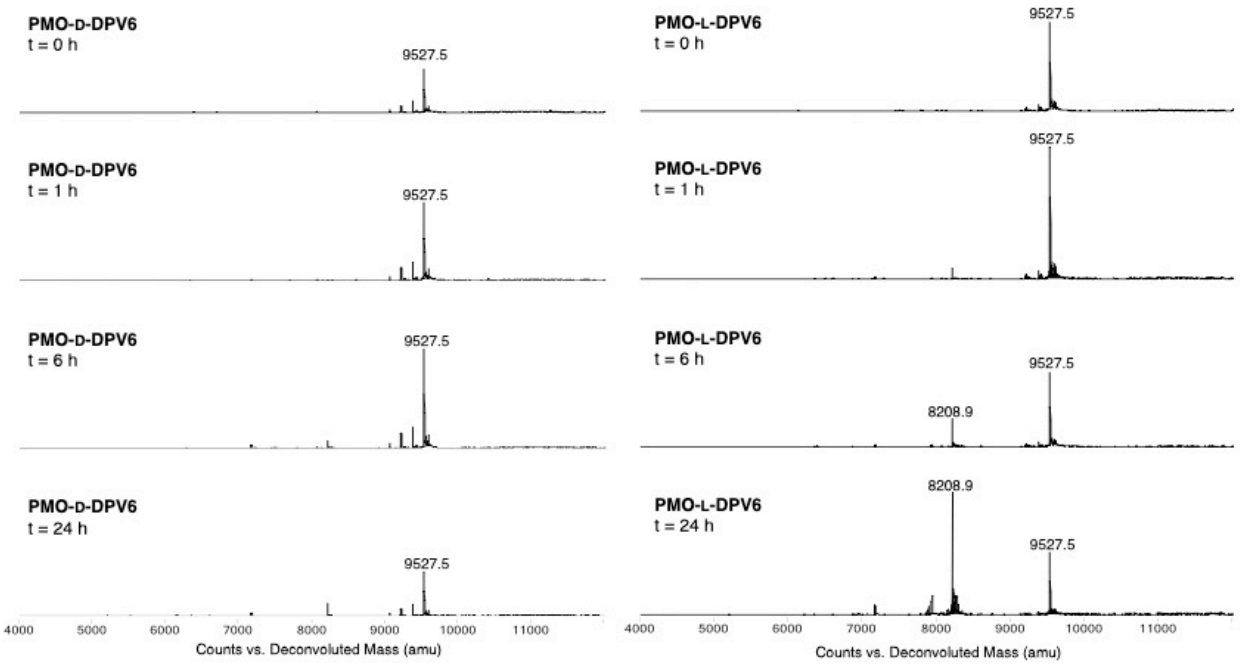


Figure 4.13 Mirror image DPV6 remains proteolytically stable. Mass spectra of samples containing PMO-D- and L-DPV6 (50 μM) following incubation with 25% human serum at 37 °C at various time points. The L-variant is degraded within 1 h, while the D-variant is stable after 6 h.

4.2.3 Mirror image PMO-peptides can be recovered from inside cells

The proteolytic stability of D-peptides simplifies the recovery of mixtures of intact constructs after being internalized into cells. While MALDI-ToF has been used previously to analyze the quantity of L-peptides and protein-peptide conjugates recovered from inside cells, it has not yet been used to profile PMO-D-peptide conjugates, or mixtures of more than three conjugates at a time.[34,35] MALDI-ToF has also been used to quantify many different delivered cargos.[38–40] The use of D-peptides would facilitate the analysis of a mixture of conjugates, because without degradation, only the parent peak would be observed. Moreover, to our knowledge, this platform has not yet been used to study peptides recovered from sub-cellular fractions. Through a series of experiments, we show that mixtures of intact PMO-D-CPPs can be recovered from the cytosol of cells and analyzed by MALDI-ToF to estimate their relative abundances without the need for isotope labeling or standard curves (Fig. 4.14A).

We recapitulated a known empirical trend suggesting more Arg residues lead to greater uptake. We began our assay with a simple model system of four polyarginine peptides with a trypsin cleavable linker and a single biotin label. HeLa cells were incubated with biotin-K-D-Arg_4, D-Arg_6,

D-Arg_8, and D-Arg_{10} for 1 h. The cells were then washed extensively with PBS and heparin and trypsinized to lift the cells as well as to cleave the L-linker on the extracellular constructs, preventing their recovery. The treatment with heparin is known to release peptides bound to the membrane but not yet internalized[20], although it is possible that membrane-bound peptides are recovered in the whole cell lysate. The whole cell lysate was then prepared using RIPA buffer. Fully intact biotinylated peptides were captured with magnetic streptavidin Dynabeads, washed, and plated directly for MALDI analysis. Also plated were Dynabeads incubated in an equimolar mixture of the same constructs as determined by UV-Vis (Fig. 4.14B).

The relative concentration of peptides on the beads can be estimated by determining the analyte's response factor (F) from the equimolar standard (Fig. 4.14B). In the standard, each analyte's concentration is 1 mM, and each analyte's response factor (F) is determined by normalizing their intensities to an internal standard (S). Here, BKr_6 was selected as the arbitrary standard, where F = 1. The response factor of each analyte should remain consistent across samples that contain the same analytes[33,37], and was used to calculate the fold change in concentration in the experimental samples. The relative concentrations [X] of the analytes, normalized to the 'internal standard' BKr_6 are shown as a bar graph (Fig. 4.14C). There is a clear increase in concentration of the constructs with more Arg residues, with BKr_{10} having 40-fold greater concentration than BKr_6. This trend of greater number of Arg residues leading to greater uptake is already well documented in the literature.[41,42]

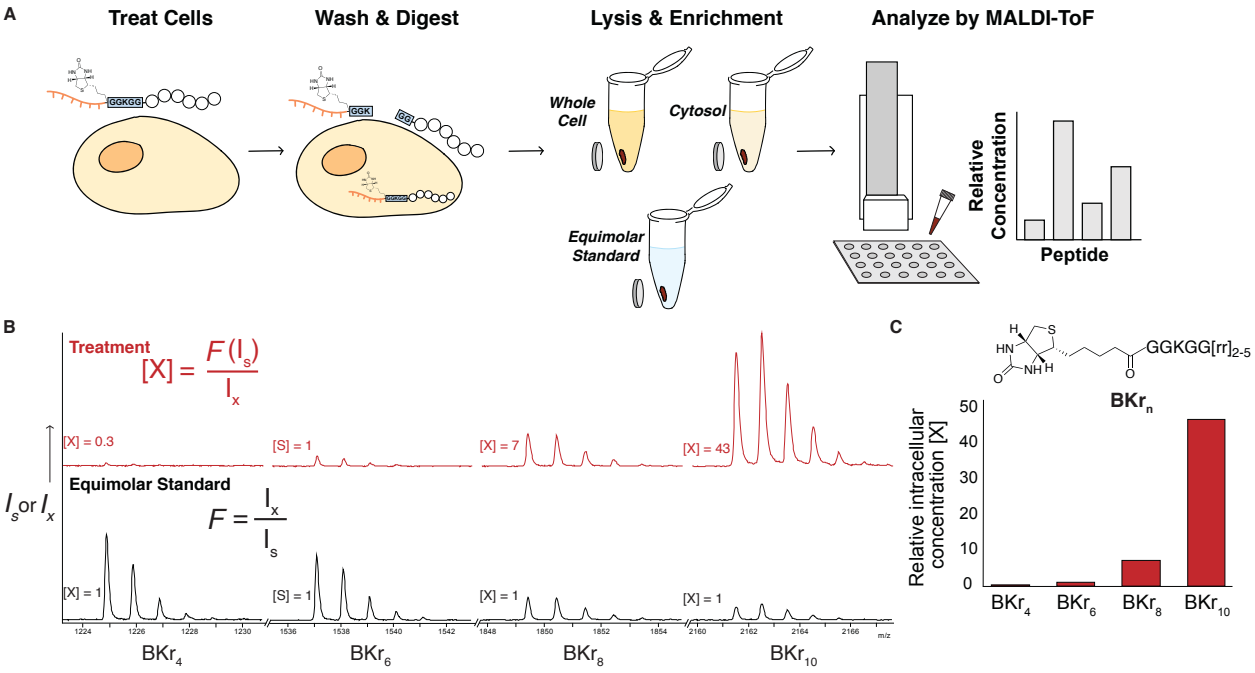


Figure 4.14 Uptake assay reveals relative concentrations of intact construct inside the cell. (A) Workflow of the uptake assay; cells are treated with PMO-D-CPPs, washed, and lysed to extract the whole cell lysate or the cytosol. A trypsin-cleavable L-linker ensures extracellular constructs are not recovered. Constructs are immobilized on magnetic streptavidin beads, washed, and plated directly for MALDI analysis. Intensities of analytes compared to an equimolar standard provide relative concentrations. (B) MALDI-TOF mass spectra displaying ions of intact biotinylated D-polyarginine peptides, isolated after internalization into HeLa cells. Spectra show ions corresponding to intact conjugates in the equimolar spike-in (black) and the whole cell lysate of cells treated with equimolar mixture (red). (C) By comparing relative intensities in the equimolar standard, the response factor (F) was determined and used to calculate the fold change in concentration in the experimental samples, shown as bar graph, normalized to BKr_4. Also shown is the equation used to determine relative concentration: I (intensity), $[X]$ (sample concentration), F (response factor), X (sample), S (standard).

4.2.4 PMO-D-CPPs can be extracted from cytosol and analyzed by MALDI

Next, we found that these PMO-CPP constructs likely enter via energy-dependent endocytosis. When evaluating CPP delivery efficiency, considering the mechanism of uptake and potential for endosomal entrapment is necessary. In contrast with CPPs alone which can enter via passive diffusion,[41,43] we hypothesized that the uptake mechanism of these constructs was likely endocytosis, considering the size of the PMO cargo as well as previous studies that we have conducted on similar conjugates.[22,23] Using a panel of chemical endocytosis inhibitors, we performed a pulse-chase EGFP 654 assay format in which cells were pre-incubated with inhibitors before treatment with the L- and D-forms of PMO-DPV7 and PMO-Bac7. Analysis by flow cytometry revealed that chlorpromazine reduced activity in a dose-dependent manner (Fig. 4.15A). Chlorpromazine is an inhibitor of clathrin-mediated endocytosis and has previously been observed

to inhibit activity of similar PMO constructs. While it is possible that multiple uptake mechanisms are occurring, these PMO-CPP conjugates are likely taken up by active transport.

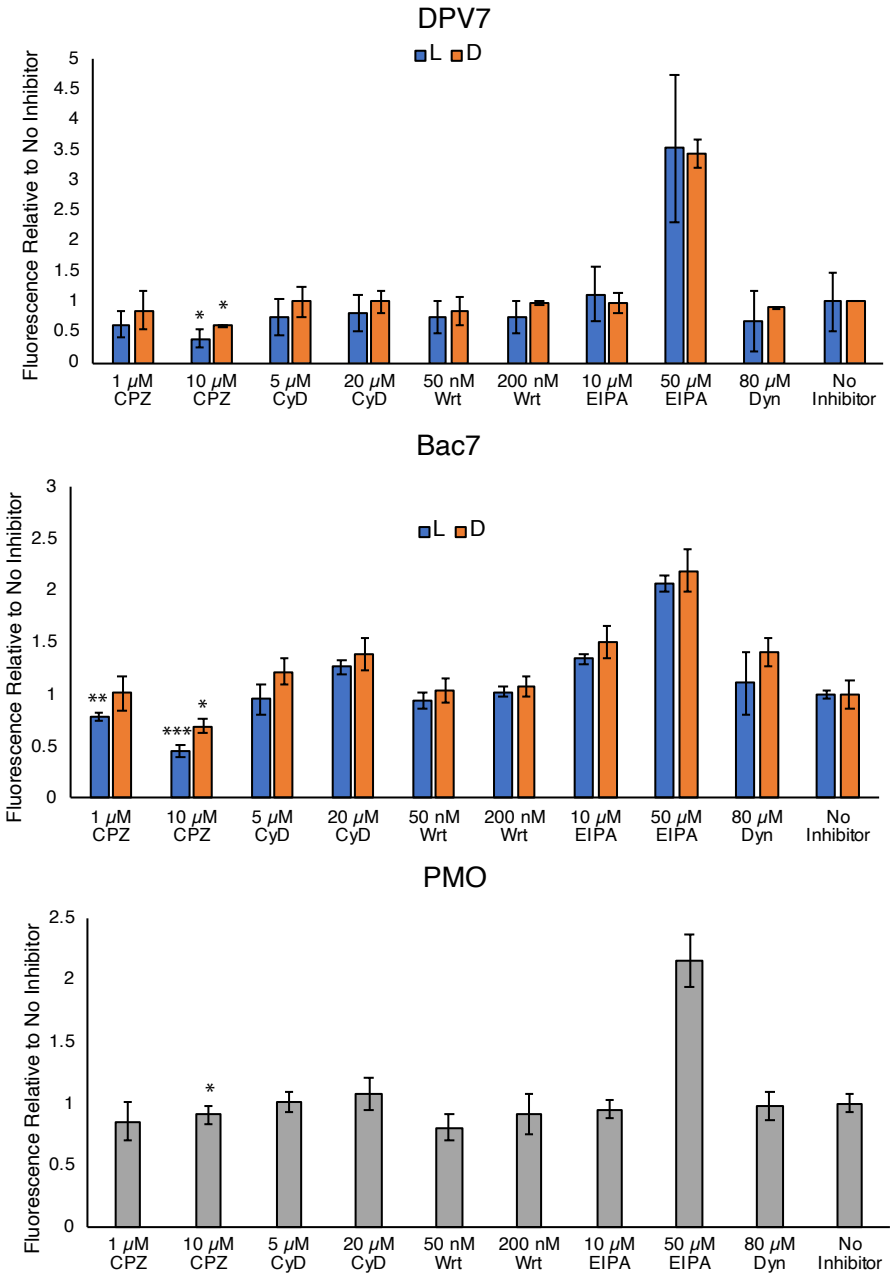


Figure 4.15 Pulse-chase EGFP Assay using several chemical endocytosis inhibitors, analyzing uptake pathways of PMO-Bac7, PMO-DPV7, and PMO alone. Concentration of PMO conjugate at 5 µM. Bars represent mean ± SD, N = 3. Statistical significance of a sample compared to the no inhibitor condition was determined using a student's paired, two-tailed t-test and indicated as *p<0.1, **p<0.01, ***p<0.001.

Knowing that PMO-CPPs enter via endocytosis, we assert that extracting the cytosol and comparing to whole cell lysate is critical when evaluating relative concentrations of constructs internalized into cells. Not all endocytosed compounds are able to escape the endosome, and endosomal entrapment would lead to less active PMO delivered into the cytosol, measured by a lower cytosolic concentration relative to whole-cell lysate. Therefore, we then extracted biotinylated PMO-CPPs from the cytosol as well as the whole cell lysate following internalization and detected them by Western blot and MALDI. Individually, we incubated PMO-D-R8 or PMO-D-Bpep at 5 μM with HeLa cells in a 12-well plate for 1 h before washing with heparin and digesting with trypsin. The cytosol was extracted using Digitonin buffer, which selectively permeabilizes the outer membrane. RIPA buffer was used to prepare whole cell lysates. To confirm cytosolic extraction, a portion of each sample was analyzed via Western blot using a cytosolic marker (Erk 1/2) and a late-endosomal marker (Rab5). Samples of cytosolic extract have markedly reduced Rab5 while all samples contain Erk 1/2 (Fig. 4.16A). Trace contamination of Rab5 is observed in the cytosolic extract although the amount is significantly reduced compared to the whole cell lysate. Finally, as with the biotinylated peptides, the PMO-CPPs were then extracted from the samples with Streptavidin-coated magnetic Dynabeads, washed extensively, and analyzed via MALDI-TOF. We indeed detected PMO-D-R8 and PMO-D-Bpep in their respective samples, presenting the first instance of an intact peptide-oligonucleotide conjugate being extracted from cells and analyzed by mass spectrometry (Fig. 4.16B-C). Moreover, we found that incubation at lower temperature appeared to inhibit cytosolic localization of PMO-D-CPPs, but resulted in equal relative concentrations between whole cell and cytosolic fractions for biotin-D-CPPs, which may be able to passively diffuse through the membrane.

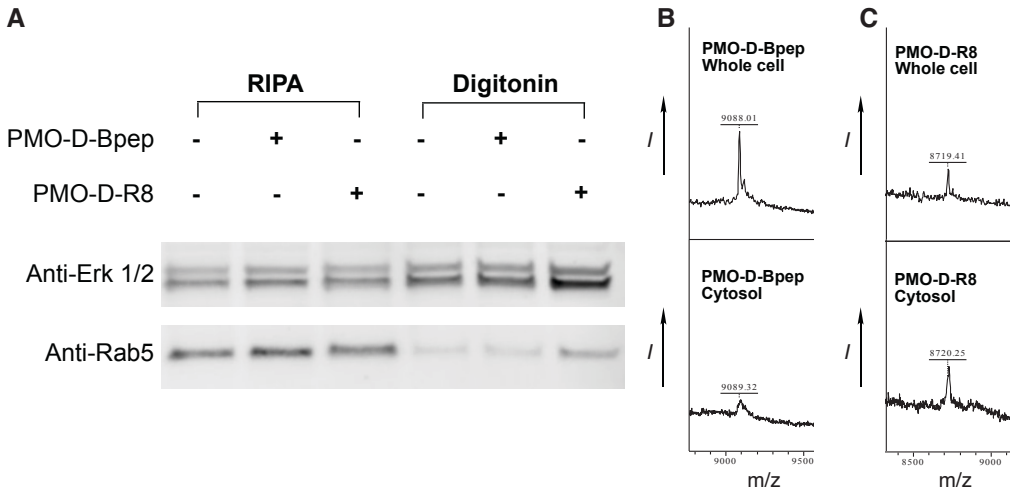


Figure 4.16 PMO-D-CPPs enter via endocytosis, and can be detected in whole cell and cytosolic lysate by MALDI-TOF. (A) Shown are PMO activities from an EGFP 654 assay in which cells are treated in a pulse-chase format with chemical endocytosis inhibitors followed by PMO-conjugates at 5 μM. Chlorpromazine (CPZ) produces a dose-dependent inhibition of PMO activity. Bars represent group mean ± SD, N = 3 distinct samples from a single biological replicate. Statistical significance is shown as *p<0.05, **p<0.005 compared to cells treated without inhibitor, determined using a student's paired two-tailed t-test. (B) Western blot demonstrating extraction of whole cell lysate and cytosolic fraction with RIPA and digitonin buffer, respectively. Erk 1/2 is a cytosolic marker, whereas Rab5 is a late endosomal marker. (C) and (D) show example MALDI spectra following uptake analysis of lysates from (B) containing PMO-D-Bpep and PMO-D-R8, respectively. Intact construct was detected in the whole cell (top) as well as cytosolic (bottom) fractions.

We confirmed by orthogonal means that the PMO-peptide conjugates entered via energy-dependent uptake and that outer membrane-bound conjugates do not contaminate lysate samples. First, an EGFP assay determined that for both PMO-D- and L-DPV7, incubation at reduced temperature negatively impacted PMO delivery (Fig. 4.17). Then, HeLa cells were incubated with three PMO-D-CPPs at 37 °C or 4 °C before washing and lysis as before. Analysis by Western blot shows presence of both cytosolic and endosomal markers in both whole cell lysates, but shows a marked absence of biotinylated construct in the 4 °C condition by Streptavidin labeling. Analysis of these samples by MALDI-TOF also shows significantly reduced signal in the 4 °C condition compared to 37C, where only PMO-D-DPV7 is detected at reduced temperature. At the same time, no construct was detected in the 4 °C cytosolic condition. By using the constructs response factor (F) from the equimolar condition, we found that PMO-D-DPV7 had the highest relative intracellular concentration, although the three constructs had very similar concentrations.

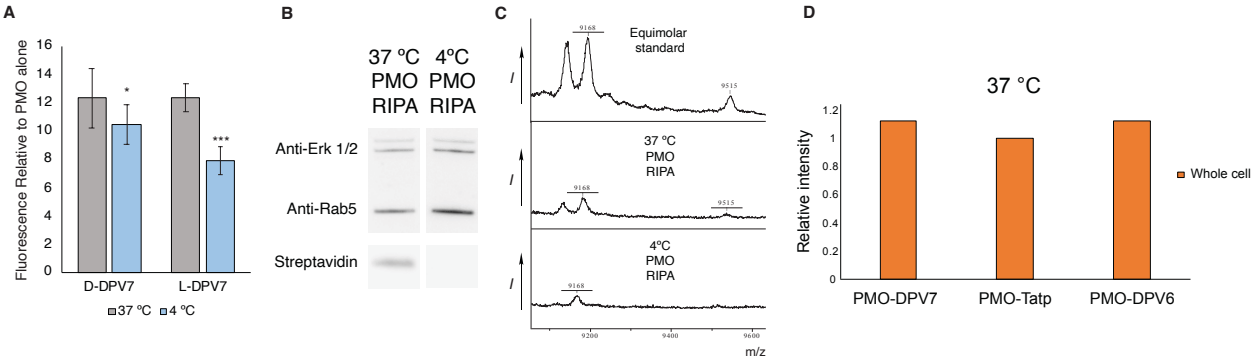


Figure 4.17 Incubation at 4 °C inhibits internalization of PMO-D-CPPs. (A) Bar graph showing fluorescence relative to PMO alone as measured in the EGFP activity assay, comparing uptake of D- and L-DPV7 at 5 μM at 37 °C to 4 °C. Bars represent group mean ± SD, N = 3 distinct samples from a single biological replicate (*$p<0.05$, ***$p<0.0005$) (B) Western blot indicating the presence of PMO-biotin-CPPs in whole cell lysate following 1 h incubation at 37 °C, but not at 4 °C. (C) MALDI-TOF mass spectra corresponding to equimolar standard (top), whole cell lysate following treatment at 37 °C (middle) and 4 °C (bottom). Only PMO-D-DPV7 is observable after incubation at 4 °C, indicating that uptake of these conjugates is arrested at low temperature. (D) Graph showing relative concentration of the three PMO-D-CPPs in whole cell lysate following incubation at 37 °C for 1 h, performed in singlicate.

Interestingly, reduced temperature did not inhibit uptake of the biotinylated-CPPs without the oligonucleotide cargo, but rather equalized the relative concentration of the whole cell and cytosolic fractions. In the same manner as the experiment with PMO-D-CPPs, biotin-D-DPV7, TATp, and DPV6 were incubated in HeLa cells at 37 °C and 4 °C. Cytosol and whole cell lysate were extracted, confirmed by Western blot (SI). Samples were then analyzed by MALDI and the relative concentrations were determined, normalized to TATp. At 37 °C, the determined relative calculations vary between the whole cell and cytosolic samples; DPV7 appears to have the highest concentration in the cytosol (Fig 4.18-4.19). However, at reduced temperature, the relative concentrations are nearly identical between the lysates. This observation is unsurprising when its considered that low temperature arrests endocytosis, meaning that the only material inside the cell likely entered through passive diffusion through the membrane to directly access the cytosol.[44,45] It is also unsurprising that CPPs with a single small biotin label are able to directly translocate, whereas peptides attached to macromolecules such as PMO cannot.[46] These experiments demonstrate that this method, in addition to determining relative internal concentration, is useful for investigating mechanisms of uptake.

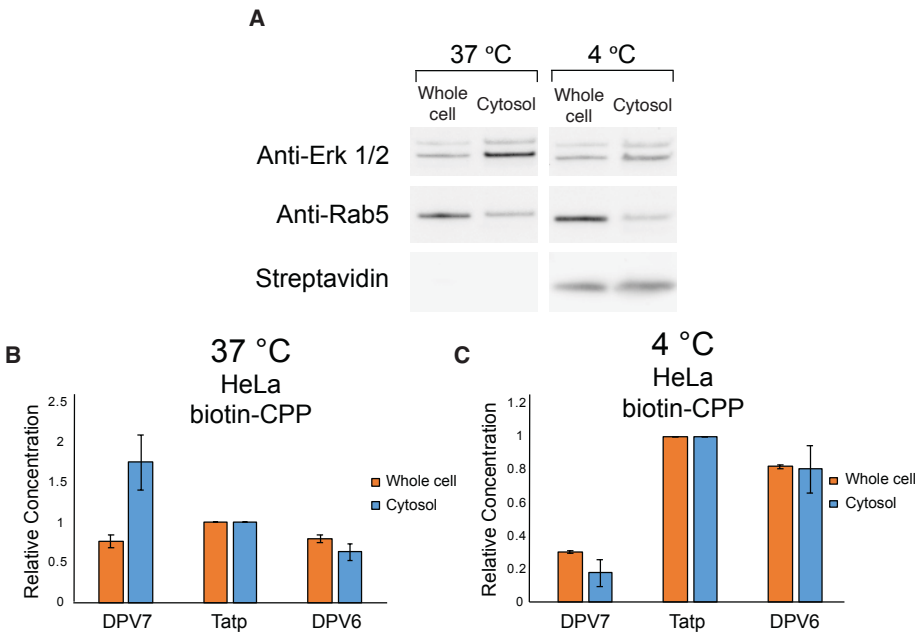


Figure 4.18 Biotin-D-CPPs were incubated with HeLa cells at 37 °C or 4 °C for 4 h. (A) Western blot showing cytosolic extraction of the samples analyzed by MALDI. Concentrations of biotin-peptide relative to TATp in the whole cell and cytosolic extracts following 4 h treatment at (B) 37 °C and at (C) 4 °C. Bars show group mean ± SD, N = 2 distinct samples from a single biological replicate. Experiment was repeated with shorter incubation time with similar results, shown in Fig 4.19.

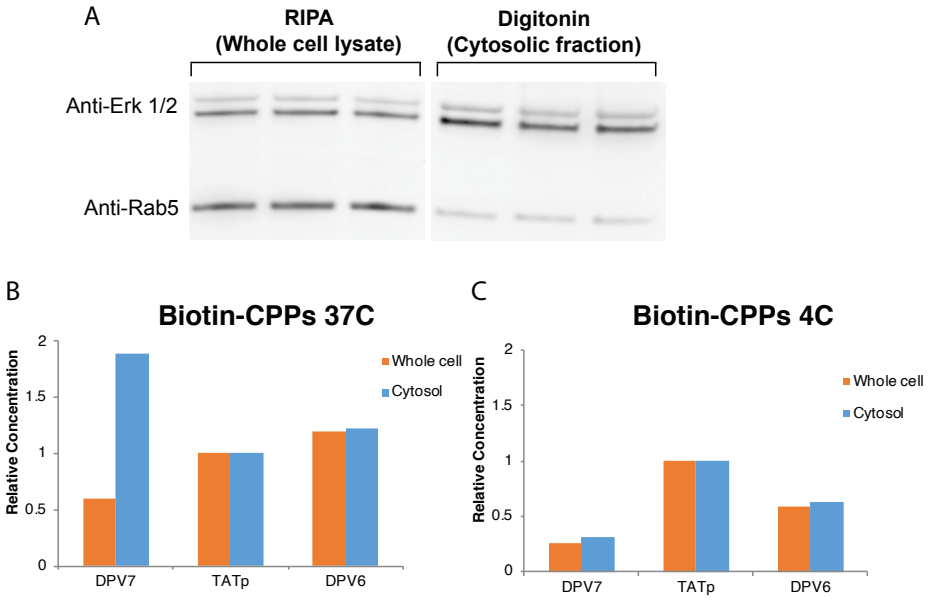


Figure 4.19 Additional Temperature experiment. Biotin-D-CPPs were incubated with HeLa cells at 37C or 4C for 1 h. (A) Western blot showing cytosolic extraction of the samples analyzed by MALDI. (B-C) Graph showing relative concentration of biotin-CPPs in the whole cell and cytosolic fractions following incubation at (B) 37C and (C) 4C as determined by MALDI-TOF. Bars represent biological singlicate.

Next, we expanded the analytes tested and profiled the uptake of six biotin-D-CPPs in the cytosol of distinct cell lines. Biotin-D-R8, TAT, Bpep, DPV7, TATp, and DPV6 were profiled by MALDI in both HeLa (Fig. 4.20A) and C2C12 mouse myoblast (Fig. 4.20B) cell lines. We observed different uptake patterns between the two cell lines; polyarginine was significantly more abundant in the C2C12 cells compared to the other peptides, and DPV6 and DPV7 were not detected in the cytosol. In HeLa, polyarginine had the highest relative concentration in both cytosol and whole cell, and DPV7 was again not detected in the cytosol. However, TAT, TATp, and DPV6 all had similar relative cytosolic concentrations. With these experiments we demonstrated that this profiling platform could determine the relative concentration of six intact peptides extracted from whole cell and cytosol of two distinct cell lines.

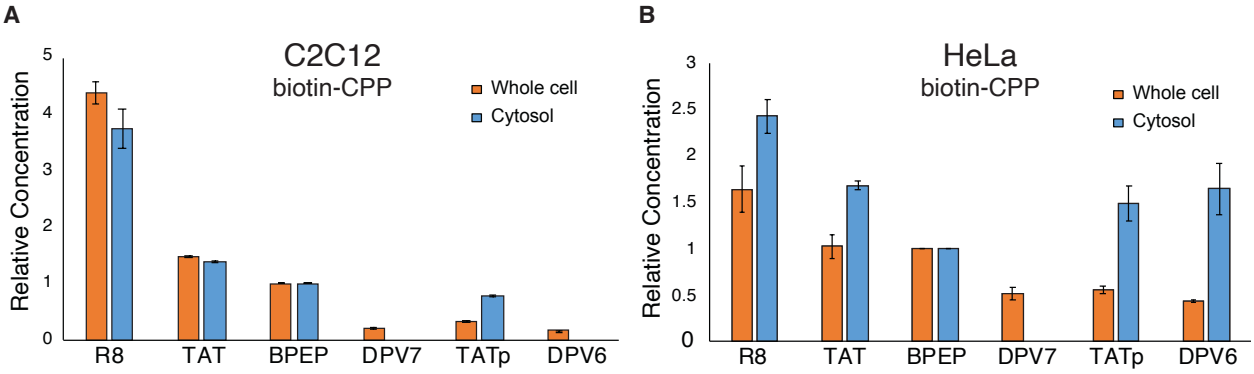


Figure 4.20 Uptake of biotin-CPPs can be profiled in different cell lines. (A) Bar graph showing concentrations of biotin-CPPs relative to BPEP in the whole cell and cytosolic extracts of C1C12 mouse myoblast cells following 1 h treatment at 37 °C. (B) Bar graph showing concentrations of biotin-CPPs in the whole cell and cytosolic extracts of HeLa cells following 1 h treatment at 37 °C. Bars represent group mean ± SD, N = 2 replicate samples. Relative concentration is normalized to Bpep.

Finally, we profiled uptake of six PMO-D-CPPs. PMO-D-R8, TAT, BPEP, DPV7, TATp, and DPV6 were profiled in HeLa cells as usual. Extraction of cytosol was confirmed by Western blot (Supporting Information Appendix I) and the samples were analyzed by MALDI. Relative concentrations were normalized to BPEP. In the whole cell extract, the relative concentrations were generally consistent across peptides, with DPV6 having 1.5-fold higher relative concentration. On the other hand, relative concentrations in the cytosol were lower for the polyarginine peptides R8 and TAT, and highest for BPEP and DPV6 (Fig. 7A). By comparing the relative concentrations in cytosolic and whole-cell lysates, the relative efficiencies demonstrate an effective metric for determining which of the six CPPs can effectively deliver PMO into the cytosol (Fig. 4.21).

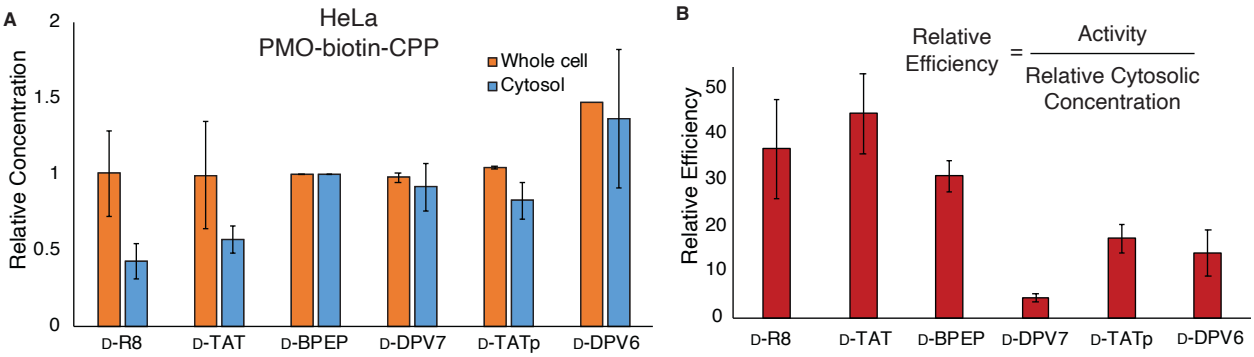


Figure 4.21 Mass spectrometry-based profiling combined with activity gives new efficiency metric for PMO-CPPs. (A) Bar graph showing concentrations of PMO-biotin-CPPs relative to BPEP in the whole cell and cytosolic extracts of Hela cells. Relative concentration is normalized to BPEP. (B) Bar graph showing relative efficiency (PMO activity / relative cytosolic concentration) of PMO-CPPs. Bars show group mean ± SD, N = 2 distinct samples from a single biological replicate, except for the whole cell condition of (A) in which N = 2 distinct samples from two independent biological replicates.

4.3 Discussion

Here we show that the all-D version of several known cell-penetrating peptides can deliver an antisense oligonucleotide to the nucleus of cells at the same efficiency as their native L-form. While the respective activities were nearly identical in cells, the D-forms were resistant to serum proteolysis. This stability enabled recovery and mass spectrometric analysis of the D-peptides from cytosolic and whole cell lysates. By comparing PMO activity to relative internalized concentrations, we obtained a metric for PMO delivery efficiency.

While it would not be expected that the mirror image of all CPPs would retain delivery activity, the striking similarities observed for the mirror image peptides selected in this study suggests that D-peptides should be studied further. Despite similar activity in cells, the difference in proteolytic stability indicates the potential for differential activities in animals. Proteolytic degradation of peptide therapeutics has been considered a major weakness limiting therapeutic investigation.[47] However, mutations to D-amino acids, or the study of entirely D-peptides, has the potential to improve pharmacokinetic properties of proteinogenic therapies. Avoidance of cleavage sites can enhance proteolytic stability and half-life in vivo, and can be achieved by integration of unnatural amino acids, including D-amino acids.[48] An all-D polyproline CPP was shown to be efficacious in mice[18], and PMO-D-CPPs were previously found to be the most stable among tested B-peptide analogs.[15] Moreover, it is likely that intact PMO-CPP conjugates are able to enter the nucleus as opposed to a cleaved PMO only.[22] Despite these promising findings, the activities of PMO-D-CPPs had not previously been explored. We found that in HeLa cells, several CPP sequences had nearly identical delivery activities in their D- and L-forms, while the D-form remained completely proteolytically stable.

Besides potential in vivo applications, the proteolytic stability of D-peptides also simplified their direct analysis by mass spectrometry following recovery from biological milieu. Previous reports demonstrated that biotinylated L-peptides could be recovered from inside whole cell lysate and quantified by MALDI-ToF, and cleavage products were identified and accounted for.[7,34] Here, by focusing only on fully intact constructs, we analyzed a mixture of six individual biotinylated D-peptides from inside cells by observing only the intact parent ions. Moreover, we demonstrated relative quantification of a mixture of biotinylated peptides recovered from cytosolic extract. Analysis of the cytosolic portion is critical when studying and developing cell-penetrating

peptides, considering that whole cell uptake does not correlate with cytosolic delivery due to the possibility of endosomal entrapment.

Finally, we showed relative quantification of intact biotinylated PMO-D-CPPs in whole cell lysate. By combining relative internal concentrations with PMO delivery activity, we obtained a metric for relative delivery efficiency. Delivery efficiency can be a useful metric for comparing CPPs delivering active therapeutic cargo as it takes into account both the activity of the cargo as well as the internal concentration. A highly efficient peptide would have high activity with low internal concentration, and thus a high relative efficiency. Using these metrics when investigating CPPs would be useful in narrowing the scope of sequences early in development to exclude sequences which accumulate in the endosomes or otherwise do not efficiently deliver active cargo.

Exploration of mirror image CPPs would enter into a largely untapped chemical space. We recently described a machine learning-based platform for the discovery of nuclear-targeting peptides containing unnatural amino acids.[23] While the unnatural residues included in this work were non-alpha-amino acids, future applications of this method could include D-amino acid substitutions, or fully mirror image CPPs in order to discover highly active, novel sequences that are completely stable.

4.4 Materials & Methods

4.4.1 Reagents and Solvents

H-Rink Amide-ChemMatrix resin was obtained from PCAS BioMatrix Inc. (St-Jean-sur-Richelieu, Quebec, Canada). 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3- oxid-hexafluorophosphate (HATU), 4-pentynoic acid, 5-azidopentanoic acid, Fmoc-β-Ala-OH, Fmoc-6-aminohexanoic acid, and Fmoc-L-Lys(N_3) were purchased from Chem-Impex International (Wood Dale, IL). PyAOP was purchased from P3 BioSystems (Louisville, KY). Fmoc-protected L-amino acids (Fmoc-Ala-OHxH_2O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(O*t*-Bu)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(O*t*-Bu)-OH; Fmoc-Gly-OH; Fmoc-His(Trt)-OH; Fmoc-Ile-OH; Fmoc-Leu-OH; Fmoc-Lys(Boc)-OH; Fmoc-Met-OH; Fmoc-Phe-OH; Fmoc-Pro-OH; Fmoc-Ser(But)-OH; Fmoc-Thr(*t*-Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(*t*-Bu)-OH; Fmoc-Val-OH), were purchased from the Novabiochem-line from Sigma Millipore. Fmoc-protected D-amino acids were purchased from Chem Impex (Wood Dale, IL). Peptide

synthesis-grade *N,N*-dimethylformamide (DMF), CH_2Cl_2, diethyl ether, and HPLC-grade acetonitrile were obtained from VWR International (Radnor, PA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO). Milli-Q water was used exclusively.

4.4.2 Liquid-chromatography mass-spectrometry

LCMS analyses were performed on either an Agilent 6520 or 6545 Accurate-Mass Q-TOF LCMS (abbreviated as 6520 or 6545) coupled to an Agilent 1260 Infinity HPLC system, or an Agilent 6550 iFunnel Q-TOF LCMS system (abbreviated as 6550) coupled to an Agilent 1290 Infinity HPLC system. Mobile phases were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LCMS methods were used for characterization:

Method A: 1-61% B over 9 min, Zorbax C3 column (6520 and 6545)

LC: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, 12-15 min 95% B; flow rate: 0.8 mL/min.
MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 300–3000 *m*/*z*. MS is on from 4 to 11 min.

Method B: 1-91% B over 9 min, Zorbax C18 column (6520 and 6545)

LC: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-91% B, 11-12 min 91-95% B, 12-15 min 95% B; flow rate: 0.8 mL/min.
MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 300–3000 *m*/*z*. MS is on from 4 to 11 min.

Method C: 1-61% B over 10 min, Phenomenex Jupiter C4 column (6550)

LC: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min 1-61% B, 12-16 min 61-90% B; 16-20 min 90% B; flow rate: 0.1 mL/min.
MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 100–1700 *m*/*z*. MS is on from 4 to 12 min.

Method D: 1-61% B over 10 min, Agilent EclipsePlus C18 column (6550)

LC: Agilent EclipsePlus C18 RRHD column: 2.1 × 50 mm, 1.8 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min, 1-61% B, 12-13 min, 61% B, 13-16 min, 1% B; flow rate: 0.1 mL/min.

MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 300–3000 m/z. MS is on from 4 to 12 min. This method was used exclusively for characterization of the modular library.

All data were processed using Agilent MassHunter software package. Y-axis in all chromatograms shown represents total ion current (TIC) unless noted.

4.4.3 General peptide preparation

Fast-flow Peptide Synthesis: Peptides were synthesized on a 0.1 mmol scale using an automated fast-flow peptide synthesizer for L-peptides and a semi-automated fast-flow peptide synthesizer for D-peptides. Automated synthesis conditions were used as previously reported.[49] Briefly, a 100 mg portion of ChemMatrix Rink Amide HYR resin was loaded into a reactor maintained at 90 ºC. All reagents were flowed at 40 mL/min with HPLC pumps through a stainless-steel loop maintained at 90 ºC before introduction into the reactor. For each coupling, 10 mL of a solution containing 0.4 M amino acid and 0.38 M HATU in DMF were mixed with 600 µL diisopropylethylamine and delivered to the reactor. Fmoc removal was accomplished using 10.4 mL of 20% (v/v) piperidine. Between each step, DMF (15 mL) was used to wash out the reactor. To couple unnatural amino acids or to cap the peptide (e.g. with 4-pentynoic acid), the resin was incubated for 30 min at room temperature with amino acid (1 mmol) dissolved in 2.5 mL 0.4 M HATU in DMF with 500 µL diisopropylethylamine. After completion of the synthesis, the resin was washed 3 times with dichloromethane and dried under vacuum.

Semi-automated synthesis was carried out as previously described.[50] 1 mmol of amino acid was combined with 2.5 mL 0.4 M HATU and 500 µL DIEA and mixed before being delivered to the reactor containing resin via syringe pump at 6 mL/min. The reactor was submerged in a water bath heated to 70 ºC. An HPLC pump delivered either DMF (20 mL) for washing or 20 % piperidine/DMF (6.7 mL) for Fmoc deprotection, at 20 mL/min.

Peptide Cleavage and Deprotection: Each peptide was subjected to simultaneous global side-chain deprotection and cleavage from resin by treatment with 5 mL of 94% trifluoroacetic acid (TFA), 2.5% thioanisole, 2.5% water, and 1% triisopropylsilane (TIPS) (v/v) at room temperature for 2 to 4 hours. The cleavage cocktail was first concentrated by bubbling N_2 through the mixture, and cleaved peptide was precipitated and triturated with 40 mL of cold ether (chilled in dry ice).

The crude product was pelleted by centrifugation for three minutes at 4,000 rpm and the ether was decanted. This wash step was repeated two more times. After the third wash, the pellet was dissolved in 50% water and 50% acetonitrile containing 0.1% TFA, filtered through a fritted syringe to remove the resin and lyophilized.

Peptide Purification: The peptides were dissolved in water and acetonitrile containing 0.1% TFA, filtered through a 0.22 µm nylon filter and purified by mass-directed semi-preparative reversed-phase HPLC. Solvent A was water with 0.1% TFA additive and Solvent B was acetonitrile with 0.1% TFA additive. A linear gradient that changed at a rate of 0.5% B/min was used. Most of the peptides were purified on an Agilent Zorbax SB C18 column: 9.4 x 250 mm, 5 µm. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

4.4.4 Preparation of PMO-peptides

PMO IVS-654 (50 mg, 8 µmol) was dissolved in 150 µL DMSO. To the solution was added a solution containing 2 equivalents of dibenzocyclooctyne acid (5.3 mg, 16 µmol) activated with HBTU (37.5µL of 0.4 M HBTU in DMF, 15 µmol) and DIEA (2.8 µL, 16 µmol) in 40 µL DMF (Final reaction volume = 0.23 mL). The reaction proceeded for 25 min before being quenched with 1 mL of water and 2 mL of ammonium hydroxide. The ammonium hydroxide hydrolyzed any ester formed during the course of the reaction. After 1 hour, the solution was diluted to 40 mL in water/acetonitrile and purified using reverse-phase HPLC (Agilent Zorbax SB C3 column: 21.2 x 100 mm, 5 µm) and a linear gradient from 2 to 60% B (solvent A: water; solvent B: acetonitrile) over 58 min (1% B / min). Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

Conjugation to peptides

PMO-DBCO (1 eq, 5 mM, water) was conjugated to azido-peptides (1.5 eq, 5 mM, water) at room temperature for 2 h. Reaction progress was monitored by LCMS and purified when PMO-DBCO was consumed. Purification was conducted using mass-directed HPLC (Solvent A: 100 mM ammonium acetate in water, Solvent B: acetonitrile) with a linear gradient that changed at a rate of 0.5% B/min, on an Agilent Zorbax SB C13 column: 9.4 x 250 mm, 5 µm. Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

4.4.5 EGFP Assay

HeLa 654 cells obtained from the University of North Carolina Tissue Culture Core facility were maintained in MEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin-streptomycin at 37 °C and 5% CO_2. 18 h prior to treatment, the cells were plated at a density of 5,000 cells per well in a 96-well plate in MEM supplemented with 10% FBS and 1% penicillin-streptomycin.

For individual peptide testing, PMO-peptides were dissolved in PBS without Ca^{2+} or Mg^{2+} at a concentration of 1 mM (determined by UV) before being diluted in MEM. Cells were incubated at the designated concentrations for 22 h at 37 °C and 5% CO_2. Next, the treatment media was removed, and the cells were washed once before being incubated with 0.25 % Trypsin-EDTA for 15 min at 37 °C and 5% CO_2. Lifted cells were transferred to a V-bottom 96-well plate and washed once with PBS, before being resuspended in PBS containing 2% FBS and 2 μg/mL propidium iodide (PI). Flow cytometry analysis was carried out on a BD LSRII flow cytometer at the Koch Institute. Gates were applied to the data to ensure that cells that were positive for propidium iodide or had forward/side scatter readings that were sufficiently different from the main cell population were excluded. Each sample was capped at 5,000 gated events.

Analysis was conducted using Graphpad Prism 7 and FlowJo. For each sample, the mean fluorescence intensity (MFI) and the number of gated cells was measured. To report activity, triplicate MFI values were averaged and normalized to the PMO alone condition. For the final set of PMO-peptides evaluated, three biological replicates were performed.

4.4.6 Endocytosis Inhibition Assay

Chemical endocytosis inhibitors were used to probe the mechanism of delivery of PMO by these peptides in a pulse-chase format. We have conducted such analysis on similar PMO-peptide constructs previously with comparable outcomes.[22] For the PMO constructs, HeLa 654 cells were preincubated with various chemical inhibitors for 30 minutes before treatment with PMO-CPP constructs for three hours. The panel included: a panel of endocytosis inhibitors including: chlorpromazine (CPZ), which is demonstrated to interfere with clathrin-mediated endocytosis; cytochalasin D (CyD), which inhibits phagocytosis and micropinocytosis; wortmannin (Wrt), which alters various endocytosis pathways by inhibiting phosphatidylinositol kinases; EIPA (5-

(N-ethyl-Nisopropyl) amiloride), which inhibits micropinocytosis; and Dynasore (Dyn), which also inhibits clathrin-mediated endocytosis.[51,52] Treatment media was then replaced with fresh media and the cells were incubated for 22 hours at 37 °C and 5% CO_2. Cells were then lifted as previously described and EGFP synthesis was measured by flow cytometry.

4.4.7 LDH Assay

Cytotoxicity assays were performed in HeLa 654 cells. Cell supernatant following treatment for flow cytometry was transferred to a new 96-well plate for analysis of LDH release. To each well of the 96-well plate containing supernatant was added CytoTox 96 Reagent (Promega). The plate was shielded from light and incubated at room temperature for 30 minutes. Equal volume of Stop Solution was added to each well, mixed, and the absorbance of each well was measured at 490 nm. The blank measurement was subtracted from each measurement, and % LDH release was calculated as % cytotoxicity = 100 × Experimental LDH Release (OD490) / Maximum LDH Release (OD490).

4.4.8 Serum stability assay

Each PMO-peptide was dissolved in PBS to a concentration of 1 mM, as confirmed by UV-Vis. PMO-peptide was then added to a solution of either PBS or PBS containing 25% human serum to a final concentration of 50 μM and incubated at 37 °C. 10 μL aliquots were removed at varying timepoints (t = 0, 1h, 6 h, 24 h), and quenched with 20 μL 1M guanidinium hydrochloride and 50 mM EDTA. 50 μL ice-cold acetonitrile was then added and the aliquots were flash frozen until LCMS analysis. Samples were thawed, and a portion of the aqueous layer was diluted before analysis by LC-qTOF. The mass spectrum for the PMO-peptides was analyzed by deconvolution, to best demonstrated whether the analyte had stayed intact or degraded.

4.4.9 Uptake Assay

<u>Cell treatment</u>

Cells were plated either in 6-well or 12-well plates at a density such that they reached 80% confluency the following day. CPP or PMO-CPP stock solutions were made fresh to 1 mM in cation-free PBS, as determined by UV-Vis. Treatment solution was then prepared by adding the

stock solution to cell media at the concentrations described. Two wells were left untreated as controls. The plates were then incubated at 37 °C and 5% CO_2 for the designated time. For the experiment to arrest energy-dependent uptake, the plate was incubated at 4 °C. Following incubation, the cells were washed three times with media, followed by 0.1 mg/mL Heparin in PBS for 5 min. Supernatant was aspirated and cells were lifted by incubating in trypsin-EDTA for 10 min at 37 °C. Trypsin was quenched by adding cell media, and cells were transferred to Eppendorf tubes and pelleted at 500 rcf for 3 min. Pellets were washed by mixing with PBS, repeated twice.

Lysis

To acquire whole cell lysate, 50 µL RIPA (1x RIPA, protease inhibitor cocktail, water) was added to the cell pellet, mixed gently, and placed on ice for 1 h. To extract the cytosol, 50 µL digitonin buffer (0.05 mg/mL digitonin, 250 mM sucrose, PBS) was added to a cell pellet, mixed very gently, and placed on ice for 10 min. Samples were then pelleted by centrifugation at 16,000 rcf for 5 min. Supernatants were transferred to new Eppendorf tubes and kept on ice. Extracted protein was quantified using Pierce Rapid Gold BCA Protein Assay Kit (Thermo Fisher). 10 µg protein from each sample was then analyzed by SDS-PAGE gel for 35 min at 165 V and then transferred to a nitrocellulose membrane soaked in 48 mM Tris, 39 mM glycine, 0.0375% SDS, 20% methanol using a TransBlot Turbo Semi-Dry Transfer Unit (BioRad) for 7 mins. The membrane was blocked at 4 °C overnight in LI-Cor Odyssey blocking buffer (PBS). The membrane was then immunostained for 1 h with anti-Erk1/2 and anti-Rab5 in TBST at room temperature. After incubation, the membrane was washed three times with TBST and incubated with the appropriate secondary antibody in TBST for 1 h at room temperature, then washed with TBST. Finally, the membrane was incubated with streptavidin-HRP for 1 h and washed with TBST. To visualize HRP, the membrane was treated with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Fisher) immediately before imaging on a ChemiDoc MP Imaging System (Bio-Rad).

MALDI-TOF

The remaining cell extracts were then used for affinity capture and MALDI-TOF analysis, following an adapted protocol.[35] 10 µL Dynabeads™ MyOne™ Streptavidin T1 (Thermo Fisher) were transferred to tubes in a magnet stand and washed with PBS. Cell extracts were added to the

corresponding bead-containing tube and rotated at 4 °C overnight. One tube contained beads that were added to an equimolar solution of peptide conjugates used in the experiment. To insure the same equivalency are in the control tube as used in the experiment, this control solution was taken directly from the combined stock solution used in the initial cell treatment. The following day, the beads slurries were washed with a series of buffers: 2 x 100 µL Buffer A (50 mM Tris-HCl (pH 7.4) and 0.1 mg/mL BSA), 2 x 100 µL Buffer B (50 mM Tris-HCl (pH 7.4), 0.1 mg/mL BSA, and 0.1% SDS), 2 x 100 µL Buffer C (50 mM Tris-HCl (pH 7.4), 0.1 mg/mL BSA, and 1 M NaCl), and 2 x 100 µL water. Beads were then incubated with 100 µL of 1 mM biotin for 2 min, before washed with 5 x 50 µL water. Supernatant was removed and the beads were brought up in 3 µL MALDI matrix (saturated alpha-Cyano-4-hydroxycinnamic acid, CHCA) and transferred to the MALDI plate to dry. Beads were analyzed by MALDI-ToF on a high-resolution Bruker Autoflex LRF Speed mass spectrometer in linear positive mode.

Relative concentrations of peptides in the mixture were determined as follows. Analytes in a mixture ionize according to their response factor (F). F was determined by normalizing the intensities of each analyte to one analyte in the control sample, where the concentration of each analyte is arbitrarily set to 1. The values of F is then used in the experimental spectra containing the same mixture of analytes to determine their relative concentrations.

$$\frac{I_x}{[X]} = F\left(\frac{I_s}{[S]}\right)$$

4.4.10 Statistics

Statistical analysis and graphing was performed using Prism (Graphpad) or Excel (Microsoft). Concentration-response curves were fitted using Prism using nonlinear regression. The listed replicates for each experiment indicates the number of distinct samples measured for a given assay. Significance for activities between constructs was determined using a student's two-sided, unpaired t-test.

4.5 Acknowledgements

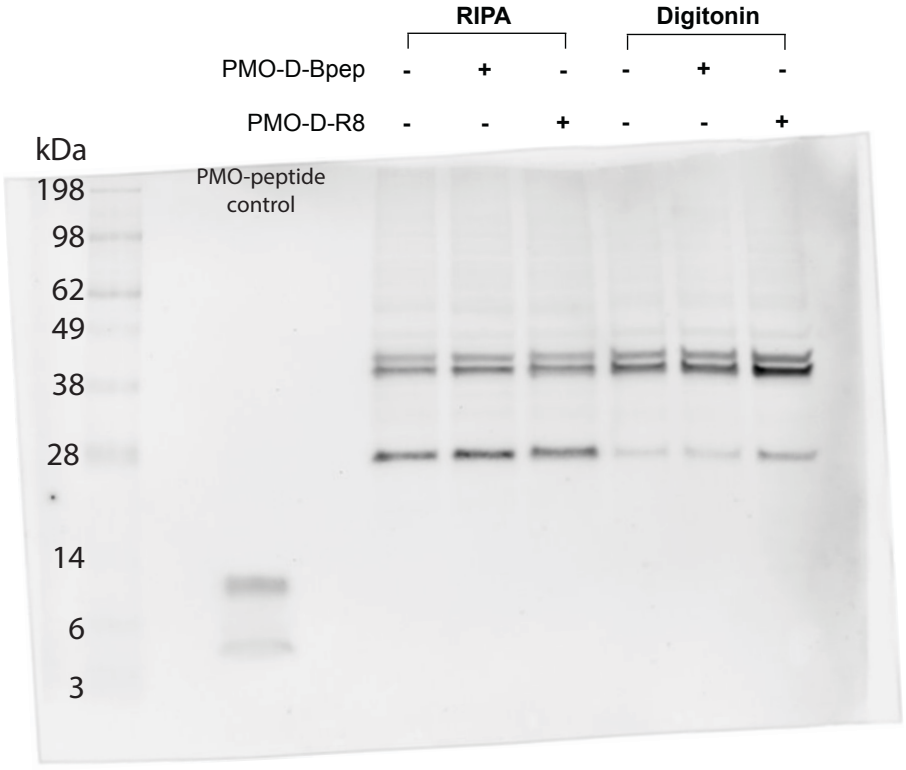
This research was funded by Sarepta Therapeutics. C.K.S. (4000057398) and C.E.F. (4000057441) acknowledge the National Science Foundation Graduate Research Fellowship (NSF Grant No. 1122374) for research support. We thank the Swanson Biotechnology Center Flow Cytometry Facility at the Koch Institute for the use of their flow cytometers, and MIT's Department of Chemistry Instrumentation Facility for the use of the MALDI-ToF.

4.6 Author contributions:

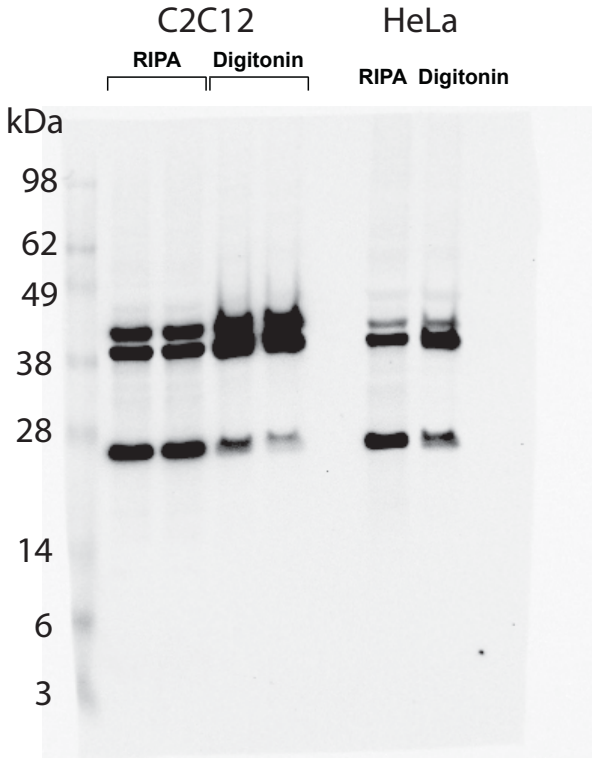
C.K.S. and B.L.P conceptualized the research. C.K.S. and C.E.F. prepared and tested PMO-peptides. C.K.S. performed uptake experiments. C.K.S. C.E.F. A.L. and B.L.P. wrote the manuscript with input from all authors.

4.7 Appendix I: Gel Images

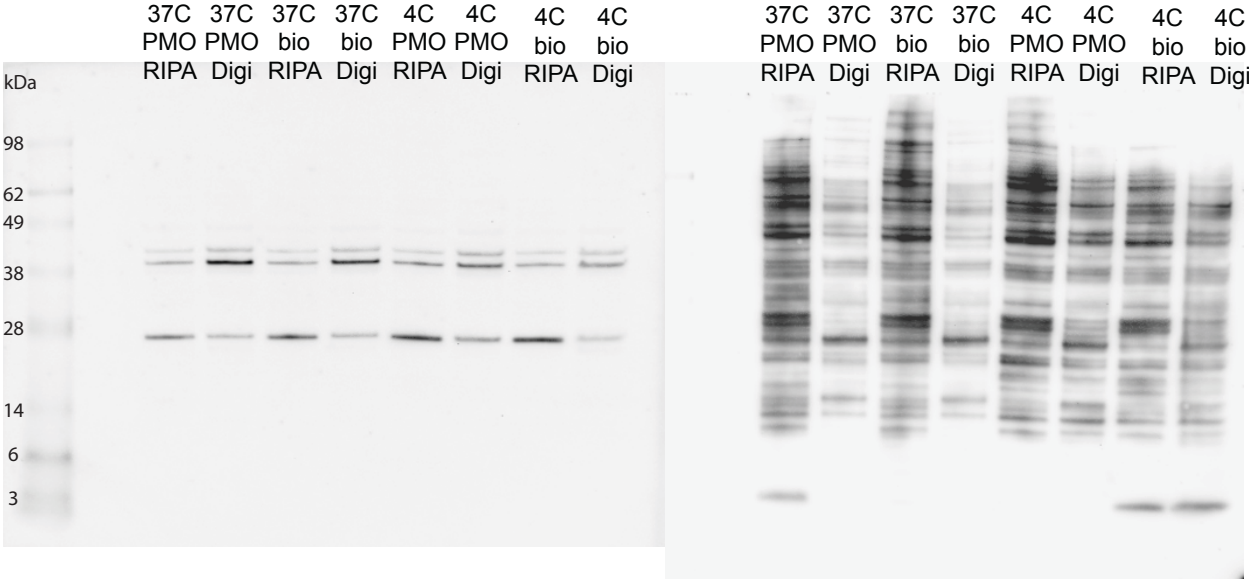
Full gel image corresponding to that shown in Figure 4.



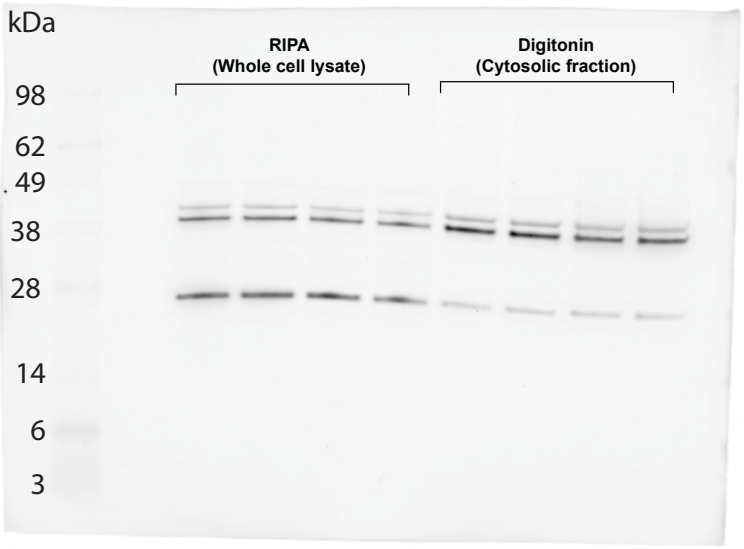
Full gel image corresponding to the experiment described in Figure 5.



Full gel image corresponding to that shown in Figure S14 and S15. IR800 (left) and Chemiluminescence (right)

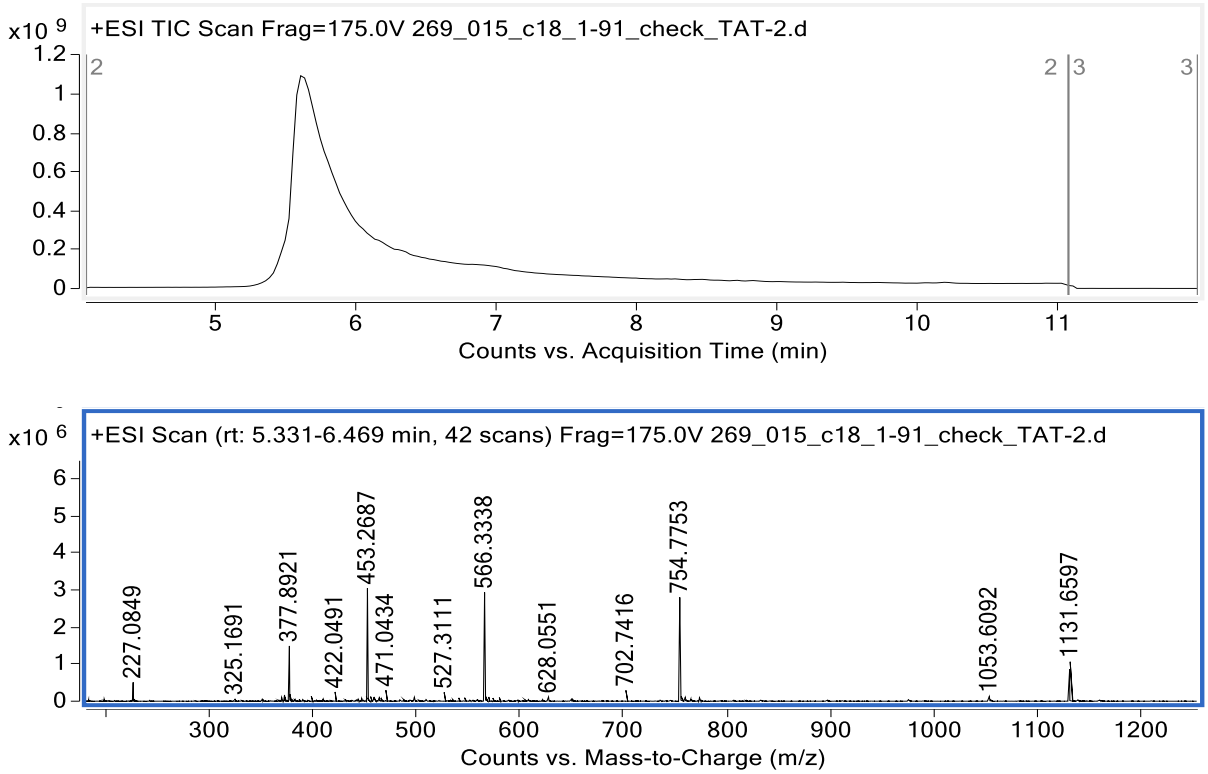


Full gel image corresponding to that shown in Figure S16

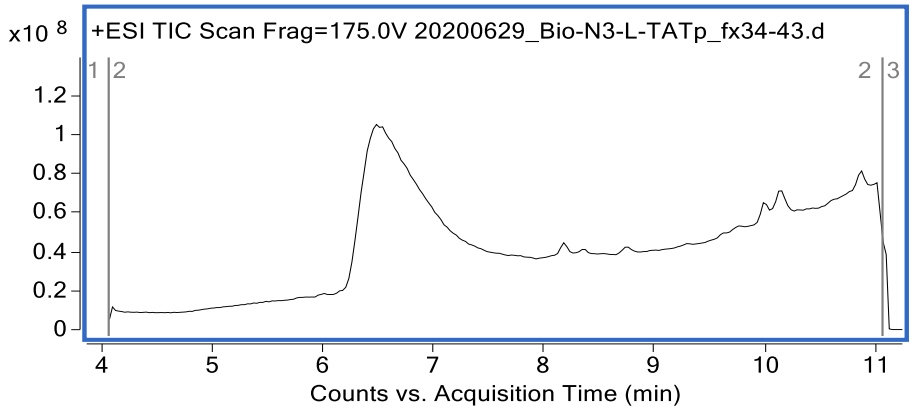


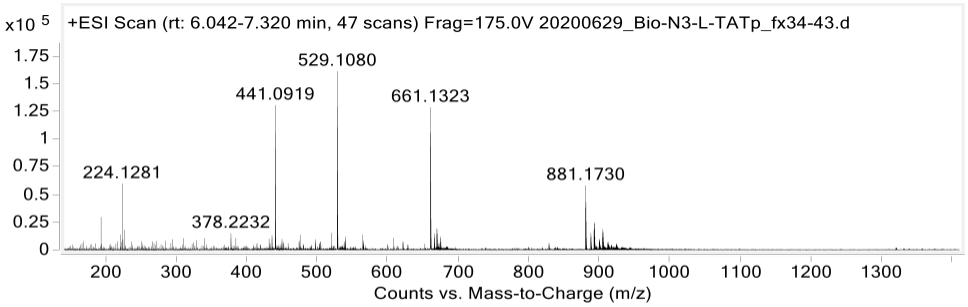
4.8 Appendix II: LC-MS Characterization

Peptide: Bio-D-TAT
Sequence: Biotin-Lys(N_3)-GGKGGWRKKRRQRRR
Calculated monoisotopic mass: 2260.3 Da
Observed monoisotopic mass: 2260.4 Da

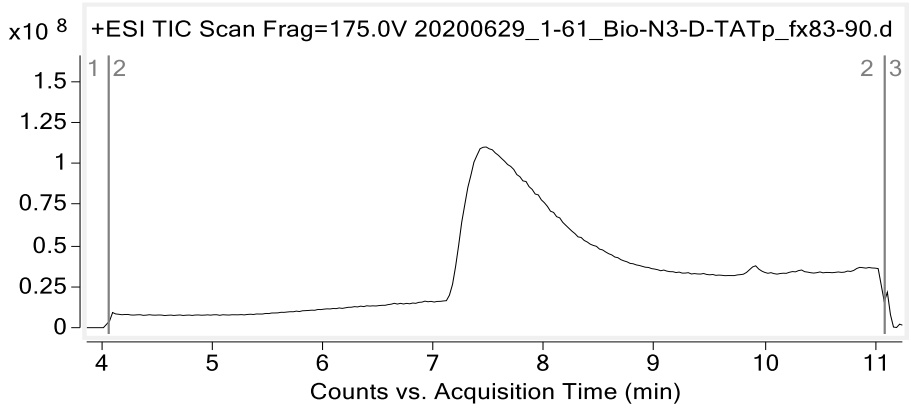


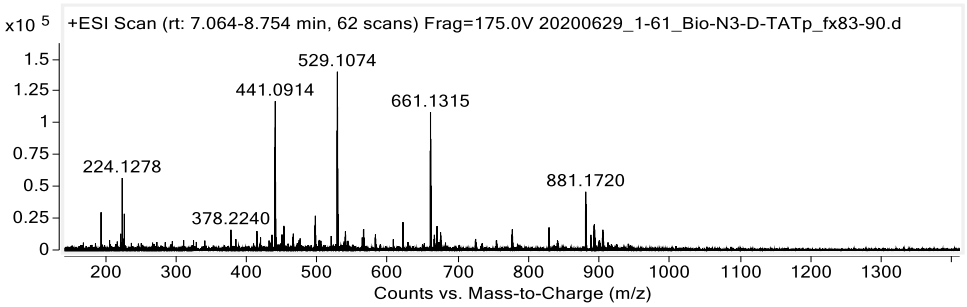
Peptide: Bio-L-TAT
Sequence: Biotin-Lys(N_3)-GGKGGWRKKRRQRRR
Calculated monoisotopic mass: 2260.3 Da
Observed monoisotopic mass: 2260.4 Da



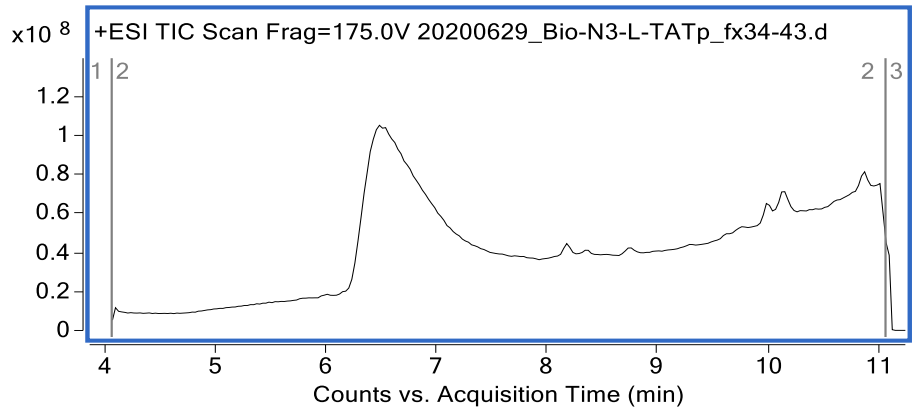


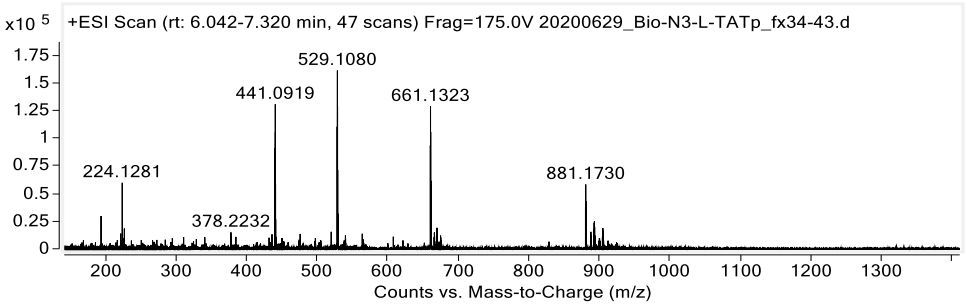
Peptide: Bio-D-TATp
Sequence: Biotin-Lys(N_3)-GGKGGWGRKKRRQRRRPPQ
Calculated monoisotopic mass: 2639.5 Da
Observed monoisotopic mass: 2639.5 Da



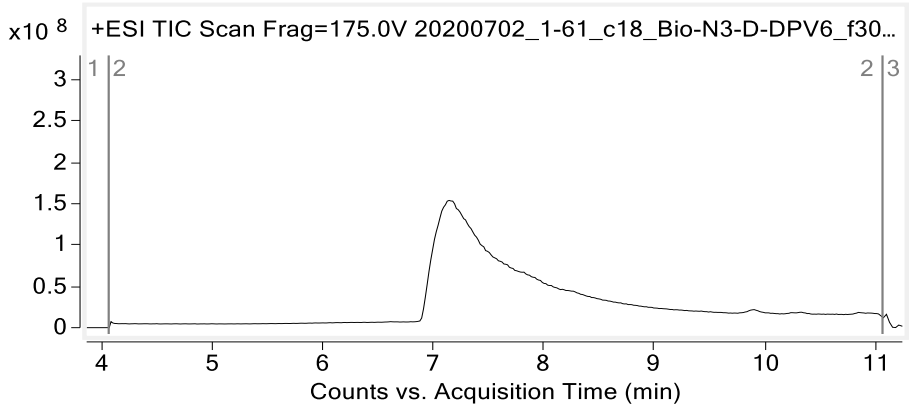


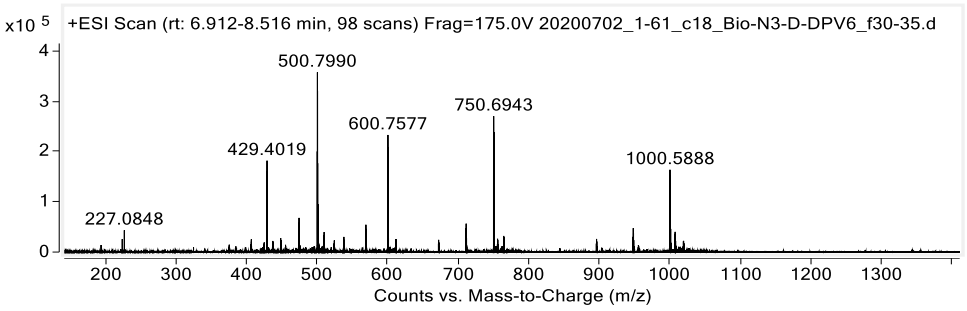
Peptide: Bio-L-TATp
Sequence: Biotin-Lys(N_3)-GGKGGWGRKKRRQRRRPPQ
Calculated monoisotopic mass: 2639.5 Da
Observed monoisotopic mass: 2639.5 Da



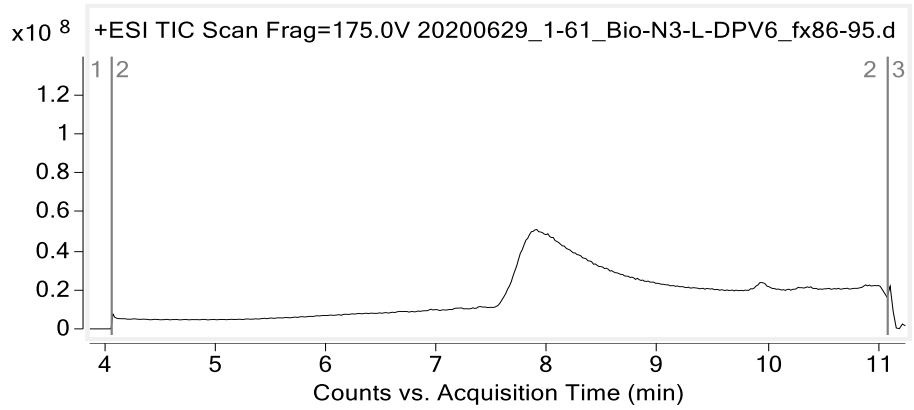


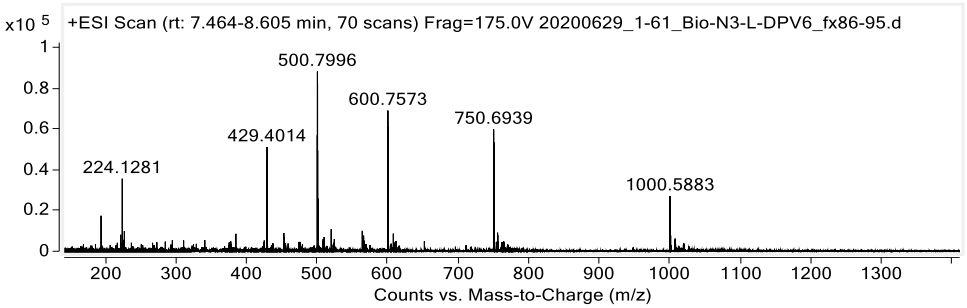
Peptide: Bio-D-DPV6
Sequence: Biotin-Lys(N_3)-GGKGGWGRPRESGKKRKRKRLKP
Calculated monoisotopic mass: 2997.7 Da
Observed monoisotopic mass: 2997.8 Da



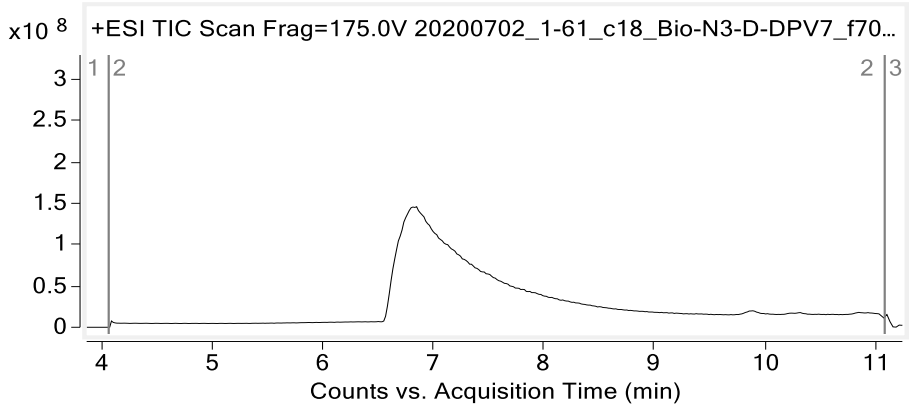


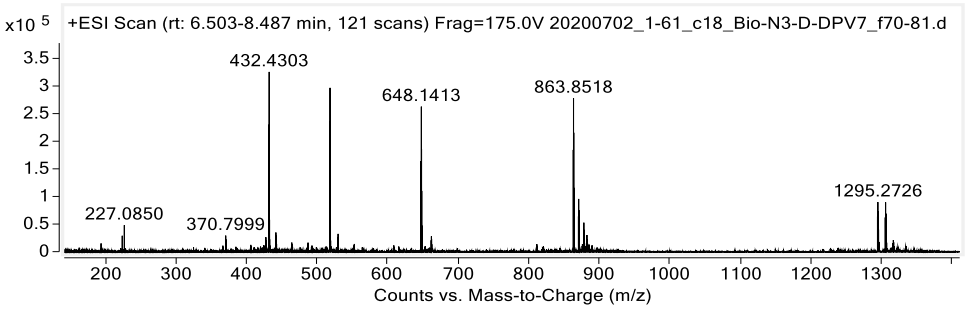
Peptide: Bio-L-DPV6
Sequence: Biotin-Lys(N_3)-GGKGGWGRPRESGKKRKRKRLKP
Calculated monoisotopic mass: 2997.7 Da
Observed monoisotopic mass: 2997.8 Da



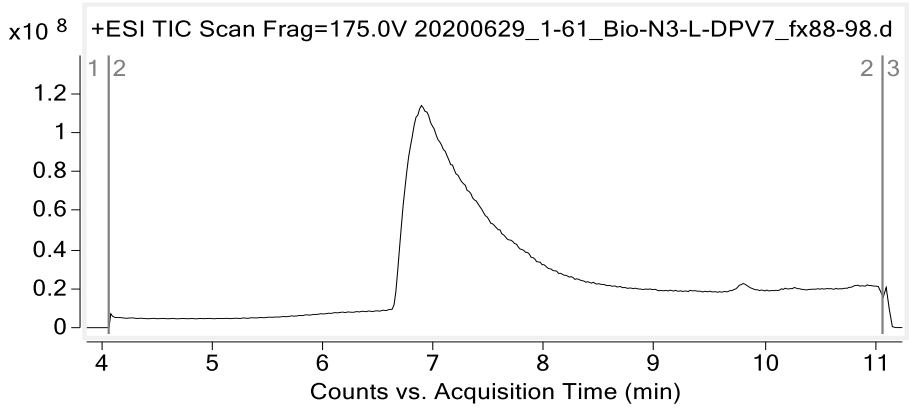


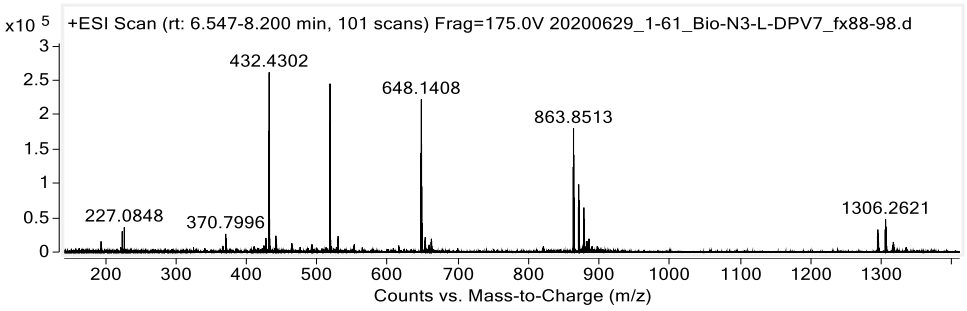
Peptide: Bio-D-DPV7
Sequence: Biotin-Lys(N_3)-GGKGGWKRKKKGKLGKKRDP
Calculated monoisotopic mass: 2587.5 Da
Observed monoisotopic mass: 2587.5 Da



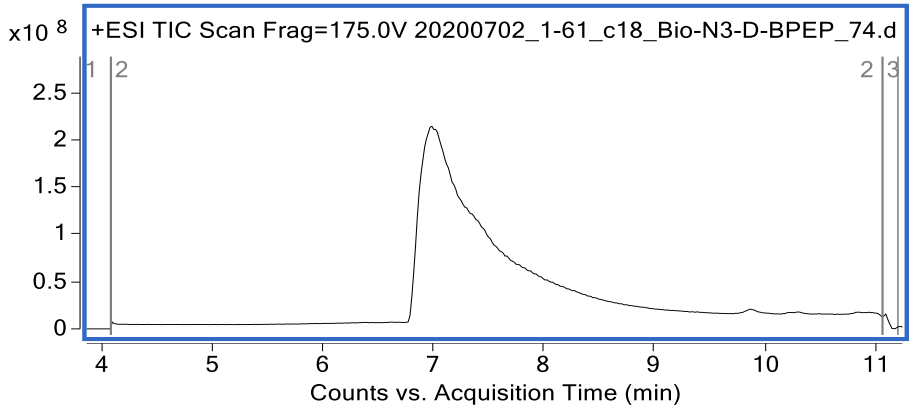


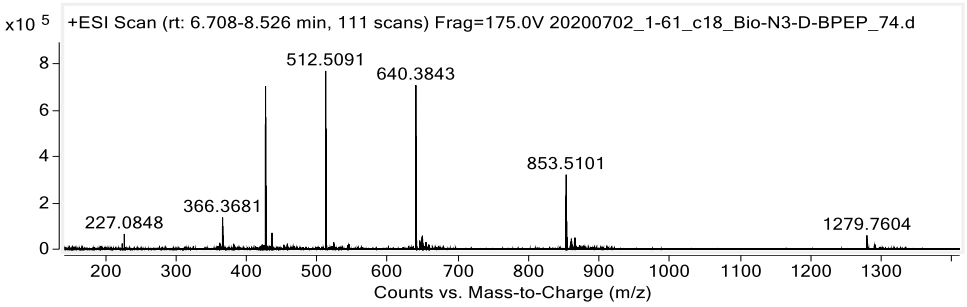
Peptide: Bio-L-DPV7
Sequence: Biotin-Lys(N_3)-GGKGGWKRKKKGKLGKKRDP
Calculated monoisotopic mass: 2587.5 Da
Observed monoisotopic mass: 2587.5 Da



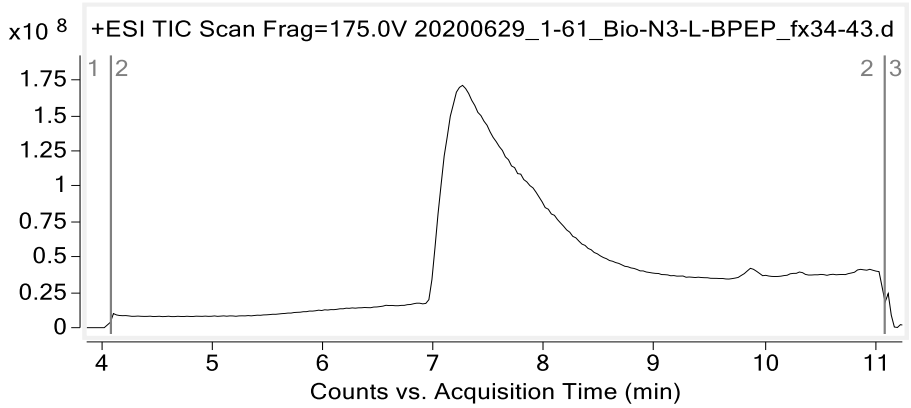


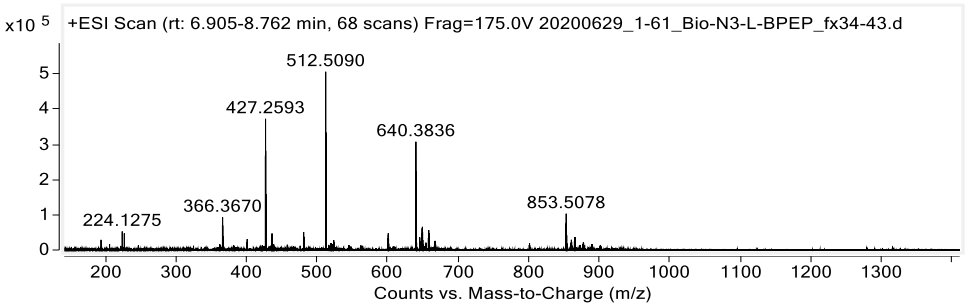
Peptide: Bio-D-BPEP
Sequence: Biotin-Lys(N_3)-GGKGGWRXRRBRRXRRBR
Calculated monoisotopic mass: 2556.5 Da
Observed monoisotopic mass: 2556.5 Da



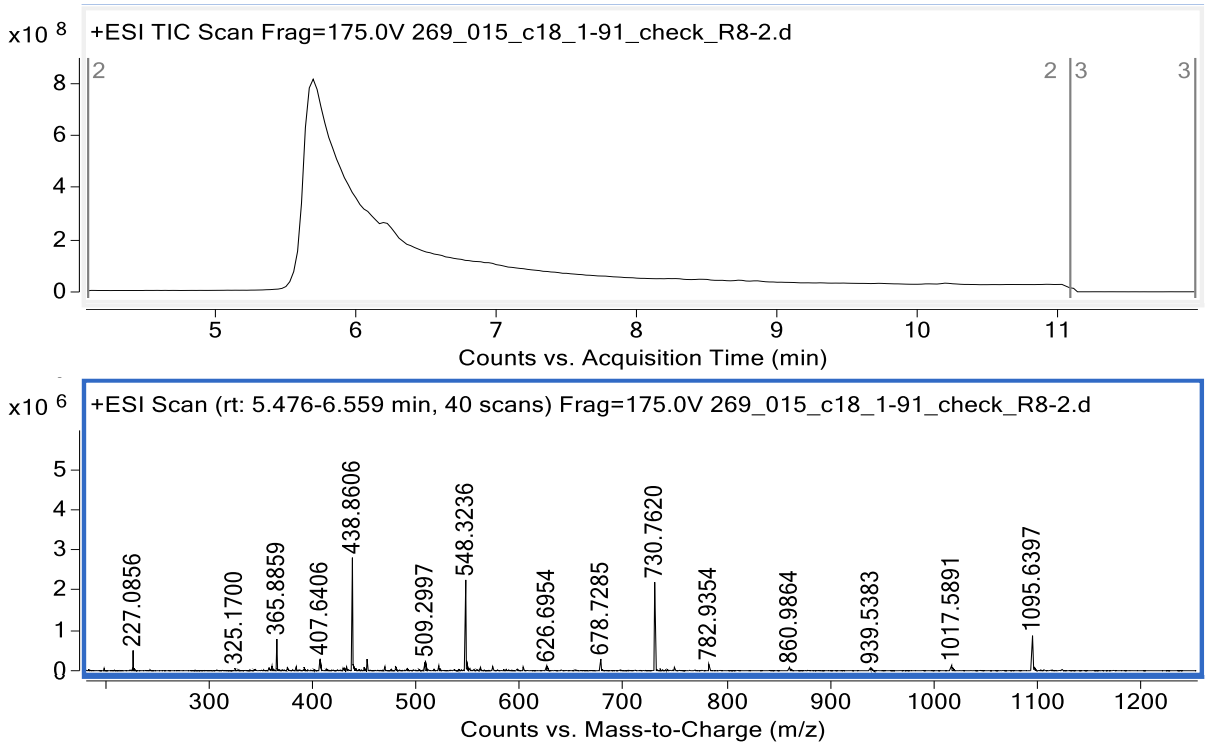


Peptide: Bio-L-BPEP
Sequence: Biotin-Lys(N_3)-GGKGGWRXRRBRRXRRBR
Calculated monoisotopic mass: 2556.5 Da
Observed monoisotopic mass: 2556.5 Da

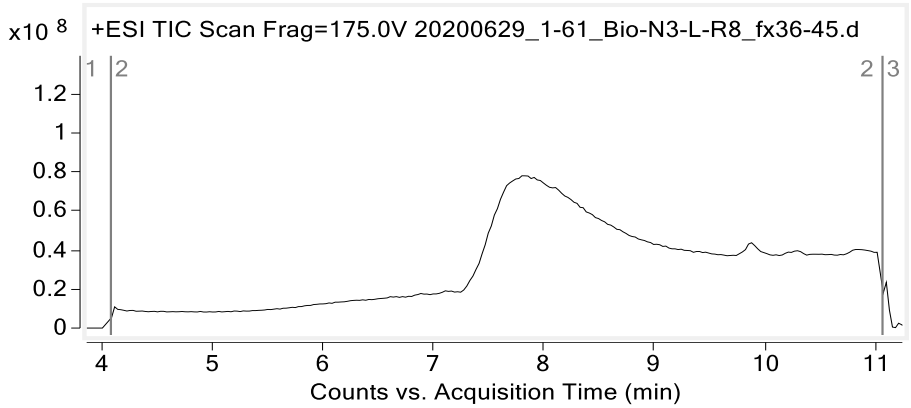


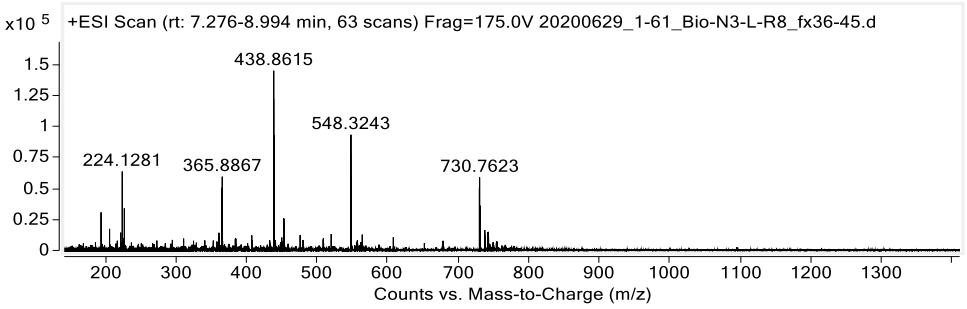


Peptide: Bio-D-R8
Sequence: Biotin-Lys(N_3)-GGKGGWRRRRRRRRR
Calculated monoisotopic mass: 2188.3 Da
Observed monoisotopic mass: 2188.3 Da

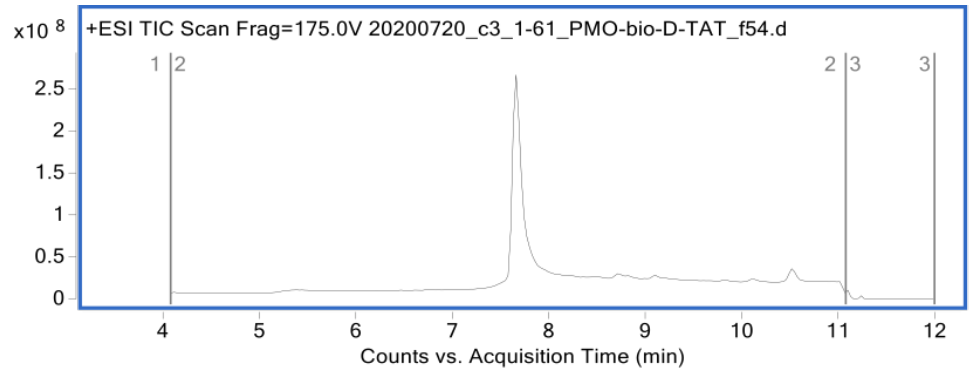


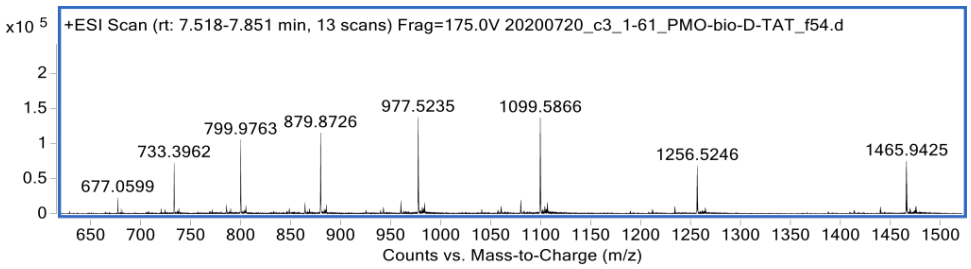
Peptide: Bio-L-R8
Sequence: Biotin-Lys(N_3)-GGKGGWRRRRRRRR
Calculated monoisotopic mass: 2188.3 Da
Observed monoisotopic mass: 2188.3 Da



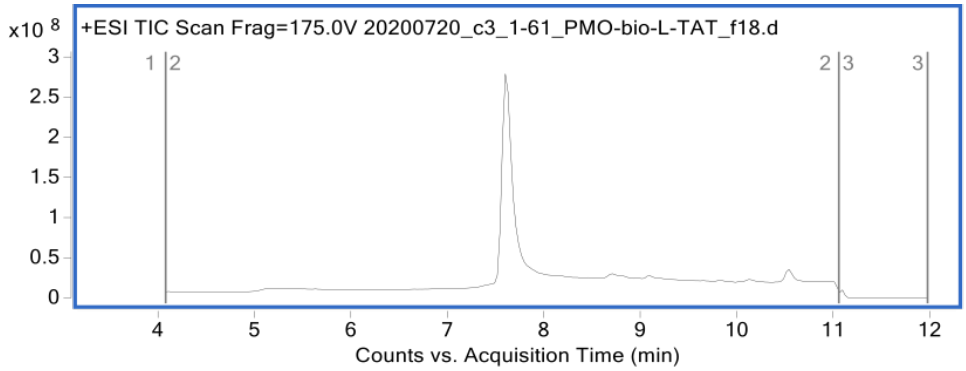


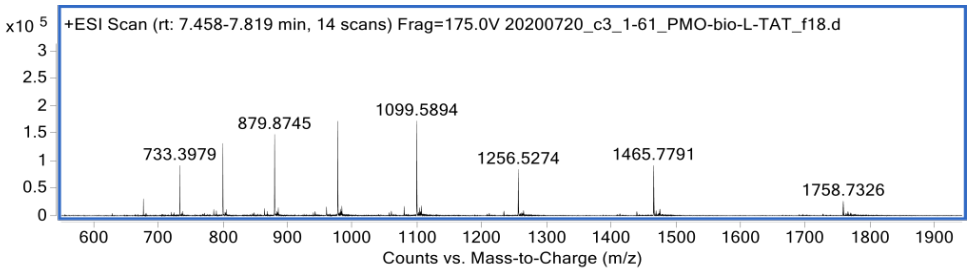
Peptide: PMO-Bio-D-TAT
Sequence: Biotin-Lys(PMO)-GGKGGWRKKRRQRRR
Calculated mass:8789.6 Da
Observed mass: 8789.8 Da



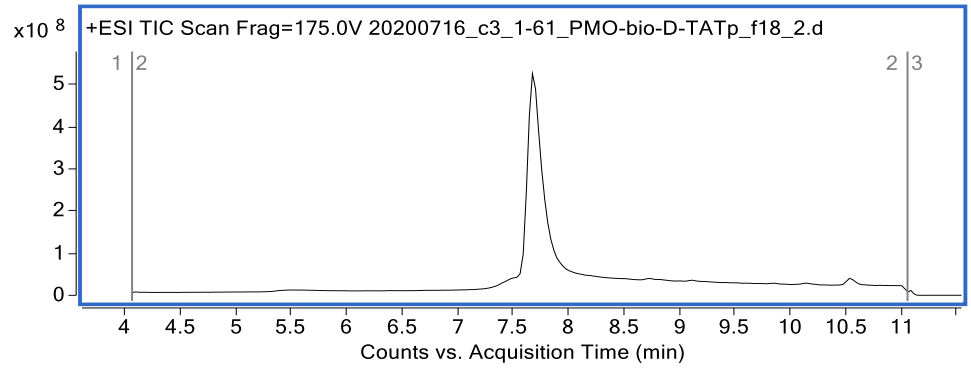


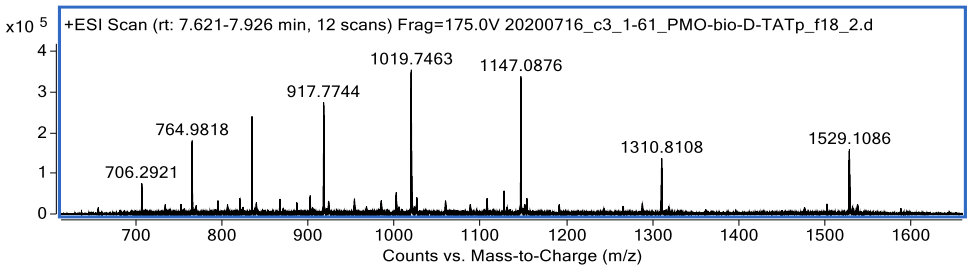
Peptide: PMO-Bio-L-TAT
Sequence: Biotin-Lys(PMO)-GGKGGWRKKRRQRRR
Calculated mass: 8789.6 Da
Observed mass: 8789.7 Da



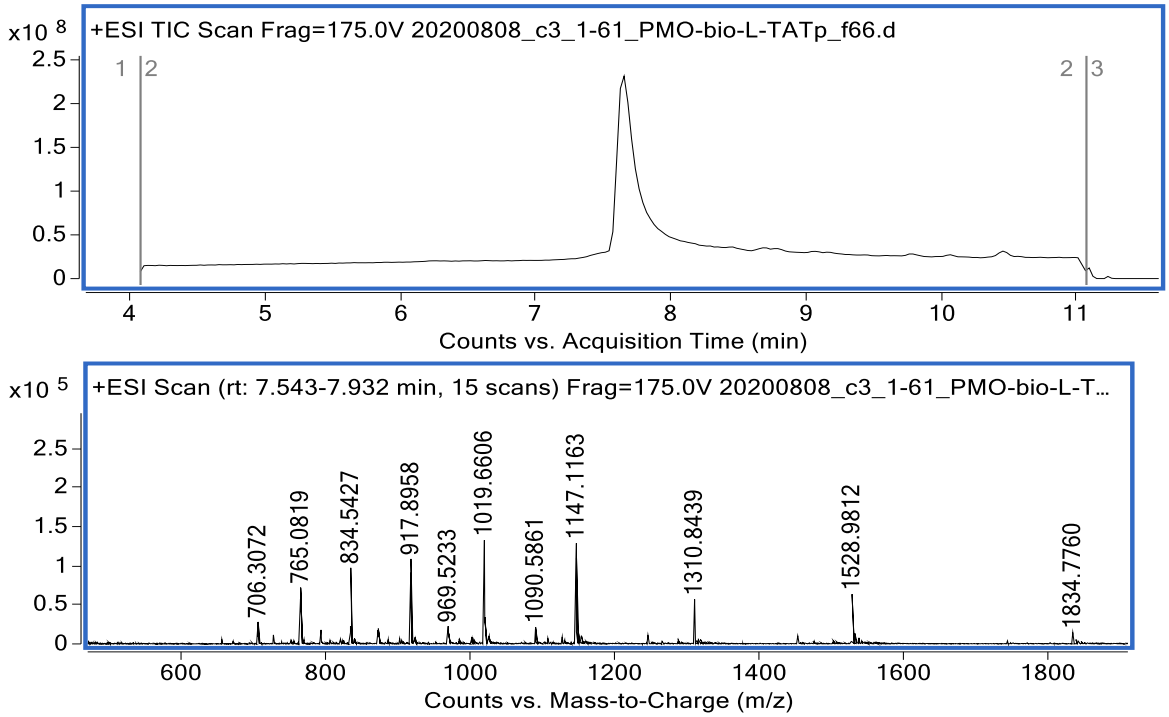


Peptide: PMO-Bio-D-TATp
Sequence: Biotin-Lys(PMO)-GGKGGWGRKKRRQRRRPPQ
Calculated mass: 9169.0 Da
Observed mass: 9169.1 Da

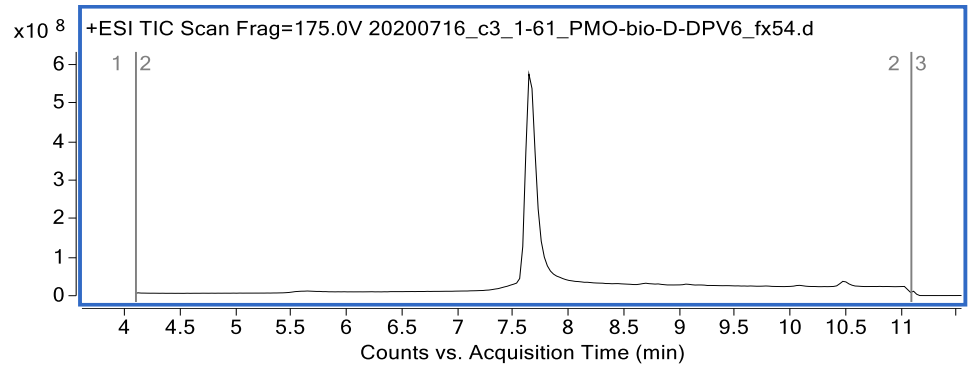


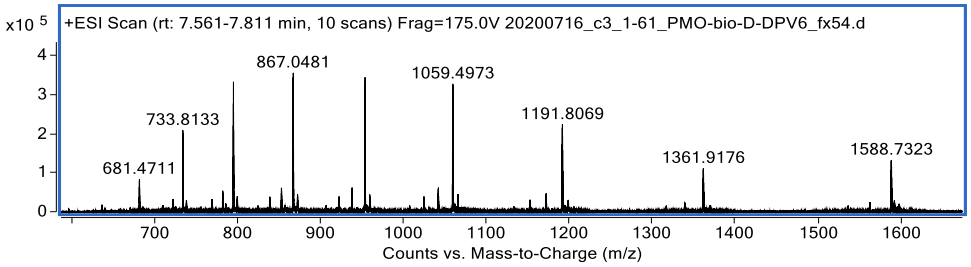


Peptide: PMO-Bio-L-TATp
Sequence: Biotin-Lys(PMO)-GGKGGWGRKKRRQRRRPPQ
Calculated mass: 9169.0 Da
Observed mass: 9169.1 Da

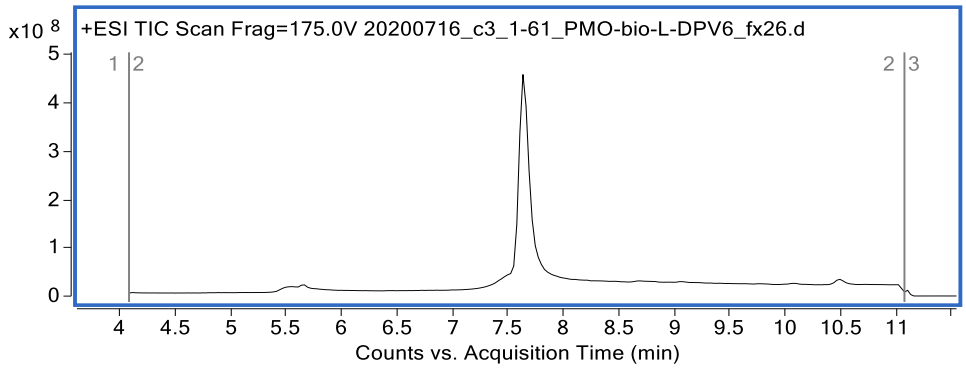


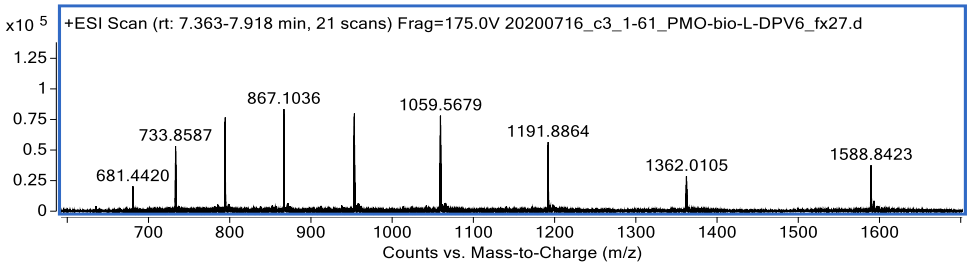
Peptide: PMO-Bio-D-DPV6
Sequence: Biotin-Lys(PMO)-GGKGGWGRPRESGKKRKRKRLKP
Calculated mass: 9527.5 Da
Observed mass: 9527.2 Da



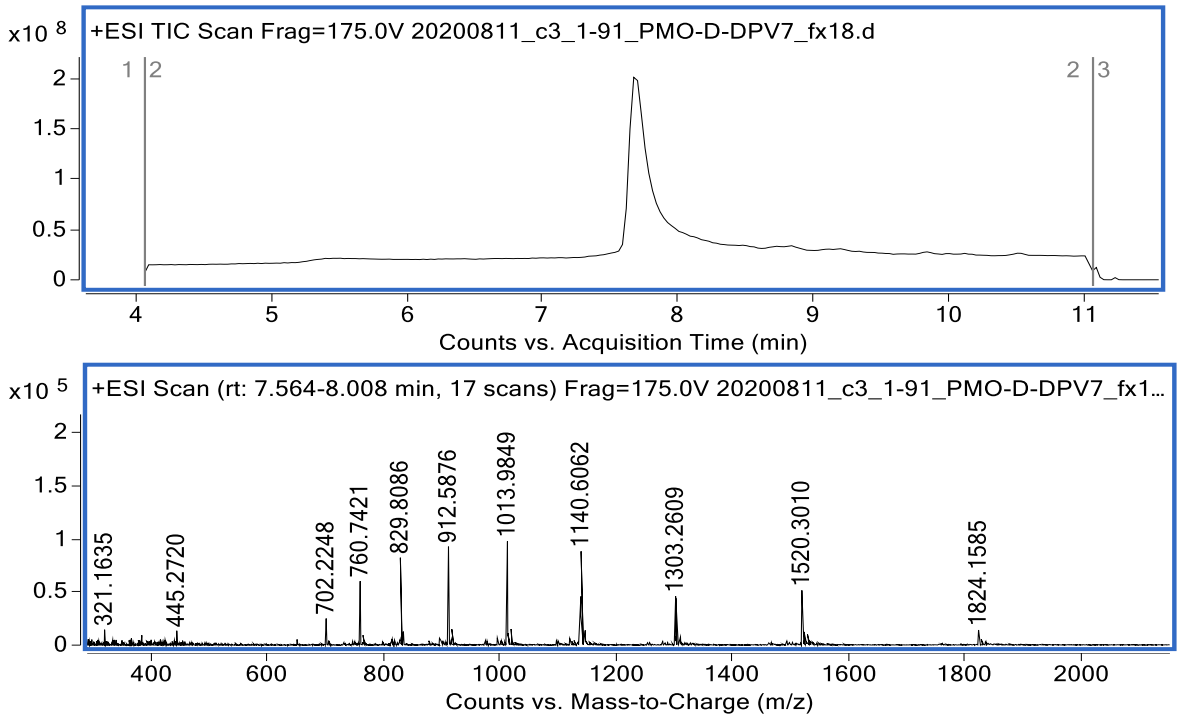


Peptide: PMO-Bio-L-DPV6
Sequence: Biotin-Lys(PMO)-GGKGGWGRPRESGKKRKRKRLKP
Calculated mass: 9527.5 Da
Observed mass: 9527.4 Da

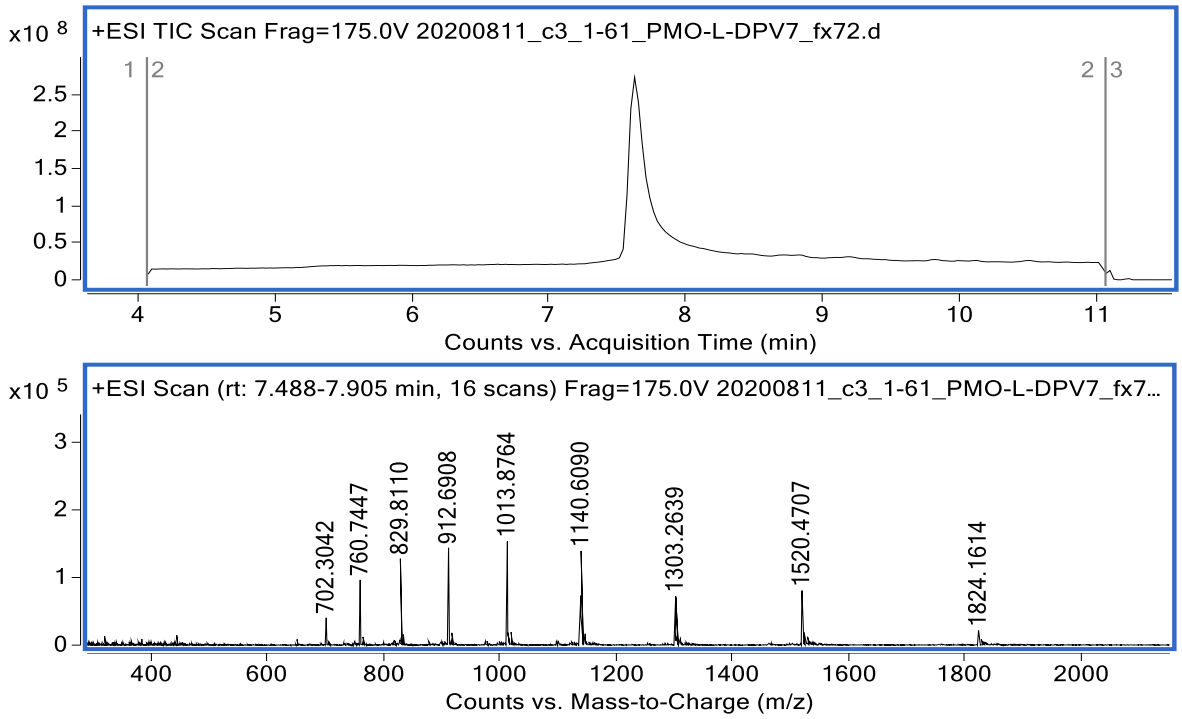




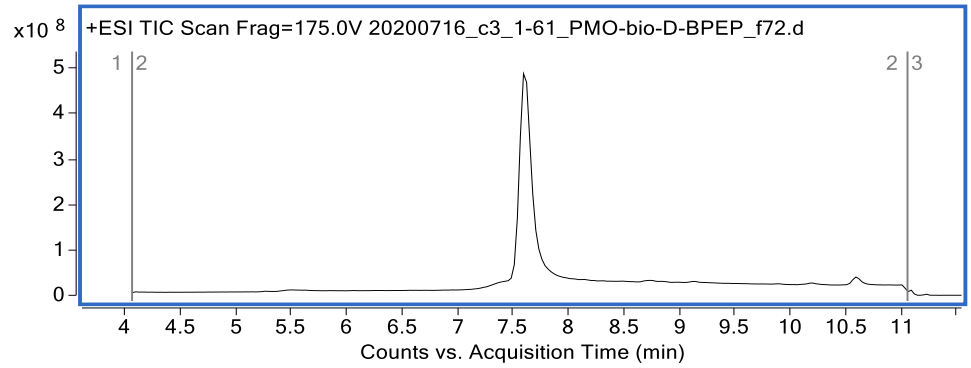
Peptide: PMO-Bio-D-DPV7
Sequence: Biotin-Lys(PMO)-GGKGGWKRKKKGKLGKKRDP
Calculated mass: 9117.1 Da
Observed mass: 9117.1 Da

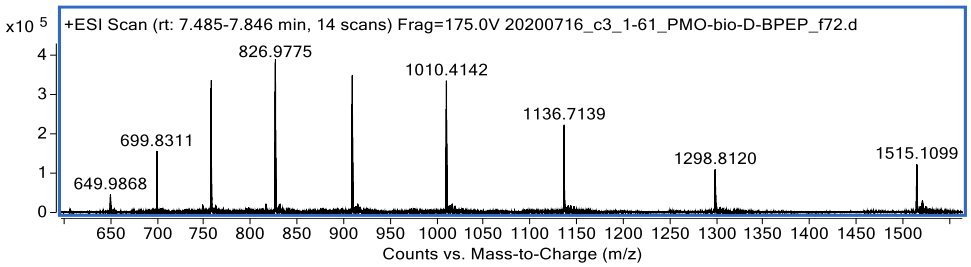


Peptide: PMO-Bio-L-DPV7
Sequence: Biotin-Lys(PMO)-GGKGGWKRKKKGKLGKKRDP
Calculated mass: 9117.1 Da
Observed mass: 9117.1 Da

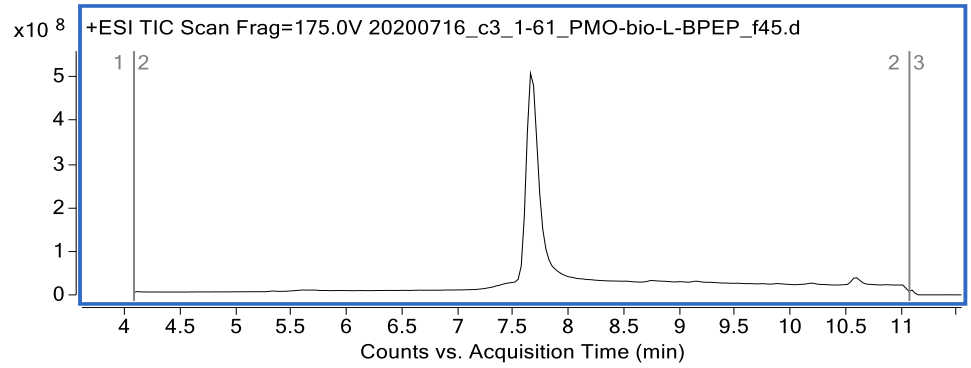


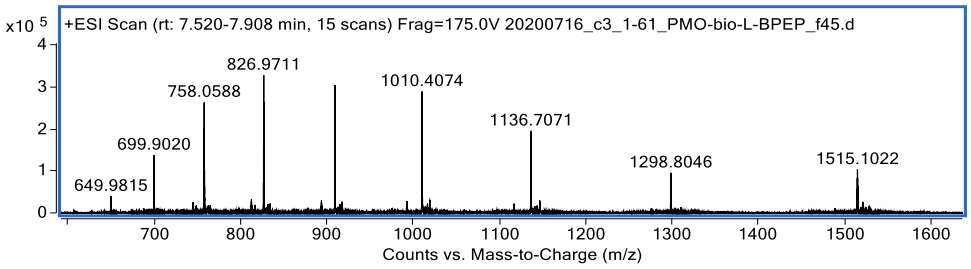
Peptide: PMO-Bio-D-BPEP
Sequence: Biotin-Lys(PMO)-GGKGGWRXRRBRRXRRBR
Calculated mass: 9086.0 Da
Observed mass: 9086.1 Da



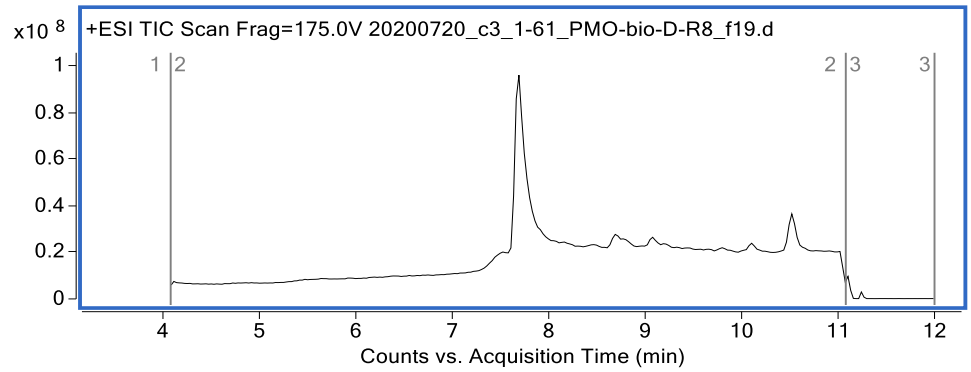


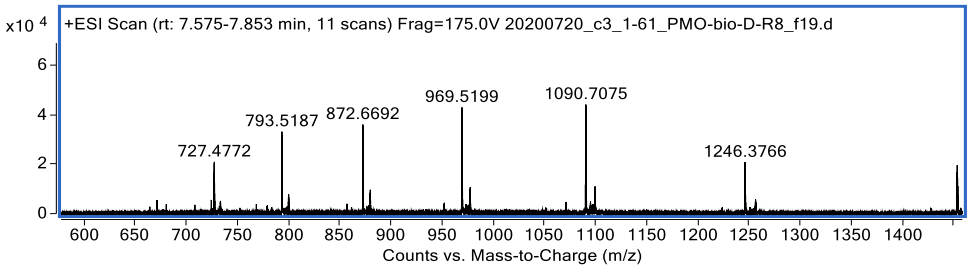
Peptide: PMO-Bio-L-BPEP
Sequence: Biotin-Lys(PMO)-GGKGGWRXRRBRRXRRBR
Calculated mass: 9086.0 Da
Observed mass: 9085.8 Da



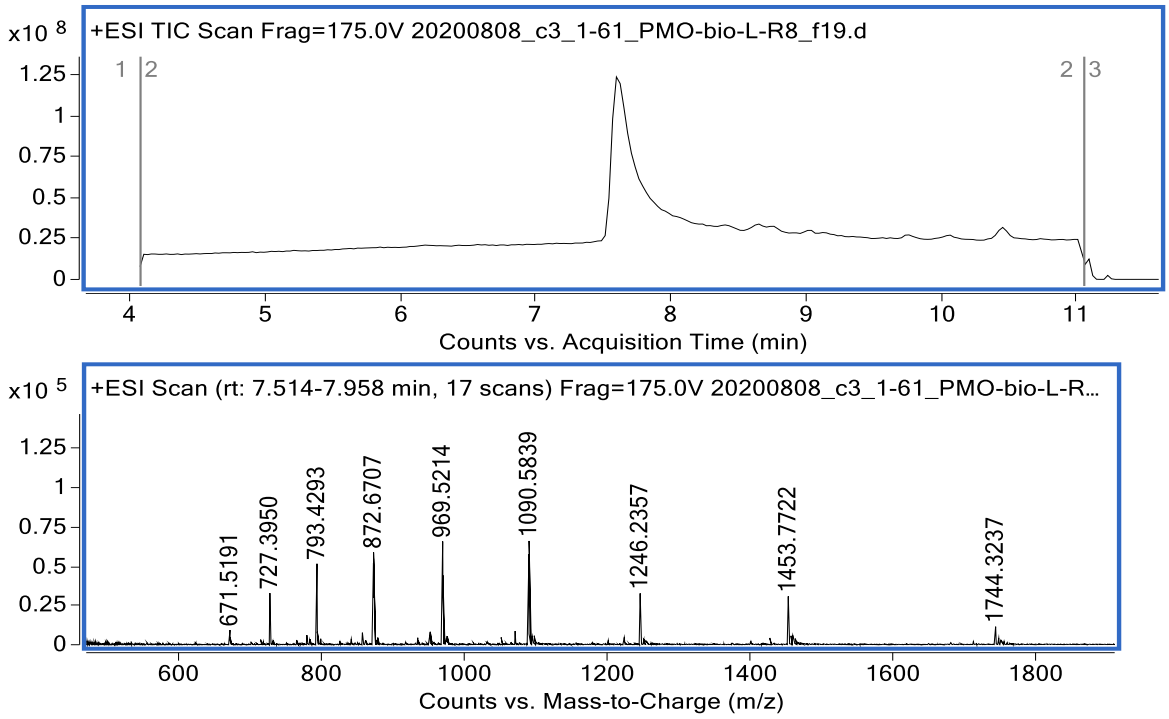


Peptide: PMO-Bio-D-R8
Sequence: Biotin-Lys(PMO)-GGKGGWRRRRRRRRR
Calculated mass: 8717.5 Da
Observed mass: 8717.7 Da





Peptide: PMO-Bio-L-R8
Sequence: Biotin-Lys(PMO)-GGKGGWRRRRRRRR
Calculated mass: 8717.5 Da
Observed mass: 8717.6 Da



4.9 References

(1) Wolfe Justin M.; Fadzen Colin M.; Holden Rebecca L.; Yao Monica; Hanson Gunnar J.; Pentelute Bradley L. Perfluoroaryl Bicyclic Cell-Penetrating Peptides for Delivery of Antisense Oligonucleotides. *Angewandte Chemie International Edition* **2018**, *0* (0). https://doi.org/10.1002/anie.201801167.

(2) Oehlke, J.; Scheller, A.; Wiesner, B.; Krause, E.; Beyermann, M.; Klauschenz, E.; Melzig, M.; Bienert, M. Cellular Uptake of an α-Helical Amphipathic Model Peptide with the Potential to Deliver Polar Compounds into the Cell Interior Non-Endocytically. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1998**, *1414* (1), 127–139. https://doi.org/10.1016/S0005-2736(98)00161-8.

(3) Margus, H.; Padari, K.; Pooga, M. Cell-Penetrating Peptides as Versatile Vehicles for Oligonucleotide Delivery. *Molecular Therapy: The Journal Of The American Society Of Gene Therapy* **2012**, *20* (3), 525–533. https://doi.org/10.1038/mt.2011.284.

(4) Reissmann Siegmund. Cell Penetration: Scope and Limitations by the Application of Cell-penetrating Peptides. *Journal of Peptide Science* **2014**, *20* (10), 760–784. https://doi.org/10.1002/psc.2672.

(5) Fischer, R.; Waizenegger, T.; Köhler, K.; Brock, R. A Quantitative Validation of Fluorophore-Labelled Cell-Permeable Peptide Conjugates: Fluorophore and Cargo Dependence of Import. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2002**, *1564* (2), 365–374. https://doi.org/10.1016/S0005-2736(02)00471-6.

(6) Lindgren, M. E.; Hällbrink, M. M.; Elmquist, A. M.; Langel, U. Passage of Cell-Penetrating Peptides across a Human Epithelial Cell Layer in Vitro. *Biochem. J.* **2004**, *377* (Pt 1), 69–76. https://doi.org/10.1042/BJ20030760.

(7) Illien, F.; Rodriguez, N.; Amoura, M.; Joliot, A.; Pallerla, M.; Cribier, S.; Burlina, F.; Sagan, S. Quantitative Fluorescence Spectroscopy and Flow Cytometry Analyses of Cell-Penetrating Peptides Internalization Pathways: Optimization, Pitfalls, Comparison with Mass Spectrometry Quantification. *Scientific Reports* **2016**, *6*, 36938. https://doi.org/10.1038/srep36938.

(8) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS Cent Sci* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.

(9) Baker, D. E. Eteplirsen. *Hosp Pharm* **2017**, *52* (4), 302–305. https://doi.org/10.1310/hpj5204-302.

(10) Lim, K. R. Q.; Maruyama, R.; Yokota, T. Eteplirsen in the Treatment of Duchenne Muscular Dystrophy. *Drug Des Devel Ther* **2017**, *11*, 533–545. https://doi.org/10.2147/DDDT.S97635.

(11) Inc, S. T. Sarepta Therapeutics Announces Positive Clinical Results from MOMENTUM, a Phase 2 Clinical Trial of SRP-5051 in Patients with Duchenne Muscular Dystrophy Amenable to Skipping Exon 51 http://www.globenewswire.com/news-release/2020/12/07/2140613/0/en/Sarepta-Therapeutics-Announces-Positive-Clinical-Results-from-MOMENTUM-a-Phase-2-Clinical-Trial-of-SRP-5051-in-Patients-with-Duchenne-Muscular-Dystrophy-Amenable-to-Skipping-Exon-5.html (accessed 2020 -12 -07).

(12) Kurrikoff, K.; Vunk, B.; Langel, Ü. Status Update in the Use of Cell-Penetrating Peptides for the Delivery of Macromolecular Therapeutics. *Expert Opinion on Biological Therapy* **2021**, *21* (3), 361–370. https://doi.org/10.1080/14712598.2021.1823368.
(13) Dintzis, H. M.; Symer, D. E.; Dintzis, R. Z.; Zawadzke, L. E.; Berg, J. M. A Comparison of the Immunogenicity of a Pair of Enantiomeric Proteins. *Proteins: Structure, Function, and Bioinformatics* **1993**, *16* (3), 306–308. https://doi.org/10.1002/prot.340160309.
(14) Kroenke, M. A.; Weeraratne, D. K.; Deng, H.; Sloey, B.; Subramanian, R.; Wu, B.; Serenko, M.; Hock, M. B. Clinical Immunogenicity of the D-Amino Acid Peptide Therapeutic Etelcalcetide: Method Development Challenges and Anti-Drug Antibody Clinical Impact Assessments. *Journal of Immunological Methods* **2017**, *445*, 37–44. https://doi.org/10.1016/j.jim.2017.03.005.
(15) Najjar, K.; Erazo-Oliveras, A.; Brock, D. J.; Wang, T.-Y.; Pellois, J.-P. An L- to d-Amino Acid Conversion in an Endosomolytic Analog of the Cell-Penetrating Peptide TAT Influences Proteolytic Stability, Endocytic Uptake, and Endosomal Escape. *J Biol Chem* **2017**, *292* (3), 847–861. https://doi.org/10.1074/jbc.M116.759837.
(16) Ma, Y.; Gong, C.; Ma, Y.; Fan, F.; Luo, M.; Yang, F.; Zhang, Y.-H. Direct Cytosolic Delivery of Cargoes in Vivo by a Chimera Consisting of D- and L-Arginine Residues. *Journal of Controlled Release* **2012**, *162* (2), 286–294. https://doi.org/10.1016/j.jconrel.2012.07.022.
(17) Henriques, S. T.; Peacock, H.; Benfield, A. H.; Wang, C. K.; Craik, D. J. Is the Mirror Image a True Reflection? Intrinsic Membrane Chirality Modulates Peptide Binding. *J. Am. Chem. Soc.* **2019**, *141* (51), 20460–20469. https://doi.org/10.1021/jacs.9b11194.
(18) Pujals, S.; Sabidó, E.; Tarragó, T.; Giralt, E. All-D Proline-Rich Cell-Penetrating Peptides: A Preliminary in Vivo Internalization Study. *Biochemical Society Transactions* **2007**, *35* (4), 794–796. https://doi.org/10.1042/BST0350794.
(19) Ito, S.; Torii, Y.; Chikamatsu, S.; Harada, T.; Yamaguchi, S.; Ogata, S.; Sonoda, K.; Wakayama, T.; Masuda, T.; Ohtsuki, S. Oral Coadministration of Zn-Insulin with d-Form Small Intestine-Permeable Cyclic Peptide Enhances Its Blood Glucose-Lowering Effect in Mice. *Mol. Pharmaceutics* **2021**, *18* (4), 1593–1603. https://doi.org/10.1021/acs.molpharmaceut.0c01010.
(20) Verdurmen, W. P. R.; Bovee-Geurts, P. H.; Wadhwani, P.; Ulrich, A. S.; Hällbrink, M.; van Kuppevelt, T. H.; Brock, R. Preferential Uptake of L- versus D-Amino Acid Cell-Penetrating Peptides in a Cell Type-Dependent Manner. *Chemistry & Biology* **2011**, *18* (8), 1000–1010. https://doi.org/10.1016/j.chembiol.2011.06.006.
(21) Knox, S. L.; Wissner, R.; Piszkiewicz, S.; Schepartz, A. Cytosolic Delivery of Argininosuccinate Synthetase Using a Cell-Permeant Miniature Protein. *ACS Cent. Sci.* **2021**, *7* (4), 641–649. https://doi.org/10.1021/acscentsci.0c01603.
(22) Fadzen, C. M.; Holden, R. L.; Wolfe, J. M.; Choo, Z.-N.; Schissel, C. K.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. *Biochemistry* **2019**. https://doi.org/10.1021/acs.biochem.9b00413.
(23) Schissel, C. K.; Mohapatra, S.; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep Learning to Design Nuclear-Targeting Abiotic Miniproteins. *Nat. Chem.* **2021**, 1–9. https://doi.org/10.1038/s41557-021-00766-3.

(24) Youngblood, D. S.; Hatlevig, S. A.; Hassinger, J. N.; Iversen, P. L.; Moulton, H. M. Stability of Cell-Penetrating Peptide−Morpholino Oligomer Conjugates in Human Serum and in Cells. *Bioconjugate Chem.* **2007**, *18* (1), 50–60. https://doi.org/10.1021/bc060138s.
(25) Sazani, P.; Gemignani, F.; Kang, S.-H.; Maier, M. A.; Manoharan, M.; Persmark, M.; Bortner, D.; Kole, R. Systemically Delivered Antisense Oligomers Upregulate Gene Expression in Mouse Tissues. *Nature Biotechnology* **2002**, *20* (12), 1228–1233. https://doi.org/10.1038/nbt759.
(26) Erazo-Oliveras, A.; Muthukrishnan, N.; Baker, R.; Wang, T.-Y.; Pellois, J.-P. Improving the Endosomal Escape of Cell-Penetrating Peptides and Their Cargos: Strategies and Challenges. *Pharmaceuticals (Basel)* **2012**, *5* (11), 1177–1209. https://doi.org/10.3390/ph5111177.
(27) Holden, P.; Horton, W. A. Crude Subcellular Fractionation of Cultured Mammalian Cell Lines. *BMC Research Notes* **2009**, *2* (1), 243. https://doi.org/10.1186/1756-0500-2-243.
(28) Peraro, L.; Deprey, K. L.; Moser, M. K.; Zou, Z.; Ball, H. L.; Levine, B.; Kritzer, J. A. Cell Penetration Profiling Using the Chloroalkane Penetration Assay. *J. Am. Chem. Soc.* **2018**, *140* (36), 11360–11369. https://doi.org/10.1021/jacs.8b06144.
(29) Schmidt, S.; Adjobo-Hermans, M. J. W.; Wallbrecher, R.; Verdurmen, W. P. R.; Bovée-Geurts, P. H. M.; van Oostrum, J.; Milletti, F.; Enderle, T.; Brock, R. Detecting Cytosolic Peptide Delivery with the GFP Complementation Assay in the Low Micromolar Range. *Angew. Chem. Int. Ed.* **2015**, *54* (50), 15105–15108. https://doi.org/10.1002/anie.201505913.
(30) Peier, A.; Ge, L.; Boyer, N.; Frost, J.; Duggal, R.; Biswas, K.; Edmondson, S.; Hermes, J. D.; Yan, L.; Zimprich, C.; Sadruddin, A.; Kristal Kaan, H. Y.; Chandramohan, A.; Brown, C. J.; Thean, D.; Lee, X. E.; Yuen, T. Y.; Ferrer-Gago, F. J.; Johannes, C. W.; Lane, D. P.; Sherborne, B.; Corona, C.; Robers, M. B.; Sawyer, T. K.; Partridge, A. W. NanoClick: A High Throughput, Target-Agnostic Peptide Cell Permeability Assay. *ACS Chem. Biol.* **2021**, *16* (2), 293–309. https://doi.org/10.1021/acschembio.0c00804.
(31) Teo, S. L. Y.; Rennick, J. J.; Yuen, D.; Al-Wassiti, H.; Johnston, A. P. R.; Pouton, C. W. Unravelling Cytosolic Delivery of Cell Penetrating Peptides with a Quantitative Endosomal Escape Assay. *Nat Commun* **2021**, *12* (1), 3721. https://doi.org/10.1038/s41467-021-23997-x.
(32) Wissner, R. F.; Steinauer, A.; Knox, S. L.; Thompson, A. D.; Schepartz, A. Fluorescence Correlation Spectroscopy Reveals Efficient Cytosolic Delivery of Protein Cargo by Cell-Permeant Miniature Proteins. *ACS Cent. Sci.* **2018**, *4* (10), 1379–1393. https://doi.org/10.1021/acscentsci.8b00446.
(33) Duncan, M. W. Practical Quantitative Biomedical Applications of MALDI-TOF Mass Spectrometry.
(34) Burlina, F.; Sagan, S.; Bolbach, G.; Chassaing, G. Quantification of the Cellular Uptake of Cell-Penetrating Peptides by MALDI-TOF Mass Spectrometry. *Angewandte Chemie International Edition* **2005**, *44* (27), 4244–4247. https://doi.org/10.1002/anie.200500477.
(35) Burlina, F.; Sagan, S.; Bolbach, G.; Chassaing, G. A Direct Approach to Quantification of the Cellular Uptake of Cell-Penetrating Peptides Using MALDI-TOF Mass Spectrometry. *Nature protocols* **2006**, *1* (1), 200.
(36) Aubry, S.; Aussedat, B.; Delaroche, D.; Jiao, C.-Y.; Bolbach, G.; Lavielle, S.; Chassaing, G.; Sagan, S.; Burlina, F. MALDI-TOF Mass Spectrometry: A Powerful Tool to Study the Internalization of Cell-Penetrating Peptides. *Biochimica et Biophysica Acta (BBA) -*

Biomembranes **2010**, *1798* (12), 2182–2189. https://doi.org/10.1016/j.bbamem.2009.11.011.
(37) Ho, H.-P.; Rathod, P.; Louis, M.; Tada, C. K.; Rahaman, S.; Mark, K. J.; Leng, J.; Dana, D.; Kumar, S.; Lichterfeld, M.; Chang, E. J. Studies on Quantitative Phosphopeptide Analysis by MALDI Mass Spectrometry Without Label, Chromatography or Calibration Curves. *Rapid Commun Mass Spectrom* **2014**, *28* (24), 2681–2689. https://doi.org/10.1002/rcm.7063.
(38) Aussedat, B.; Sagan, S.; Chassaing, G.; Bolbach, G.; Burlina, F. Quantification of the Efficiency of Cargo Delivery by Peptidic and Pseudo-Peptidic Trojan Carriers Using MALDI-TOF Mass Spectrometry. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2006**, *1758* (3), 375–383. https://doi.org/10.1016/j.bbamem.2006.01.012.
(39) Bode, S. A.; Thévenin, M.; Bechara, C.; Sagan, S.; Bregant, S.; Lavielle, S.; Chassaing, G.; Burlina, F. Self-Assembling Mini Cell-Penetrating Peptides Enter by Both Direct Translocation and Glycosaminoglycan-Dependent Endocytosis. *Chem. Commun.* **2012**, *48* (57), 7179–7181. https://doi.org/10.1039/C2CC33240J.
(40) Aussedat, B.; Dupont, E.; Sagan, S.; Joliot, A.; Lavielle, S.; Chassaing, G.; Burlina, F. Modifications in the Chemical Structure of Trojan Carriers: Impact on Cargo Delivery. *Chem. Commun.* **2008**, No. 12, 1398–1400. https://doi.org/10.1039/B800433A.
(41) Brock, R. The Uptake of Arginine-Rich Cell-Penetrating Peptides: Putting the Puzzle Together. *Bioconjugate Chem.* **2014**, *25* (5), 863–868. https://doi.org/10.1021/bc500017t.
(42) Fuchs, S. M.; Raines, R. T. Polyarginine as a Multifunctional Fusion Tag. *Protein Science* **2005**, *14* (6), 1538–1544. https://doi.org/10.1110/ps.051393805.
(43) Richard, J. P.; Melikov, K.; Vives, E.; Ramos, C.; Verbeure, B.; Gait, M. J.; Chernomordik, L. V.; Lebleu, B. Cell-Penetrating Peptides. A Reevaluation of the Mechanism of Cellular Uptake. *J. Biol. Chem.* **2003**, *278* (1), 585–590. https://doi.org/10.1074/jbc.M209548200.
(44) Maiolo, J. R.; Ferrer, M.; Ottinger, E. A. Effects of Cargo Molecules on the Cellular Uptake of Arginine-Rich Cell-Penetrating Peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2005**, *1712* (2), 161–172. https://doi.org/10.1016/j.bbamem.2005.04.010.
(45) Patel, L. N.; Zaro, J. L.; Shen, W.-C. Cell Penetrating Peptides: Intracellular Pathways and Pharmaceutical Perspectives. *Pharmaceutical Research* **2007**, *24* (11), 1977–1992. https://doi.org/10.1007/s11095-007-9303-7.
(46) Boisguérin, P.; Deshayes, S.; Gait, M. J.; O'Donovan, L.; Godfrey, C.; Betts, C. A.; Wood, M. J. A.; Lebleu, B. Delivery of Therapeutic Oligonucleotides with Cell Penetrating Peptides. *Advanced Drug Delivery Reviews* **2015**, *87*, 52–67. https://doi.org/10.1016/j.addr.2015.02.008.
(47) Böttger, R.; Hoffmann, R.; Knappe, D. Differential Stability of Therapeutic Peptides with Different Proteolytic Cleavage Sites in Blood, Plasma and Serum. *PLOS ONE* **2017**, *12* (6), e0178943. https://doi.org/10.1371/journal.pone.0178943.
(48) Werner, H. M.; Cabalteja, C. C.; Horne, W. S. Peptide Backbone Composition and Protease Susceptibility: Impact of Modification Type, Position, and Tandem Substitution. *Chembiochem* **2016**, *17* (8), 712–718. https://doi.org/10.1002/cbic.201500312.
(49) Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Liu, S.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by

Automated Flow Chemistry. *Science* **2020**, *368* (6494), 980–987. https://doi.org/10.1126/science.abb2491.
(50) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014**, *15* (5), 713–720. https://doi.org/10.1002/cbic.201300796.
(51) Kjeken, R.; Mousavi, S. A.; Brech, A.; Griffiths, G.; Berg, T. Wortmannin-Sensitive Trafficking Steps in the Endocytic Pathway in Rat Liver Endothelial Cells. *Biochem J* **2001**, *357* (Pt 2), 497–503.
(52) Dutta, D.; Donaldson, J. G. Search for Inhibitors of Endocytosis: Intended Specificity and Unintended Consequences. *Cellular Logistics* **2012**, *2* (4), 203–208. https://doi.org/10.4161/cl.23967.

5 Chapter 5: In-cell Penetration Selection—Mass Spectrometry Produces Noncanonical Peptides for Antisense Delivery

The work presented in this chapter has been reproduced from the following manuscript:

Schissel, C.K.*; Farquhar, C. E.*; Loas, A.; Malmberg, A.B.; Pentelute, B.L. In-cell penetration selection—mass spectrometry produces noncanonical peptides for antisense delivery. *(Manuscript in preparation)*

**These authors contributed equally.*

5.1 Introduction

After 30 years of investigation, therapies involving cell-penetrating peptides (CPPs) are beginning to advance to late-stage clinical trials.[1,2] These sequences, composed typically of fewer than 20 amino acids and endowed with diverse physicochemical properties, are able to penetrate the cellular membrane and at times deliver otherwise non-penetrant cargo.[3] Because of these properties, CPPs have potential applications for the treatment of disease, including cancer, genetic disorders, inflammation, and diabetes. SRP-5051, a peptide covalently attached to an antisense phosphorodiamidate morpholino oligomer (PMO) cargo, demonstrated higher tissue exposure and activity in patients taking a monthly dose compared to patients taking weekly doses of the antisense oligomer alone.[4] Despite these recent advances, to our knowledge no CPP-based therapy has reached the commercial market yet.[5]

While there are several limitations that have slowed the clinical advancement of CPPs, one that we are particularly interested in addressing is the empirical design of novel, more efficient sequences. Historically, CPPs, also known as protein transduction domains (PTDs), were derived from transmembrane portions of viral and transcriptional proteins. For example, the polyarginine peptide TAT was derived from the HIV-trans-activator of transcription protein and was found to penetrate into the nucleus and target gene expression.[6,7] From this and similar sequences, synthetic peptides could be designed, including some tailored for delivery of PMO cargo such as Bpep, which relies on arginine to trigger uptake and the unnatural residues β-alanine and 6-amino-hexanoic acid to trigger endosomal escape.[8] Beyond empirical design using derivatives of polyarginine sequences, the rational design of new sequences remains challenging. Methods involving some rational design include synthetic molecular evolution[9,10] and in silico methods.[11–14] The latter include our own recent work that leverages machine learning to design new sequences using a model trained with a combinatorial library tested for the desired activity: nuclear localization.[15–17] Finally, another common strategy involves screening platforms employing libraries from phage or mRNA display.[18,19] For example, a screening platform identified several "phylomer" CPPs from bacterial and viral genomes that were then shown to deliver antisense cargo in vivo.[20] Still, a persistent limitation with these approaches is the difficulty of incorporating D-chiral or unnatural amino acids, which would provide access to an augmented chemical space. Unnatural amino acids are more easily incorporated into synthetic one-bead one-compound (OBOC) libraries, although discovery of CPPs by these methods often relies on synthetic

vesicles.[21] Improved screening platforms for discovery of enhanced CPPs could be developed by using unnatural peptides to access greater chemical diversity and proteolytic stability, and by incorporating biologically relevant screening conditions into the protocol, such as in-cell selection and inclusion of the specific cargo to be delivered.

Classic affinity selection involves screening peptide ligands from synthetic libraries (OBOC), phage or mRNA display against immobilized protein targets, and decoding hits.[22,23] These methods advanced to biologically relevant conditions in on-cell selection platforms for the discovery of new ligands with affinity for the external surface of cells and tissues.[24–26] Again, biological display techniques are restricted to the use of mostly natural amino acids, limiting the resulting library diversity and proteolytic stability[27,28], and even those mirror image techniques that allow D-peptide discovery still have difficulty incorporating non-canonical residues.[29,30] Screening of a synthetic one-bead one-compound (OBOC) library eases the incorporation of non-canonical and D-residues. Our group has recently demonstrated that in vivo affinity selection-mass spectrometry (AS-MS) could identify an erythrocyte-targeting D-peptide.[31] Such label-free techniques applied to the cell surface allow for the discovery of novel, non-canonical, D-peptide binders without the addition of display scaffolds or encoding tags.

While most works have focused on affinity screening at the cell surface, there has been some success pushing these techniques to discover peptides that cross the cell membrane. As mentioned, phage display and encoded peptide libraries have been used to discover novel cell penetrating peptides, but these methods have limited advancements for discovery of peptides that deliver cargo to subcellular compartments.[19,20] Recently, the first example of a DNA-encoded small molecule library screen inside living cells resulted in several chemical motifs that bind the over-expressed protein targets inside oocytes.[32] This strategy of in-cell selection (rather than on-cell) would be beneficial for discovery of peptides that can deliver macromolecules to the cytosol.

Here we have combined CPP library design and AS-MS selection approaches into a new method: in-cell penetration selection-mass spectrometry (in-cell PS-MS, Fig. 5.1). Bringing together our expertise in both MS-based selection methodologies and in-cell localization, this technique enables direct recovery of "hit" peptides that deliver a specific type of antisense cargo into the cytosol of cells. Our PS-MS methodology allows the detection of non-canonical peptide-cargo conjugates in-cell, with only the addition of a small biotin handle for extraction. In addition, the PS-MS platform allows additional spatial resolution, separating peptides extracted from whole

cell lysates from those extracted from the cytosolic fraction. Resulting peptides are tested in a validation assay selects for peptides that can effectively deliver the cargo to the nucleus, giving an additional layer of spatial resolution. The PMO-CPP library used in the screening demonstrated antisense delivery activity, not necessarily due to a few highly active sequences but due to the activity of the library as a whole. This method led to the discovery of a potent antisense delivery peptide, Pep1a, isolated from the cytosol of cells. This peptide was more active and more efficiently localized to the nucleus compared to peptides that were isolated from the whole cell extracts, which include endosomes.

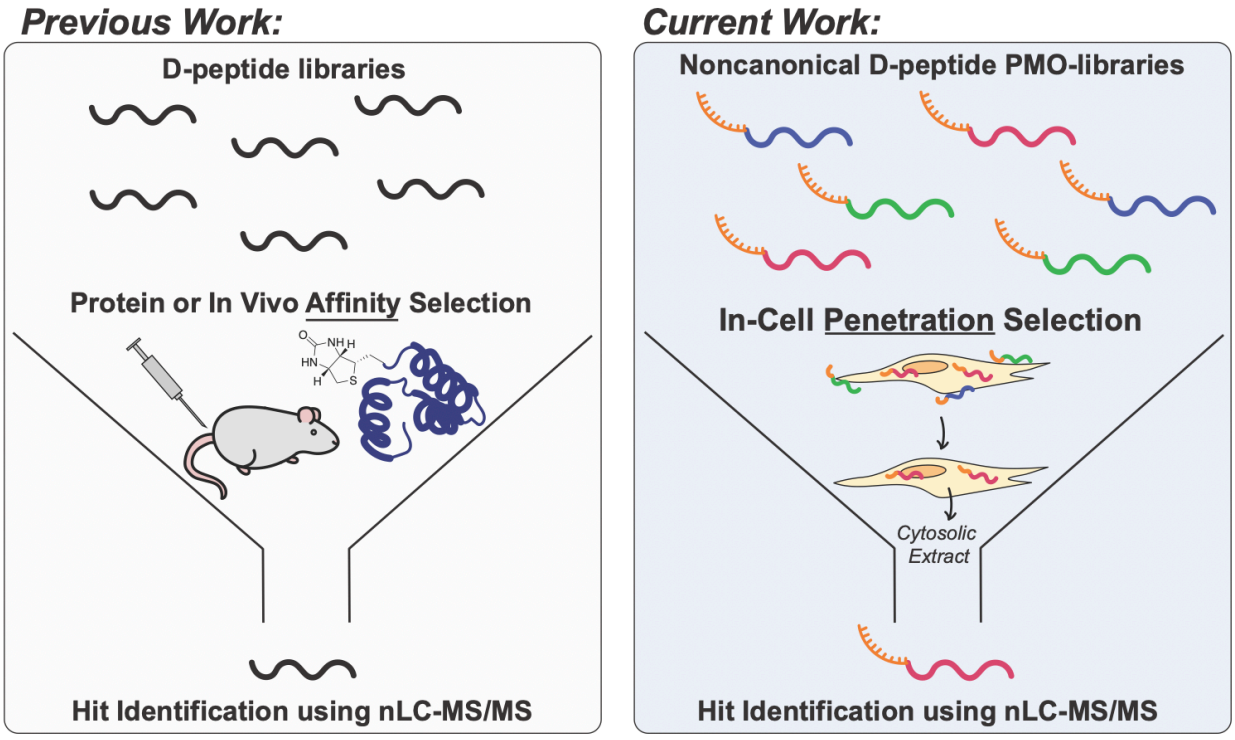


Figure 5.1 In-cell penetration selection—mass spectrometry identifies noncanonical peptides that access the cytosol. Building on prior methods for identifying binders to proteins and cells (left), in-cell PS-MS identifies noncanonical peptides that carry macromolecular cargo into the cytosol of cells.

5.2 Results

5.2.1 Library preparation

The library was prepared with a "CPP-like" C-terminal sequence and six variable positions containing D- and unnatural amino acids (Fig. 5.2A). Split-and-pool synthesis afforded 0.016 μg of peptide per bead for a low-redundancy, 95,000-member library with a theoretical diversity greater than 10^8. A KWKK motif, derived from the established cell-penetrating peptide penetratin[33], was installed at the C-terminus to give the library a boost in activity. We have previously shown that these fixed constraints and C-terminal charge also increase peptide recovery in AS-MS.[22,34] Unnatural amino acids were chosen to expand the chemical diversity and potentially enhance cell penetration of the library peptides. The library includes unnatural residues with non-α backbones to promote endosomal escape (γ-aminobutyric acid and β-alanine),[8] residues with hydrophobic and aromatic functionality to increase membrane penetration (homoleucine, norleucine, naphthylalanine, and diphenylalanine),[35,36] and additional charged residues and arginine analogues to enhance membrane penetration (diaminobutyric acid, aminopiperidine-carboxylic acid, aminomethylphenylalanine, and 2-amino-4-guanidinobutanoic acid)[37,38] (Fig. 5.2B). The oxidative cleavable linker isoseramox was installed by reductive amination immediately following the variable region as previously reported,[39] followed by a trypsin cleavage site to prevent the recovery of non-internalized peptides. Finally, azidolysine and biotin capped the N-terminus of the sequences to allow for PMO conjugation and affinity capture, respectively. Following cleavage from the resin, a portion of the library was conjugated by azide-alkyne cycloaddition to a model PMO derivatized with dibenzocyclooctyne (DBCO), monitored by LC-MS. Quality control analysis of the library by Orbitrap nano-liquid chromatography-tandem mass spectrometry (nLC-MS/MS) confirmed successful synthesis and exhibited a range of incorporated residues (Fig. 5.2C, Appendix III). This library design ensured the isolation of 10-mer peptides with a native N-terminus, suitable for sequencing via tandem mass spectrometry.[39]

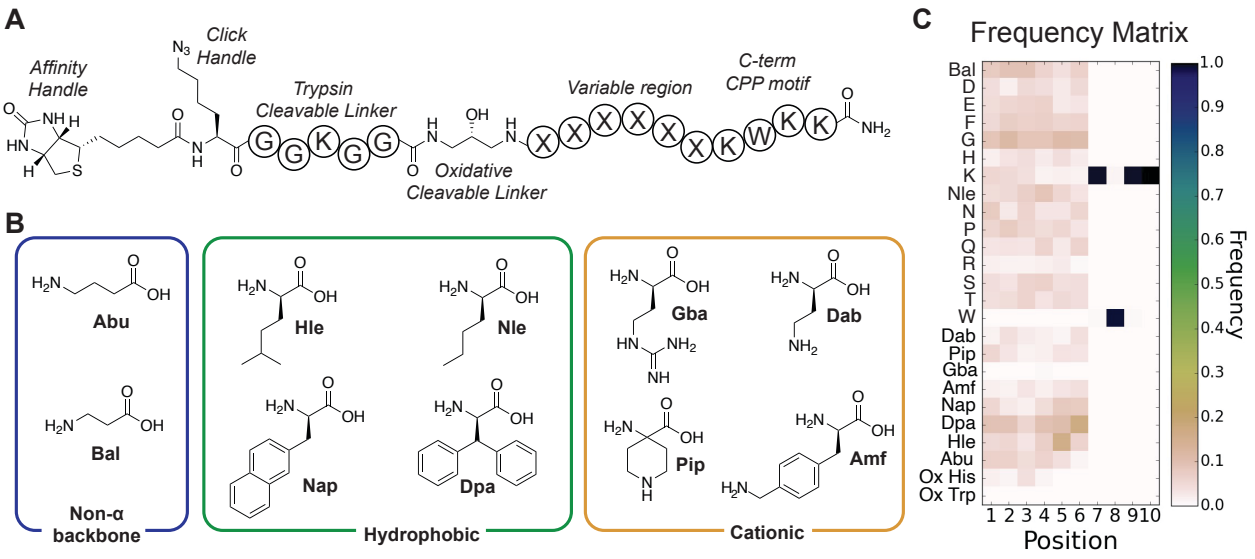


Figure 5.2 A combinatorial library was prepared with unnatural and D-amino acids. (A) Design of the library. An N-terminal biotin and azidolysine provide an affinity handle and conjugation handle, respectively. A trypsin-cleavable linker prevents isolation of extracellular conjugates, and isoseramox cleavable linker permits oxidative cleavage of the conjugates from streptavidin beads. Finally, there are six variable positions within the library peptides, with a "CPP-like" motif capping the C-terminus. (B) Structures of the unnatural monomers used. All natural-backbone monomers were in D-form. (C) Heat map of the quality control showing relative abundance of the various amino acids of the sequence up to the isoseramox linker. Positions 7-10 show the KWKK motif, with positions 1-6 showing the varied composition of the variable region. Abu (γ-aminobutyric acid), Bal (β-alanine), Hle (homoleucine), Nle (norleucine), Nap (naphthylalanine), Dpa (diphenylalanine), Dab (diaminobutyric acid), Pip (aminopiperidine-carboxylic acid), Amf (aminomethylphenylalanine), and Gba (2-amino-4-guanidinobutanoic acid).

We then performed a series of activity experiments to confirm that the peptides within the library had nuclear-localizing activity. The phenotypic assay used correlates with the amount of active PMO delivered to the nucleus by resulting in corrective splicing to produce enhanced green fluorescent protein (EGFP), quantified by flow cytometry. First, PMO-library aliquots demonstrated a concentration-dependent increase in activity (Fig. 5.3A). At the same time, the library at these concentrations did not exhibit any membrane disruption or toxicity as determined by a lactate dehydrogenase (LDH) release assay (Fig. 5.3B). We also found that testing the same concentrations of library aliquots containing different amounts of sequences, thereby increasing diversity, showed no difference in activity (Fig. 5.3C).

The PMO delivery activity of the library is likely energy-dependent, similar to PMO-CPP conjugates previously investigated.[15,16,40] A 1,000-member portion of the PMO-library was incubated with cells at 4 ˚C, conditions that arrest energy-dependent uptake. After incubation with the PMO-CPP conjugates, each well was washed extensively with PBS and heparin in order to disrupt and remove membrane-bound conjugates.[41] The cells were warmed back up to 37 ˚C and the assay continued in standard format and analyzed by flow cytometry. The significant decrease in library PMO delivery (relative to PMO alone) at 4 ˚C for both 5 μM and 20 μM library incubation conditions demonstrates energy-dependent uptake for the PMO-CPPs (Fig. 5.3D). Previous PMO-CPPs discovered in our laboratory entered cells via clathrin-mediated endocytosis, as demonstrated through incubation with a panel of chemical endocytosis inhibitors, suggesting this could be a likely mechanism of energy-dependent uptake for a PMO-CPP library. [15,16,40]

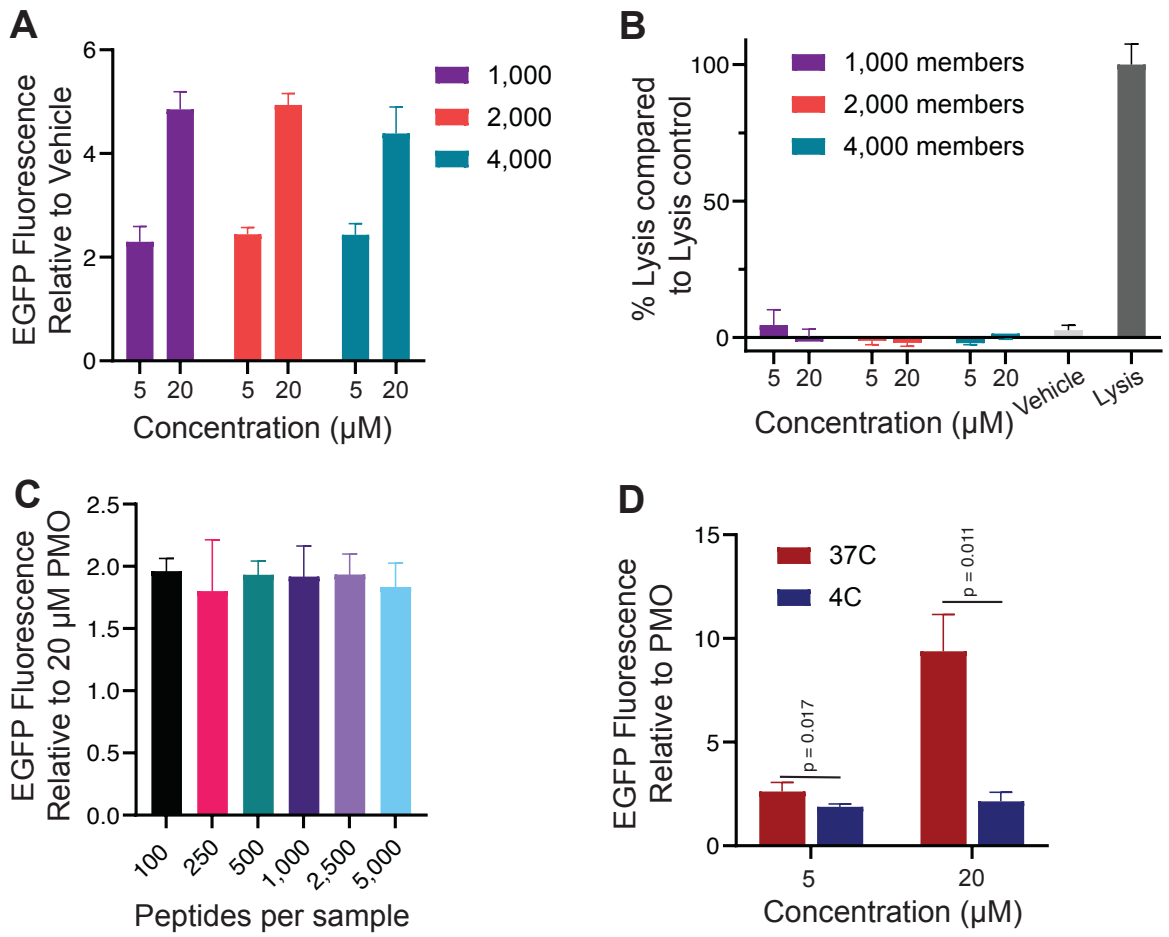


Figure 5.3 The CPP library can deliver PMO regardless of member size and enters via active transport. (A) HeLa 654 cells were treated with 5 or 20 μM of PMO-Library containing ~1000, ~2000, or ~4000 members for 22 h prior to flow cytometry. Results are given as the mean EGFP fluorescence of cells treated with PMO-peptide relative to the fluorescence of cells alone. PMO-Library samples show concentration-dependent PMO delivery at all library sizes tested. (B) HeLa 654 cells were treated with 5 or 20 μM PMO-Library for 22 h, then tested for LDH released into the cell media. Results are given as LDH release above vehicle relative to fully lysed cells. No compounds showed LDH release significantly above vehicle-treated cells. (C) HeLa 654 cells were treated with 20 μM PMO-Library of varying member sizes or 20 μM PMO alone for 22 h prior to flow cytometry. Results are given relative to the fluorescence of PMO-treated cells. There is no significant difference in EGFP fluorescence between the libraries of different sizes. (D) HeLa 654 cells were pre-incubated at 4 ˚C or 37 ˚C for 30 min prior to treatment with 20 μM PMO-Library or 20 μM PMO alone for 2 h at the indicated temperature. After treatment, cells were washed with 0.1 mg/mL heparin and incubated in media for 22 h prior to flow cytometry. Results are given relative to the fluorescence of PMO-treated cells. There was a significant difference between the 4 ˚C and 37 ˚C treatment conditions at 5 μM ($p = 0.017$) and 20 μM ($p = 0.011$) of PMO-library.

5.2.2 In-cell penetration selection-mass spectrometry

We subjected the library to the in-cell penetration selection-mass spectrometry platform (in-cell PS-MS) to discover sequences that are present in the whole cell lysate and in the cytosol (Fig. 5.4). Here we are profiling for sequences that access the cytosol, making them more likely to access the nucleus and the RNA target of the PMO cargo. The protocol for extracting biotinylated sequences was adapted from our recent method of profiling mixtures of PMO-D-CPPs from the cytosol and whole cell using MALDI-ToF.[42,43] Confluent HeLa cells in a 12-well plate were treated with 20 μM of biotin-library or PMO-biotin-library (~10^3 members, 3.5 nmol individual peptide per bead) for 1 h, before being washed with PBS and heparin to dissociate membrane-bound conjugates. Cells were then lifted and extracellular conjugates digested with Trypsin, pelleted, and washed with PBS. Cells were gently lysed using either RIPA buffer (for whole cell extraction) or digitonin buffer (for cytosolic extraction).[42,44] Exclusion of endosomes in the cytosolic fraction was confirmed by Western blot, in which the endosomal marker, Rab5, is absent (Fig. 5.5). In addition to the experimental samples, half of the no-treatment lysates were spiked with library as positive controls (Appendix IV).

Biotinylated species in the lysates were affinity captured with magnetic streptavidin beads, and ultimately released by oxidative cleavage using brief incubation with sodium periodate. We had previously used this cleavable linker to recover a single PMO-CPP conjugate from inside cells, and it was found to reliably cleave library peptides from streptavidin beads to isolate native peptides for sequencing by mass spectrometry.[39] The isolated peptides were desalted by solid-phase extraction and analyzed via Orbitrap tandem mass spectrometry using a mixed fragmentation method optimized for cationic peptides, consisting of electron-transfer dissociation (ETD), higher-energy ETD, and higher-energy collisional dissociation (HCD). Sequences matching the library design were then identified using a Python script.[22,34]

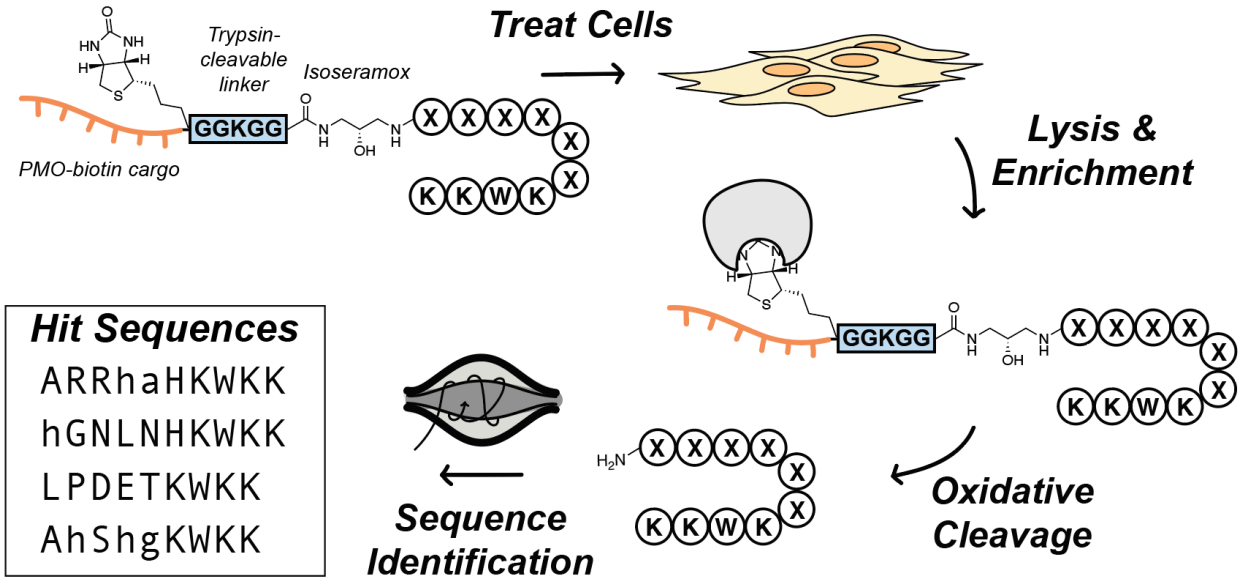


Figure 5.4 Workflow of in-cell penetration selection-mass spectrometry. HeLa cells were treated with 20 µM PMO-biotin-library or biotin-library (1,000 members) for 1 h at 37 ˚C. Cells were then extensively washed with PBS and 0.1 mg/mL heparin before lysis with RIPA (whole cell extract) or digitonin (cytosolic extract). Lysates were incubated with magnetic streptavidin beads, and the C-terminal native peptides were cleaved from the beads under oxidative conditions. The peptides were desalted through solid-phase extraction and sequenced by nLC-MS/MS. Hit PMO-delivering sequences were then identified as those peptides found only in the PMO-library fractions that do not overlap with peptides found in the cell only control or the samples treated with the biotin-library.

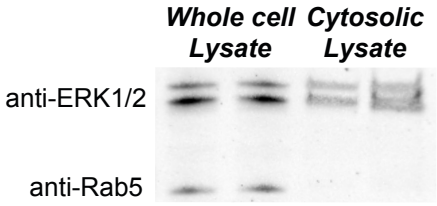


Figure 5.5 Extraction of the cytosol was verified via Western blot. The protein in the no treatment control lysates were analyzed via sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot to visualize presence of ERK1/2 (cytosolic marker) and Rab5 (an endosomal marker). Rab5 is observed in the whole cell lysate but not in the cytosolic extract.

5.2.3 Hit peptides identified from PS-MS show nuclear PMO delivery, but are not solely responsible PMO delivery activity of library

Several hit peptides were selected for experimental validation and showed differential activities depending on the fraction in which they were found. We selected two sequences found in the cytosolic extract (Pep1a, Pep1b) and two from the whole cell extract (Pep1c, Pep1d), with sequences shown in Fig. 5.6. These peptides were synthesized via semi-automated solid-phase fast-flow peptide synthesis[45] with identical sequences to the library design with the exception of a D-Ser residue to replace the isoseramox linker. These sequences were tested first in a concentration-response EGFP assay. The sequences extracted from the cytosol showed significantly increased activity compared to the sequences from the whole cell lysate, with Pep1a showing an EC50 of 43 μM compared to Pep1c with EC50 of 380 μM (Fig. 5.7A). It was also confirmed that these sequences did not exhibit membrane toxicity at the concentrations tested (Fig. 5.7B). The peptides showed a positive correlation between charge and activity, with the highest performing peptide (Pep1a) having a charge of +7, compared to Pep1c with a charge of +2. This trend of positive charge correlating with PMO activity has been observed consistently in our lab.[15–17,40] Although the hit peptides do not show higher activity than the parent peptide penetratin, they do show significantly less toxicity and membrane disruption, demonstrating their potential utility as PMO delivery vehicles. We also compared Pep1a and Pep1c to the known endosomal escape peptide Bpep[42], composed of eight Arg residues interspaced with non-α-backbone residues β-alanine and 6-aminohexanoic acid, which shows an EC50 closer to 3 μM (Fig. 5.7C). Interestingly, Pep1a shares some similar motifs to Bpep, namely two Arg residues flanked by two non-α-backbone residues. If these motifs are responsible for endosomal escape, it is not surprising that Pep1a was found in the cytosol and confirmed to have significant PMO delivery activity.

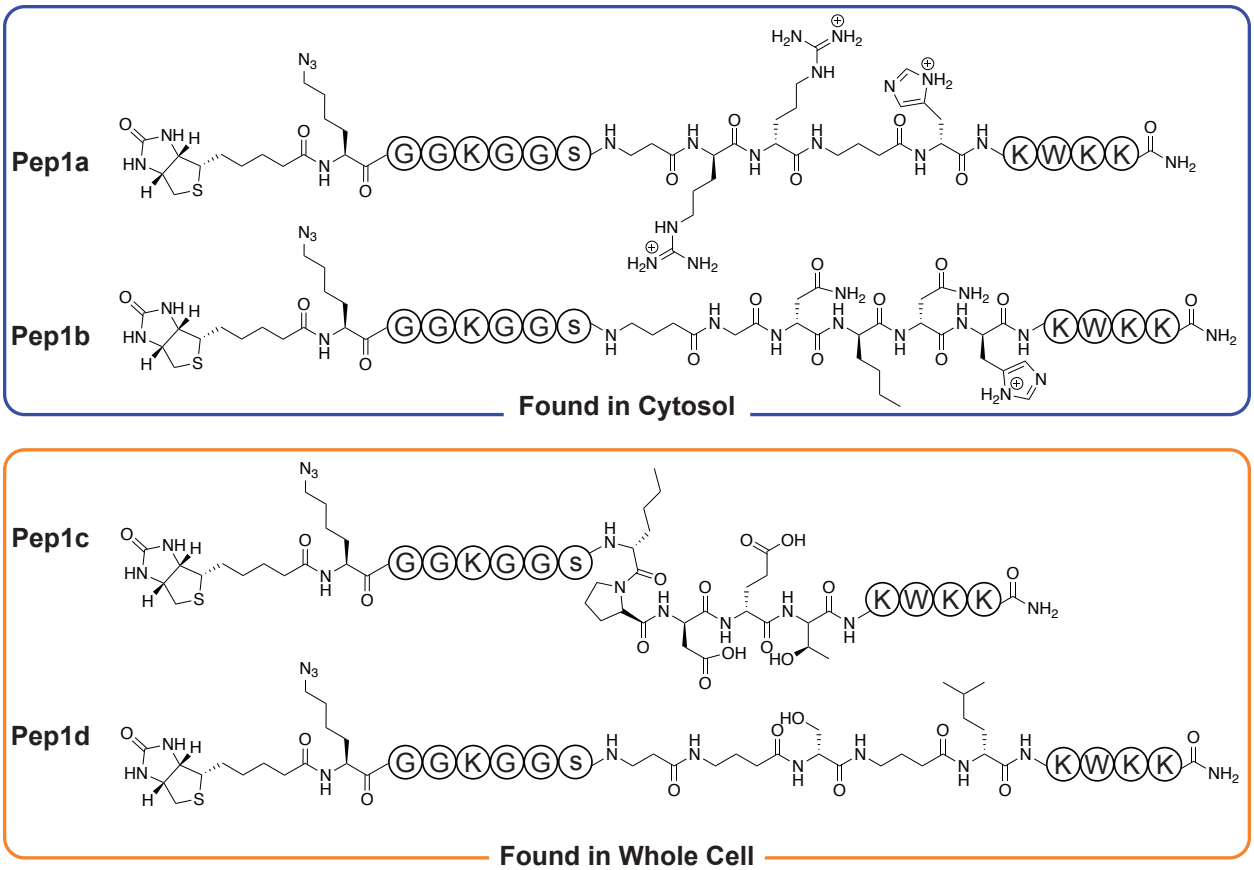


Figure 5.6 Candidate Peptides. Shown are the sequences of the four candidate peptides, grouped by their extraction from either cytosol or whole-cell lysate. Aside from the biotin and azidolysine cap, the fixed regions of the sequence are shown within circles while the regions unique to each sequence are fully drawn structures. The "s" within the sixth circle represents the isoseramox cleavable linker in the extracted and sequenced library peptides, or D-serine in the PMO-peptides synthesized for hit validation.

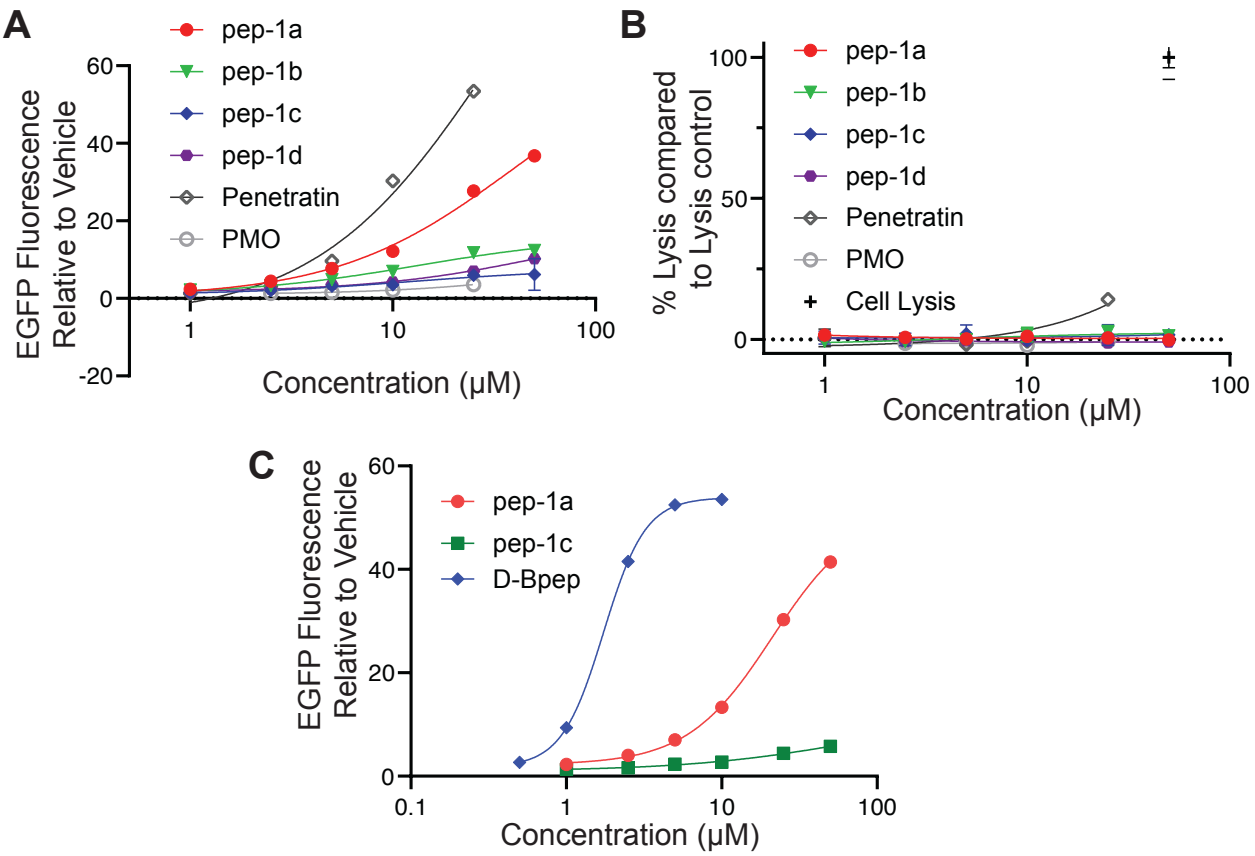


Figure 5.7 Activities of candidate peptides. (A) HeLa 654 cells were treated with 1, 2.5, 5, 10, 25, or 50 μM PMO-CPP for 22 h prior to flow cytometry. Results are given as the mean EGFP fluorescence of cells treated with PMO-peptide relative to the fluorescence of cells treated with vehicle only. All 4 peptides show similar activity across biological replicates (not shown). (B) Cell supernatant from (A) was tested for LDH release. Results are given as percent LDH release above vehicle relative to fully lysed cells. Only penetratin showed significant ($p = 0.008$) LDH release above vehicle-treated cells. (C) HeLa 654 cells were treated with 0.5, 1, 2.5, 5, 10, 25, or 50 μM PMO-CPP for 22 h. Pep1a has a significantly lower EC50 (42.8 μM) compared to pep1c with EC50 of 378 μM ($p = 0.0267$).

The bioactive hit peptides are not solely responsible for the PMO delivery exhibited by the entire 1,000-member library. Within a 20 μM treatment dose of a ~1,000 member library, each individual peptide would be present at ~20 nM, a concentration at which no single peptide is known to deliver PMO cargo. To investigate whether overall library PMO delivery efficacy could be due to a few highly active peptides, we treated HeLa 654 cells with a 250-member library at 20 μM and compared the activity to HeLa cells treated with the same library with a penetrant peptide (either Pep1a or the positive control D-Bpep) spiked in at roughly the concentration of the individual library members (Fig. 5.8A). There was no significant difference in PMO delivery between the 250-member library alone and the library with potent peptides spiked in, showing that

the overall, combined library penetration is unlikely to be affected by the activity of a few members. To further confirm this finding, we repeated the experiment with a larger library of 2,500 members and spiked in penetrant peptides at 10-fold higher concentrations than the individual library members, which also showed no significant change in library PMO delivery.

Instead of a few highly active members, library penetration is more likely the result of many cationic peptides acting in tandem. There has been previous evidence that at high concentrations (> 20 μM), cell entry of highly cationic peptides can be caused by non-specific flooding via non-endocytic pathways, via a positive feed-back loop that involves alteration of the plasma membrane composition.[46] We demonstrated that the library enters cells in an energy-independent manner, ruling out this and other mechanisms of energy-independent non-specific library entry. However, the concept of multiple cationic peptides acting in tandem at the plasma membrane suggests that such an effect could be responsible for the overall penetration of the library, from an "ensemble" of cationic peptides at the cell membrane. To test this hypothesis, we modeled the library on a much smaller scale, using only five peptides. Since we demonstrated that libraries ranging in size from 100 to 5,000 members show similar activity, we expect that even a small model may represent the activity of library peptides. Thus, we tested the four hit library candidates, as well as a "library peptide" found in the quality control sequencing of the library, but not extracted from cells, both individually and in combination for PMO delivery (Fig. 5.8B). The 5 μM "combined peptides" sample contains each individual peptide at 1 μM, yet this five-member library shows significantly more PMO delivery than any of the individual peptides at 1 μM. In fact, the PMO delivery of the peptides in combination at 5 μM total peptide more closely matches the averaged values of all five peptides individually at 5 μM, further suggesting that the activity observed from the library is due to an ensemble effect from the activity of many cationic individual peptides and not due to a few highly active sequences.

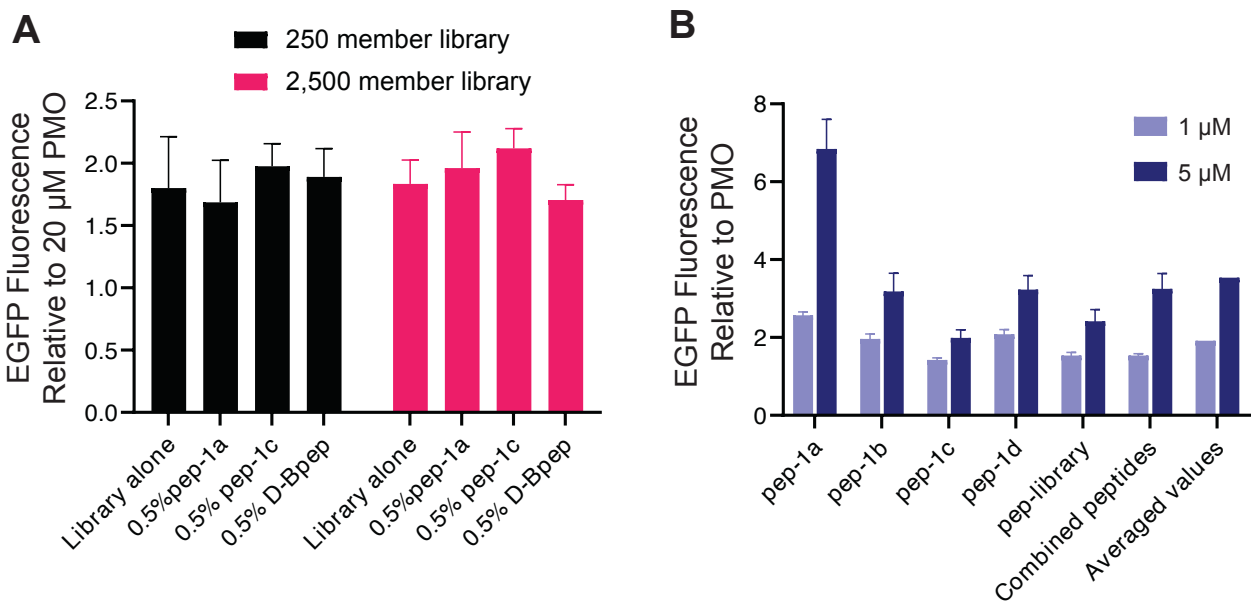


Figure 5.8 Activity of one peptide does not influence library activity. (A) HeLa 654 cells were treated with 20 µM PMO-Library or 19.9 µM PMO-Library and 0.1 µM PMO-CPP for 22 h prior to flow cytometry. Results are given as the mean EGFP fluorescence of cells treated with PMO-peptide relative to the fluorescence of cells treated with vehicle only. All 3 PMO-CPP treatment conditions are not significantly different from the library alone. (B) HeLa 654 cells were treated with 1 or 5 µM PMO-CPP or a combined solution of 5 PMO-CPPs for 22 h prior to flow cytometry. Results are given as the mean EGFP fluorescence of cells treated with PMO-peptide relative to the fluorescence of cells treated with vehicle only. Indicated concentration represents total PMO-CPP present in the sample. All PMO-CPPs at 1 µM showed significantly less PMO delivery activity than 5 µM of the combined peptides ($p < 0.05$).

5.2.4 Hit peptide demonstrates high endosomal escape activity and PMO delivery

These PMO-CPPs likely enter the cell via clathrin-mediated endocytosis as we have found for previous constructs. Pep1a was tested with a series of chemical endocytosis inhibitors in a pulse-chase format EGFP assay, in which HeLa 654 cells were pre-incubated with inhibitors to arrest various endocytosis pathways before PMO-CPPs were added. Following 3 h co-incubation, cells were washed extensively with heparin to dissociate membrane-bound constructs.[15,16,40] Activity of Pep1a was impacted by 10 μM chlorpromazine (Fig. 5.9A). Chlorpromazine is a known inhibitor of clathrin-mediated endocytosis, indicating that this conjugate may participate in this pathway. Moreover, the 4 ˚C condition also significantly impacted the activities of each conjugate, indicating that, like the entire library sample (Fig. 5.3D), uptake of the four hit peptides is energy-dependent, and the peptides are most likely entering the cells through endocytosis (Fig. 5.9B).

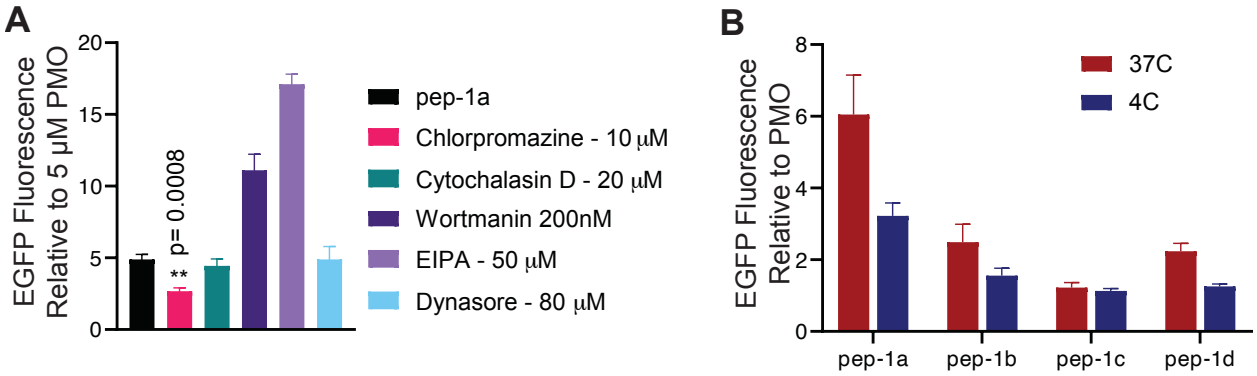


Figure 5.9 Hit peptides likely deliver PMO via clathrin-mediated endocytosis. (A) Plot of EGFP mean fluorescence intensity relative to PMO for cells treated with different endocytosis inhibitors. The cells were pre-incubated for 30 min with the indicated compound and then 5 μM PMO-Pep1a was added. After treatment with the construct for 3 h, the cells were washed with 0.1 mg/mL heparin and the media was exchanged for fresh, untreated media for 22 h prior to flow cytometry. At 10 μM chlorpromazine, EGFP fluorescence significantly decreased ($p = 0.0008$). (B) Plot of EGFP mean fluorescence intensity relative to PMO for cells incubated with PMO-CPPs at 4 ˚C or 37 ˚C. The cells were pre-incubated for 30 min at 4 ˚C or 37 ˚C, followed by the addition of PMO-peptide conjugate to each well at a concentration of 5 μM. After incubation at 4˚C or 37 ˚C for 2 h, the cells were washed with 0.1 mg/mL heparin and the media was exchanged for fresh, untreated media for 22 h prior to flow cytometry.

We further investigated the differences in activity between the sequences found in the cytosol versus the whole cell lysate and compared them to a benchmark compound, PMO-D-Bpep, using flow cytometry.[42] For this purpose, several SulfoCy5-labeled PMO-CPPs were generated and tested to ensure the fluorophore did not impact PMO delivery activity (Fig. 5.10A). Comparing the results of the EGFP assay of conjugates with and without the fluorophore, no significant differences were found between the constructs' EC50 values (Fig. 5.10B-D).

In addition, we looked at the uptake and nuclear delivery of the PMO-SulfoCy5-CPPs by monitoring both the EGFP fluorescence from the EGFP assay and the Cy5 fluorescence from the total uptake of the fluorescent analogs. Each conjugate demonstrated similar concentration-dependent increases in both EGFP and Cy5 fluorescence (Fig. 5.11A-C). Pep1c showed low fluorescence signal in both the EGFP and Cy5 channels. On the other hand, Pep1a showed higher fluorescence signals in each channel, indicating greater uptake and nuclear localization compared to Pep1c. Interestingly, D-Bpep showed greater EGFP fluorescence but slightly diminished Cy5 fluorescence than Pep1a, indicating that D-Bpep may access the nucleus more efficiently once taken up in endosomes, but with a lower total uptake compared to Pep1a. This pattern has been observed previously, in that the endosomal escape motifs present in Bpep enhance endosomal escape while diminishing cellular uptake.[8]

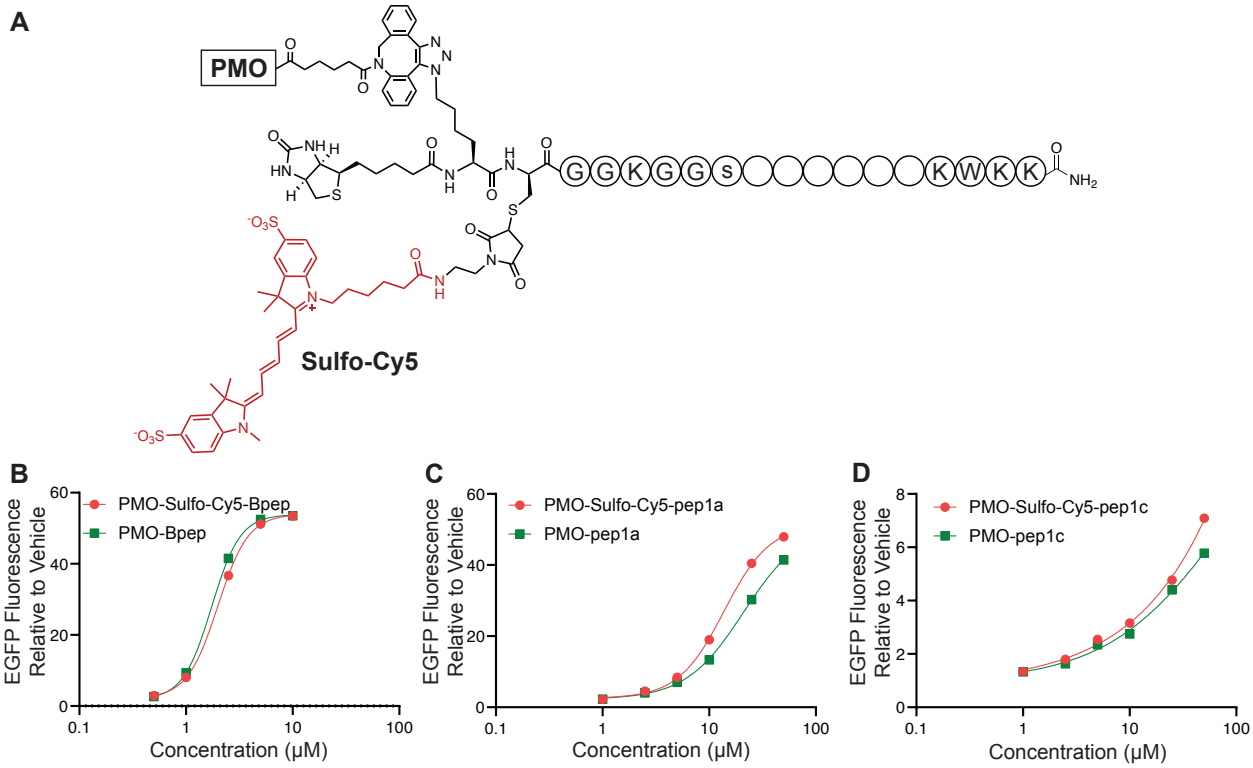


Figure 5.10 SulfoCy5 label does not impact PMO delivery. (A) Sequences of PMO-SulfoCy5-CPP constructs, with the N-terminal cargo fully drawn out (Z). Lowercase letters denote D-amino acids. (B-D) HeLa 654 cells were treated with 1, 2.5, 5, 10, 25, or 50 μM PMO-CPP or PMO-SulfoCy5-CPP for 22 h prior to flow-cytometry. Results are given as the mean EGFP fluorescence of cells treated with PMO-peptide relative to the fluorescence of cells treated with vehicle only. (B) Treatment with D-Bpep constructs. (C) Treatment with Pep1a constructs. (D) Treatment with Pep1c constructs. Pep1a showed no significant difference with the addition of SulfoCy5 at 5 μM and below ($p<0.05$), while Pep1c and D-Bpep showed no significant difference between the SulfoCy5 and standard constructs at 25 μM and below ($p<0.05$). No peptides had significant differences in EC50 between SulfoCy5 and unlabeled constructs. Bal, B (beta-Alanine), Abu (γ-aminobutyric acid), Dab (diaminobutyric acid), Nle (Norleucine), X (6-aminohexanoic acid).

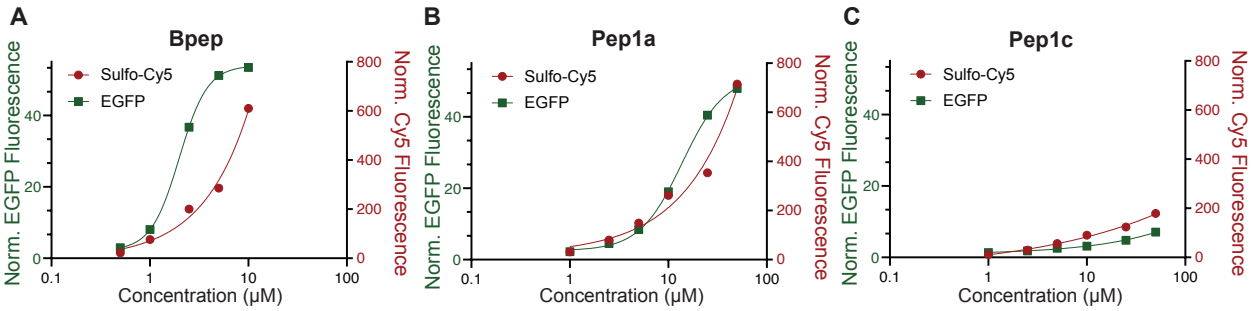


Figure 5.11 Uptake vs nuclear delivery. HeLa 654 cells were treated with 1, 2.5, 5, 10, 25, or 50 μM PMO-SulfoCy5-CPP for 22 h prior to flow cytometry. Results are given as the mean fluorescence of cells treated with PMO-SulfoCy5-peptide relative to the fluorescence of cells treated with vehicle only for each channel. (A) Treatment with D-Bpep constructs. (B) Treatment with Pep1a constructs. (C) Treatment with Pep1c constructs.

The uptake of the fluorescent conjugates into HeLa cells was also evaluated via confocal microscopy. HeLa cells were treated with 5 μM or 25 μM of SulfoCy5-labeled conjugates for 30 min, followed by a wash in complete media. Hoechst and LysoTracker Green in complete media were added to the cells immediately prior to imaging. D-Bpep was again used as a control here as it is a known highly active sequence able to escape the endosome and localize to the nucleus. Indeed, at both concentrations diffuse fluorescence is observed in the cytosol and nucleus of PMO-Bpep-treated cells, in addition to punctate fluorescence colocalizing with LysoTracker, indicating accumulation in endosomes (Fig. 5.12). Pep1a also demonstrates nuclear localization, especially at the higher concentration (Fig. 5.12B). In contrast, Pep1c shows significantly reduced overall fluorescence inside the cell, and exclusively as punctate fluorescence within endosomes.

The hit peptides discovered by PS-MS likely do not permeabilize the endosomal membrane to allow release of other cargoes. HeLa cells were first preincubated with DEAC-k5, an endosomal-localizing peptide composed of D-lysine residues.[47] The DEAC-k5 is visible as blue puncta in the no-CPP treatment control, indicating the expected endosomal localization (Fig. 5.13). After treatment with PMO-SulfoCy5-CPP constructs, the DEAC-k5 continues to occupy endosomes, indicating that Pep1a and Pep1c do not non-specifically permeabilize the endosome to release other endosomal cargo.

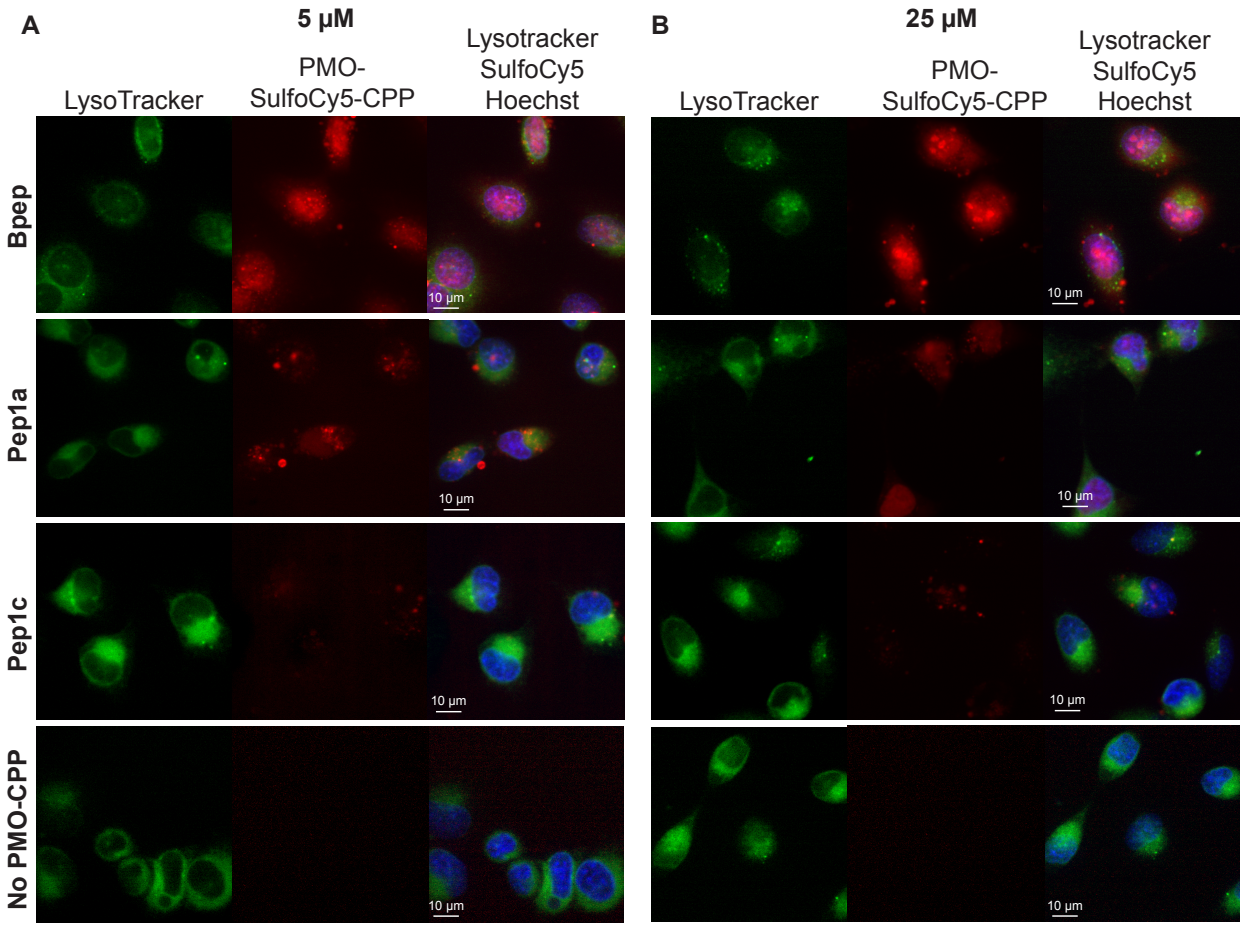


Figure 5.12 Pep1a localizes to cytosol and nucleus. Confocal micrographs of HeLa cells treated with (A) 5 µM or (B) 25 µM PMO-SulfoCy5-Bpep, PMO-SulfoCy5-Pep1a, or PMO-SulfoCy5-Pep1c. Hoechst labels the nuclei and Lyostracker Green labels the endosomes. SulfoCy5-labeled PMO-CPPs can be observed in the endosomes, cytosol, and nuclei of the cells.

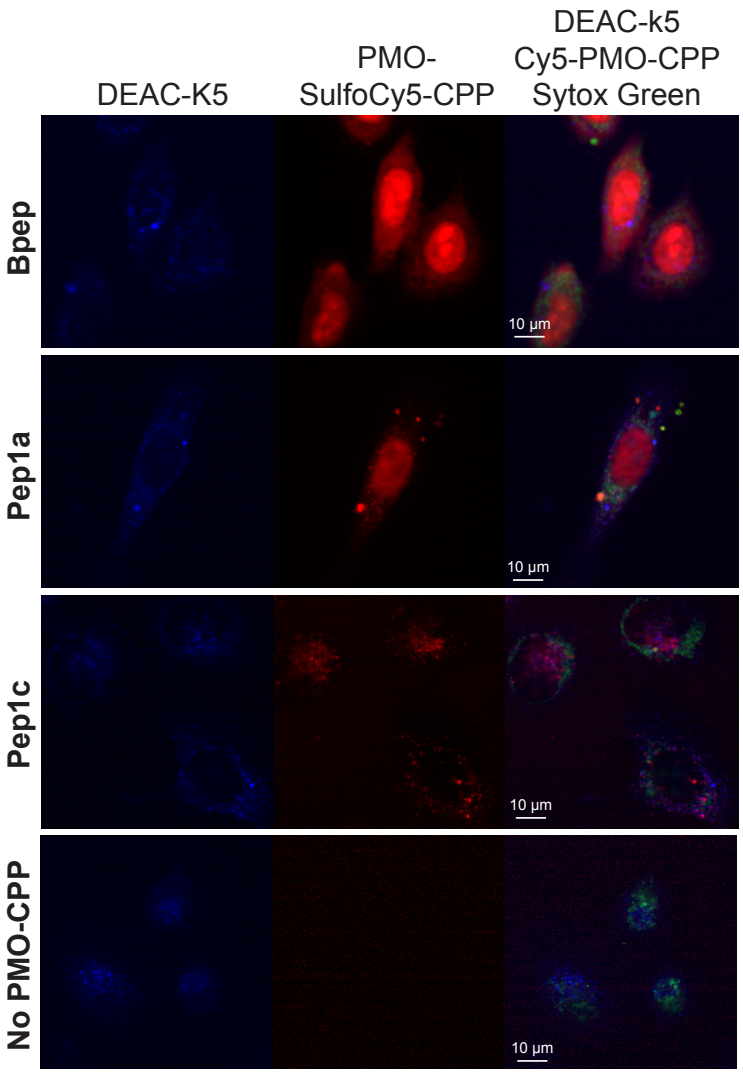


Figure 5.13 PMO-CPPs do not appear to permeabilize endosomes for general cargo release. Confocal micrographs of HeLa cells treated with 50 µM DEAC-k5 (endosome localizing peptide) followed by 25 µM of PMO-SulfoCy5-CPPs. All conjugates demonstrate fluorescent puncta likely due to accumulation in endosomes, but Bpep and Pep1a show intense nuclear staining, indicating endosomal escape. However, the DEAC-k5 appears to remain as puncta and does not show diffuse fluorescence in cytosol or nucleus. Sytox Green was added to exclude observation of dead cells.

5.3 Discussion

Affinity selection-mass spectrometry (AS-MS) techniques have traditionally been used to probe protein-protein interactions in vitro.[23] Our group has recently shown that chemical libraries may reach the diversity of other display techniques for identification of peptide binders to proteins[22], and that this strategy can be applied for cell-surface selection in vivo.[31] In this work, we have added an additional spatial element to this strategy, by extracting the cytosol for in-cell selection of fully synthetic peptide libraries conjugated to a model antisense cargo. By comparing these sequences to those found in a whole cell extract, we can exclude sequences that accumulate in the endosomes.

In-cell PS-MS combined with subcellular fractionation resulted in the identification of a novel, abiotic peptide capable of accessing the cytosol and delivering PMO to the nucleus. Pep1a, like the positive control peptide Bpep, was able to deliver PMO to the nucleus by escaping endosomes. Furthermore, Pep1a does not appear to permeabilize the endosome to allow the escape of other endosomal cargo, nor does it demonstrate cell membrane toxicity. All peptides discovered through this novel platform demonstrated lower toxicity than the CPP penetratin, which contributed the fixed "CPP-like" C-terminal region in the library. Endowed with lower toxicity and superior chemical diversity provide by the noncanonical residues, the peptides discovered with the in-cell PS-MS platform show advantages over the library's parent peptide. The few active PMO-CPPs individually sequenced and validated are not solely responsible for the overall cell penetration of the library, however. In fact, it is more likely that the PMO delivery arises from the combined activity of a number of peptides at low concentration, including the hits discovered with our platform.

Current investigations in our laboratory aim to further combine this method with orthogonal approaches in order to better focus the selection on successful peptides. As such, we envisage using chromatographic fractionation of the peptide libraries to enrich for penetrant peptides within the library before conducting the PS-MS uptake assay. This chromatographic pre-enrichment would allow for screening of an overall less active, but more diverse library, perhaps obviating the need for a pre-installed fixed C-terminal penetrating motif.

5.4 Materials & Methods

5.4.1 Reagents and Solvents

H-Rink Amide-ChemMatrix resin was obtained from PCAS BioMatrix Inc. (St-Jean-sur-Richelieu, Quebec, Canada) and TentaGel was obtained from Rapp Polymere (Tuebingen, Germany). 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium-3- oxid-hexafluorophosphate (HATU), Fmoc-L-Lys(N_3), Fmoc-β-Ala-OH, Fmoc-D-norleucine, *N*α-Fmoc-*N*γ-Boc-D-2,4-diaminobutyric acid, Fmoc-D-homoleucine, Fmoc-3,3-diphenyl-D-alanine, Fmoc-3-(1-naphthyl)-D-alanine, Fmoc-4-(Boc-aminomethyl)-D-phenylalanine, 1-Boc-piperidine-4-Fmoc-amino-4-carboxylic acid, Fmoc-γ-aminobutyric acid, and 2-(Fmoc-amino)-4-(bis-Boc-guanidino)-D-butyric acid were purchased from Chem-Impex International (Wood Dale, IL). Fmoc-protected D-amino acids (Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(O*t*-Bu)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(O*t*-Bu)-OH; Fmoc-Gly-OH; Fmoc-His(Trt)-OH; Fmoc-Lys(Boc)-OH; Fmoc-Phe-OH; Fmoc-Pro-OH; Fmoc-Ser(But)-OH; Fmoc-Thr(*t*-Bu)-OH; Fmoc-Trp(Boc)-OH), were purchased from the Novabiochem-line from MilliporeSigma. Sulfo-Cyanine5 maleimide was purchased from Lumiprobe (Cockeysville, MD), and 7-diethylaminocoumarin-3-carboxylic acid was purchased from AAT Bioquest (Sunnyvale, CA). Dibenzocyclooctyne acid was purchased from Click Chemistry Tools (Scottsdale, AZ). Cytochalasin D was obtained from Santa Cruz Biotech. Peptide synthesis-grade *N,N*-dimethylformamide (DMF), CH_2Cl_2 (DCM), diethyl ether, and HPLC-grade acetonitrile were obtained from VWR International (Radnor, PA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO). Milli-Q water was used exclusively. Hoechst 33342 and LysoTracker™ Green were purchased from ThermoFisher Scientific (Walthan, MA). The LDH Assay kit was purchased from Promega (Madison, WI).

5.4.2 Liquid chromatography—mass spectrometry

LC-MS analyses were performed on an Agilent 6550 iFunnel Q-TOF LC-MS system (abbreviated as 6550) coupled to an Agilent 1290 Infinity HPLC system. Mobile phases were: 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LC-MS method was used for characterization:

Method A: 1-61% B over 6 min, Zorbax C3 column (6550)

LC: Agilent EclipsePlus C18 RRHD column: 2.1 × 50 mm, 1.8 μm, column temperature: 40 °C, gradient: 0-1 min 1% B, 1-6 min, 1-61% B, 6-7 min, 91% B, 7-8 min, 1% B; flow rate: 0.5 mL/min. MS: Positive electrospray ionization (ESI) extended dynamic range mode in mass range 300–3000 m/z. MS is on from 1 to 6 min.

All data were processed using Agilent MassHunter software package. Y-axis in all chromatograms shown represents total ion current (TIC) unless noted.

5.4.3 General peptide preparation

Fast-flow Peptide Synthesis: Peptides were synthesized on a 0.1 mmol scale using an automated fast-flow peptide synthesizer for L-peptides and a semi-automated fast-flow peptide synthesizer for D-peptides.[45] Automated synthesis conditions were used as previously reported.[48] Briefly, a 100 mg portion of ChemMatrix Rink Amide HYR resin was loaded into a reactor maintained at 90 ºC. All reagents were flowed at 40 mL/min with HPLC pumps through a stainless-steel loop maintained at 90 ºC before introduction into the reactor. For each coupling, 10 mL of a solution containing 0.4 M amino acid and 0.38 M HATU in DMF were mixed with 600 μL of diisopropylethylamine and delivered to the reactor. Fmoc removal was accomplished using 10.4 mL of 20% (v/v) piperidine. Between each step, DMF (15 mL) was used to wash out the reactor. To couple unnatural amino acids or to cap the peptide (e.g., with 4-pentynoic acid), the resin was incubated for 30 min at room temperature with amino acid (1 mmol) dissolved in 2.5 mL of 0.4 M HATU in DMF with 500 μL of diisopropylethylamine (DIEA). After completion of the synthesis, the resin was washed 3 times with dichloromethane and dried under vacuum.

Semi-automated synthesis was carried out as previously described.[45] 1 mmol of amino acid was combined with 2.5 mL of 0.4 M HATU and 500 μL of DIEA and mixed before being delivered to the reactor containing resin via syringe pump at 6 mL/min. The reactor was submerged in a water bath heated to 70 ºC. An HPLC pump delivered either DMF (20 mL) for washing or 20% piperidine/DMF (6.7 mL) for Fmoc deprotection, at 20 mL/min.

Peptide Cleavage and Deprotection: Each peptide was subjected to simultaneous global side-chain deprotection and cleavage from resin by treatment with 5 mL of 94% trifluoroacetic acid (TFA), 2.5% thioanisole, 2.5% water, and 1% triisopropylsilane (TIPS) (v/v) at room temperature for 2 to 4 h. The cleavage cocktail was first concentrated by bubbling N_2 through the mixture, and cleaved peptide was precipitated and triturated with 40 mL of cold ether (chilled in dry ice). The

crude product was pelleted by centrifugation for three minutes at 4,000 rpm and the ether was decanted. This wash step was repeated two more times. After the third wash, the pellet was dissolved in 50% water and 50% acetonitrile containing 0.1% TFA, filtered through a fritted syringe to remove the resin and lyophilized.

Peptide Purification: The peptides were dissolved in water and acetonitrile containing 0.1% TFA, filtered through a 0.22 μm nylon filter and purified by mass-directed semi-preparative reversed-phase HPLC. Solvent A was water with 0.1% TFA additive and Solvent B was acetonitrile with 0.1% TFA additive. A linear gradient that changed at a rate of 0.5% B/min was used. Most of the peptides were purified on an Agilent Zorbax SB C18 column: 9.4 x 250 mm, 5 μm. Based on target ion mass data recorded for each fraction, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

Preparation of PMO-Peptides

PMO IVS2-654 (50 mg, 8 μmol) obtained from Sarepta Therapeutics was dissolved in 150 μL DMSO. To the solution was added a solution containing 2 equivalents of dibenzocyclooctyne acid (5.3 mg, 16 μmol) activated with HBTU (37.5 μL of 0.4 M HBTU in DMF, 15 μmol) and DIEA (2.8 μL, 16 μmol) in 40 μL DMF (Final reaction volume = 0.23 mL). The reaction proceeded for 25 min before being quenched with 1 mL of water and 2 mL of ammonium hydroxide. The ammonium hydroxide hydrolyzed any ester formed during the course of the reaction. After 1 hour, the solution was diluted to 40 mL in water/acetonitrile and purified using reverse-phase HPLC (Agilent Zorbax SB C3 column: 21.2 x 100 mm, 5 μm) and a linear gradient from 2 to 60% B (solvent A: water; solvent B: acetonitrile) over 58 min (1% B / min). Using mass data about each fraction from the instrument, only pure fractions were pooled and lyophilized. The purity of the fraction pool was confirmed by LC-MS.

Conjugation to peptides

PMO-DBCO (1 eq, 5 mM, water) was conjugated to azido-peptides (1 eq, 5 mM, water) at room temperature for 2 h. Reaction progress was monitored by LC-MS and additional stock of 5 mM azido-peptide was added until all PMO-DBCO was consumed. The purity of the final construct was confirmed by LC-MS.

5.4.4 Preparation of peptide libraries

Split-and-pool synthesis was carried out on 300 μm TentaGel resin (0.23 mmol/g) for a 95,000 member library. Splits were performed by suspending the resin in DCM and dividing it evenly (via

pipetting) among 22 plastic fritted syringes on a vacuum manifold. Couplings were carried out as follows: solutions of Fmoc-protected amino acids (10 equivalents relative to the resin loading), PyAOP (0.38 M in DMF; 0.9 eq. relative to amino acid), and DIEA (1.1 eq. for histidine; 3 eq. for all other amino acids) were each added to individual portions of resin. Couplings were allowed to proceed for 60 min. Resin portions were recombined and washed with DCM and DMF. Fmoc removal was carried out by treatment of the resin with 20% piperidine in DMF (1x flow wash; 2x 5 min batch treatments). Resin was washed again with DMF and DCM before the next split.

5.4.5 EGFP Assay

HeLa 654 cells obtained from the University of North Carolina Tissue Culture Core facility were maintained in MEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin-streptomycin at 37 °C and 5% CO_2. 18 h prior to treatment, the cells were plated at a density of 5,000 cells per well in a 96-well plate in MEM supplemented with 10% FBS and 1% penicillin-streptomycin.

For individual peptide testing, PMO-peptides were dissolved in PBS without Ca^{2+} or Mg^{2+} at a concentration of 1 mM (determined by UV) before being diluted in MEM. Cells were incubated at the designated concentrations for 22 h at 37 °C and 5% CO_2. Next, the treatment media was removed, and the cells were washed once before being incubated with 0.25 % Trypsin-EDTA for 15 min at 37 °C and 5% CO_2. Lifted cells were transferred to a V-bottom 96-well plate and washed once with PBS, before being resuspended in PBS containing 2% FBS and 2 μg/mL propidium iodide (PI). Flow cytometry analysis was carried out on a BD LSRII flow cytometer. Gates were applied to the data to ensure that cells that were positive for propidium iodide or had forward/side scatter readings that were sufficiently different from the main cell population were excluded. Each sample was capped at 5,000 gated events.

Analysis was conducted using Graphpad Prism 7 and FlowJo. For each sample, the mean fluorescence intensity (MFI) and the number of gated cells was measured. To report activity, triplicate MFI values were averaged and normalized to the PMO alone condition. For the final set of PMO-peptides evaluated, three biological replicates were performed.

5.4.6 Endocytosis Inhibition Assay

Chemical endocytosis inhibitors were used to probe the mechanism of delivery of PMO by these peptides in a pulse-chase format. We have conducted such analysis on similar PMO-peptide

constructs previously with comparable outcomes.[40] For the PMO constructs, HeLa 654 cells were preincubated with various chemical inhibitors for 30 minutes before treatment with PMO-CPP constructs for three hours. The panel of endocytosis inhibitors included: 10 μM chlorpromazine (CPZ), which is demonstrated to interfere with clathrin-mediated endocytosis; 20 μM cytochalasin D (CyD), which inhibits phagocytosis and micropinocytosis; 200 nM wortmannin (Wrt), which alters various endocytosis pathways by inhibiting phosphatidylinositol kinases; 50 μM EIPA (5-(*N*-ethyl-*N*-isopropyl)amiloride), which inhibits micropinocytosis; and 80 μM Dynasore (Dyn), which also inhibits clathrin-mediated endocytosis.[49,50] Treatment media was then replaced with fresh media and the cells were incubated for 22 h at 37 °C and 5% CO_2. Cells were then lifted as previously described and EGFP synthesis was measured by flow cytometry.

5.4.7 LDH Assay

Cytotoxicity assays were performed in HeLa 654 cells. Cell supernatant following treatment for flow cytometry was transferred to a new 96-well plate for analysis of LDH release. To each well of the 96-well plate containing supernatant was added CytoTox 96 Reagent (Promega). The plate was shielded from light and incubated at room temperature for 30 min. Equal volume of Stop Solution was added to each well, mixed, and the absorbance of each well was measured at 490 nm. The measurement of vehicle-treated cells was subtracted from each measurement, and % LDH release was calculated as % cytotoxicity = 100 × Experimental LDH Release (OD490) / Maximum LDH Release (OD490).

5.4.8 Microscopy

HeLa cells were plated at a density of 8,000 cells/well in a 96-well cover glass-bottomed plate the day before the experiment. For standard localization imaging, cells were treated with PMO-Sulfo-Cy5-CPP conjugates at 5 μM or 25 μM for 30 min at 37 °C and 5% CO_2. Each well was washed with media and incubated in fresh media for 1 h before Hoechst (nuclear) and Lysotracker Green (endosomal) fluorescent tracking dyes were added, and imaged immediately. The endosomal release experiment was adapted from a previously reported protocol.[47] Cells were treated with 50 μM 7-Diethylaminocoumarin-3-carboxylic acid (DEAC)-k5 for 1 h at 37 °C and 5% CO_2 before being washed with media. Then, PMO-Sulfo-Cy5-CPP conjugates were added as before. Sytox Green was added immediately before imaging in order to exclude observation of

nonviable cells. Imaging was performed at the Whitehead Institute's Keck Imaging Facility on an RPI Spinning Disk Confocal Microscope at 40x objective.

5.4.9 Uptake Assay

Cell treatment

Cells were plated either in 6-well or 12-well plates at a density such that they reached 80% confluency the following day. CPP or PMO-CPP stock solutions were made fresh to 1 mM in cation-free PBS, as determined by UV-Vis. Treatment solution was then prepared by adding the stock solution to cell media at the concentrations described. Two wells were left untreated as controls. The plates were then incubated at 37 °C and 5% CO_2 for the designated time. For the experiment to arrest energy-dependent uptake, the plate was incubated at 4 °C. Following incubation, the cells were washed three times with media, followed by 0.1 mg/mL heparin in PBS for 5 min. Supernatant was aspirated and cells were lifted by incubating in trypsin-EDTA for 10 min at 37 °C. Trypsin was quenched by adding cell media, and cells were transferred to Eppendorf tubes and pelleted at 500 rcf for 3 min. Pellets were washed by mixing with PBS, repeated twice.

Lysis

To acquire whole cell lysate, 50 µL RIPA buffer, protease inhibitor cocktail, water) was added to the cell pellet, mixed gently, and placed on ice for 1 h. To extract the cytosol, 50 µL digitonin buffer (0.05 mg/mL digitonin, 250 mM sucrose, PBS) was added to a cell pellet, mixed very gently, and placed on ice for 10 min. Samples were then pelleted by centrifugation at 16,000 rcf for 5 min. Supernatants were transferred to new Eppendorf tubes and kept on ice. Extracted protein from the cell-only control samples was quantified using Pierce Rapid Gold BCA Protein Assay Kit (Thermo Fisher). 10 µg protein from each sample was then analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel for 35 min at 165 V and then transferred to a nitrocellulose membrane soaked in 48 mM Tris, 39 mM glycine, 0.0375% SDS, 20% methanol using a TransBlot Turbo Semi-Dry Transfer Unit (BioRad) for 7 min. The membrane was blocked at 4 °C overnight in LI-Cor Odyssey blocking buffer in PBS. The membrane was then immunostained for 1 h with anti-Erk1/2 and anti-Rab5 (Cell Signaling) in PBS-Tween at room temperature. After incubation, the membrane was washed three times with TBST and incubated

with the appropriate secondary antibody in TBST for 1 h at room temperature, then washed with TBST. The membrane was imaged on a ChemiDoc MP Imaging System (Bio-Rad).

Penetration Selection

10 µL Dynabeads™ MyOne™ Streptavidin T1 (Thermo Fisher) were transferred to tubes in a magnet stand and washed with PBS. Cell extracts were added to the corresponding bead-containing tube and rotated at 4 °C for 2 h. To two of the cell only sample lysates was added each 0.5 µL of PMO-biotin-library and 0.5 µL of biotin-library, and was combined with 50 µL of Streptavidin beads. Following pulldown, the beads were washed with 6 M guanidinium chloride (GuHCl, pH 6.8, 2 x 200 µL) and suspended in 100 µL PBS. Then, NaIO4 (1 mM in H_2O, 2 µL) was added and the beads incubated for 5 min in absence of light, followed by quench solutions: Na_2SO_3 (100 mM, 5 µL) and NH_2OH (100 mM, 5 µL). The supernatants were transferred to new tubes and the beads were washed with 6 M GuHCl (2 x 100 µL). Pooled supernatant fractions were then desalted by solid-phase extraction (SPE) using C18 ZipTips, lyophilized, and rehydrated in 10 µL 1 M GuHCl in water containing 0.1% formic acid.

Orbitrap LC-MS/MS

Analysis was performed on an EASY-nLC 1200 (Thermo Fisher Scientific) nano-liquid chromatography handling system connected to an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific). Samples were run on a PepMap RSLC C18 column (2 µm particle size, 15 cm × 50 µm ID; Thermo Fisher Scientific, P/N ES901). A nanoViper Trap Column (C18, 3 µm particle size, 100 Å pore size, 20 mm × 75 µm ID; Thermo Fisher Scientific, P/N 164946) was used for desalting. The standard nano-LC method was run at 40 °C and a flow rate of 300 nL/min with the following gradient: 1% solvent B in solvent A ramping linearly to 41% B in A over 55 min, where solvent A = water (0.1% FA), and solvent B = 80% acetonitrile, 20% water (0.1% FA). Positive ion spray voltage was set to 2200 V. Orbitrap detection was used for primary MS, with the following parameters: resolution = 120,000; quadrupole isolation; scan range = 150–1200 m/z; RF lens = 30%; AGC target = 250%; maximum injection time = 100 ms; 1 microscan. Acquisition of secondary MS spectra was done in a data-dependent manner: dynamic exclusion was employed such that a precursor was excluded for 30 s if it was detected four or more times within 30 s (mass tolerance: 10.00 ppm); monoisotopic precursor selection used to select for

peptides; intensity threshold was set to 2×10^4; charge states 2–10 were selected; and precursor selection range was set to 200–1400 m/z. The top 15 most intense precursors that met the preceding criteria were subjected to subsequent fragmentation. Two fragmentation modes— higher-energy collisional dissociation (HCD), and electron-transfer/higher-energy collisional dissociation (EThcD)—were used for acquisition of secondary MS spectra. Detection was performed in the Orbitrap (resolution = 30,000; quadrupole isolation; isolation window = 1.3 m/z; AGC target = 2×10^4; maximum injection time = 100 ms; 1 microscan). For HCD, a stepped collision energy of 3, 5, or 7% was used. For EThcD, a supplemental activation collision energy of 25% was used.

De novo peptide sequencing and filtering

De novo peptide sequencing of the acquired data was performed in PEAKS 8 (BioInformatics Solutions Inc.). Using PEAKS, spectra were prefiltered to remove noise, and sequenced. All non-canonical amino acids were sequenced as post-translational modifications based on the canonical amino acid most closely matching their molecular mass. Norleucine and β-Alanine were sequenced as leucine and alanine, respectively. In addition, His-oxide and Met-oxide were allowed as variable post-translational modifications. Twenty candidate sequence assignments were created for each secondary scan.

An automated Python-based routine was used for postprocessing data analysis to eliminate noise, synthetic impurities, duplicates, resolve certain sequencing ambiguities, and to select the best candidate sequence assignment for each MS/MS scan. The script eliminates all sequence candidates of length other than 9 or 10 and all candidates not bearing the C-terminal KWKK motif. Next, for each remaining spectrum, a single candidate is kept, discarding all other peptides with lower sequencing scores from PEAKS (average local confidence (ALC) scores, from 0 to 99), and duplicate sequences are labeled as non-unique. Finally, the resulting unique sequence assignments are refined further by excluding prominent synthetic impurities that were not eliminated in the previous steps. If two unique sequences have an identifiable main product/side product relationship, the side product is eliminated. In this way, peptides containing oxidized Met residues, deamidation of Gln to Glu, which occasionally happens during saponification of PAM ester, sodium adducts, and a few less prominent side-reactions are identified, and their corresponding sequences are discarded.

5.4.10 Statistics

Statistical analysis and graphing was performed using Prism (Graphpad) or Excel (Microsoft). Concentration-response curves were fitted using Prism using nonlinear regression. The listed replicates for each experiment indicates the number of distinct samples measured for a given assay. Significance for activities between constructs was determined using a student's two-sided, unpaired t-test.

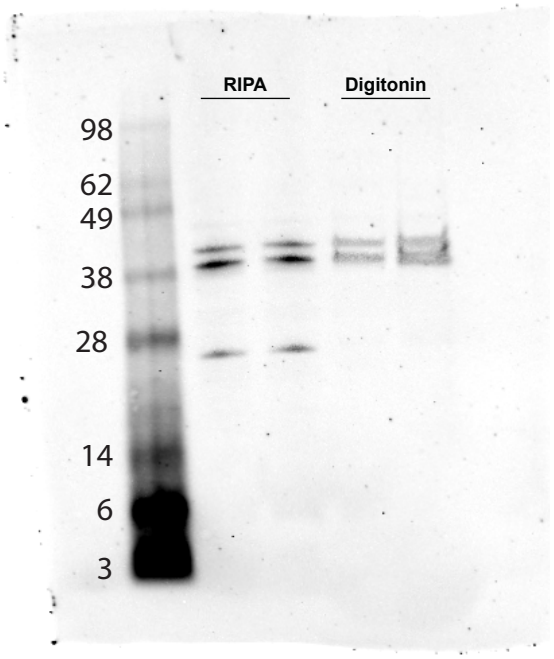
5.5 Acknowledgements

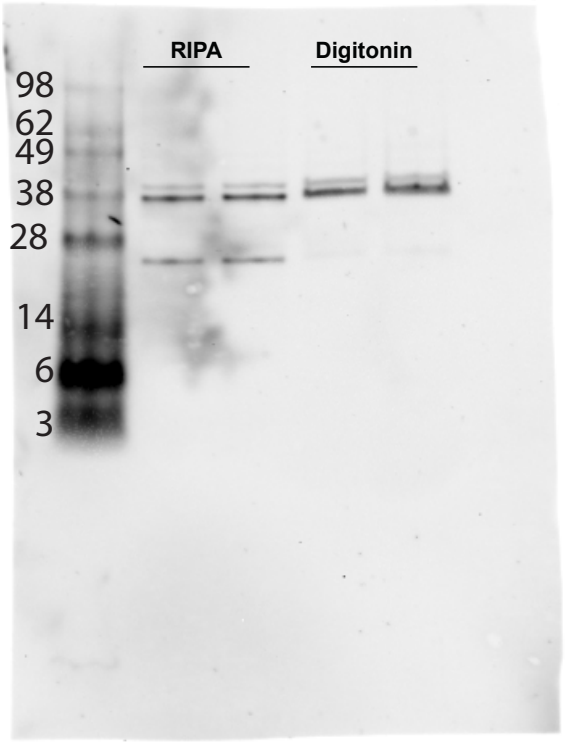
This research was funded by Sarepta Therapeutics. C.K.S. (4000057398) and C.E.F. (4000057441) acknowledge the National Science Foundation Graduate Research Fellowship (NSF Grant No. 1122374) for research support. We acknowledge support from the Swanson Biotechnology Center Flow Cytometry Facility at the Koch Institute for Integrative Cancer Research at MIT through the use of their flow cytometers (NCI Cancer Center Support Grant P30-CA14051). We thank Cassandra Rogers and Brandyn Braswell at the Whitehead Institute's Keck Imaging facility for use of the confocal microscope.

5.6 Author Contributions

C.K.S. and B.L.P conceptualized the research. C.K.S. and C.E.F. performed experiments and analyzed the data. C.K.S. C.E.F. A.L. and B.L.P. wrote the manuscript with input from all authors.

5.7 Appendix I: Gel Images

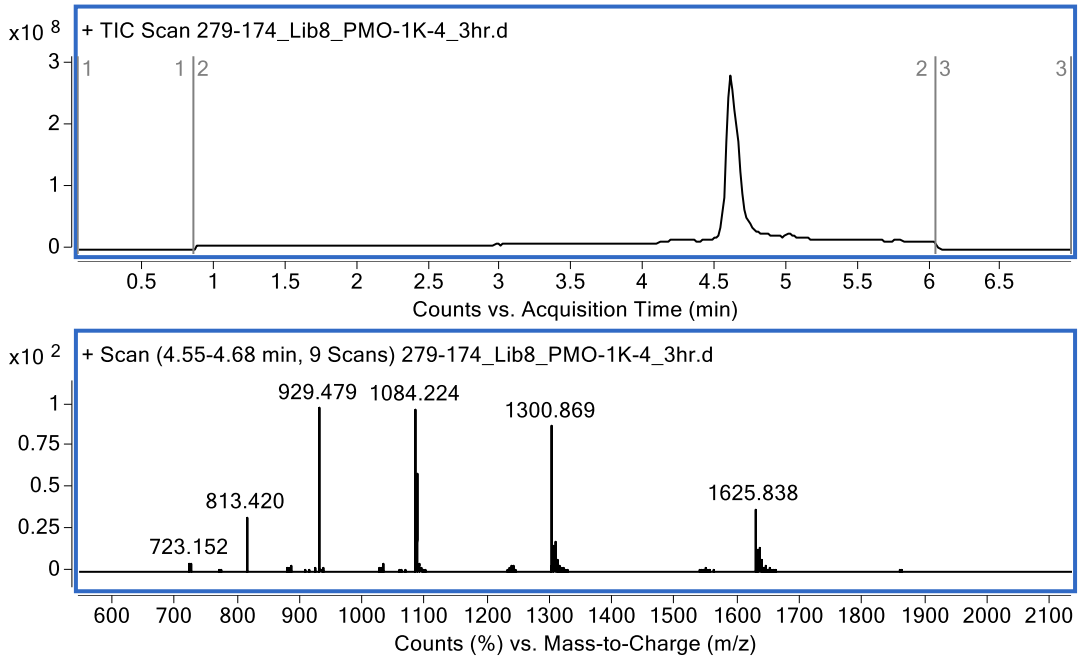




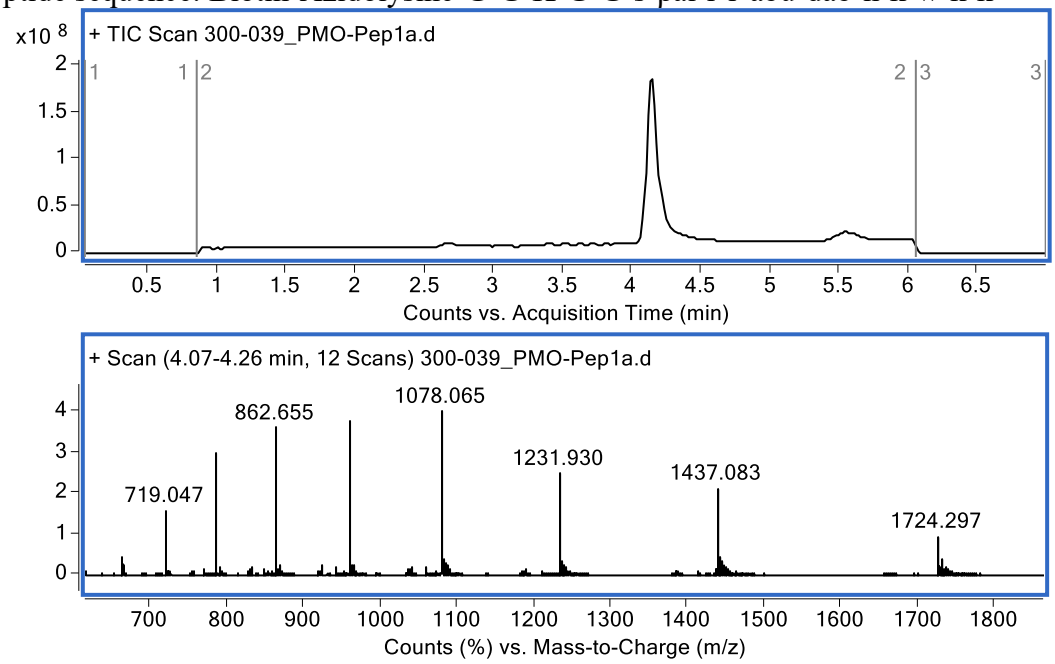
5.8 Appendix II: LC-MS Characterization

Note: Chromatograms were obtained using Method A

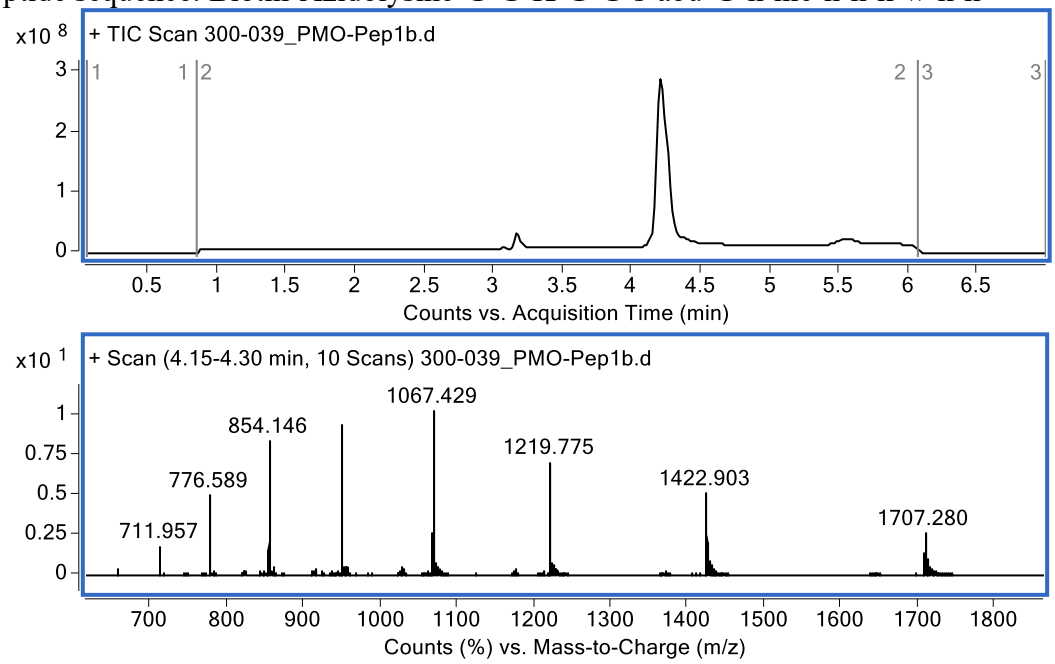
PMO-DBCO
Mass Expected: 6500.0
Mass Observed: 6499.9



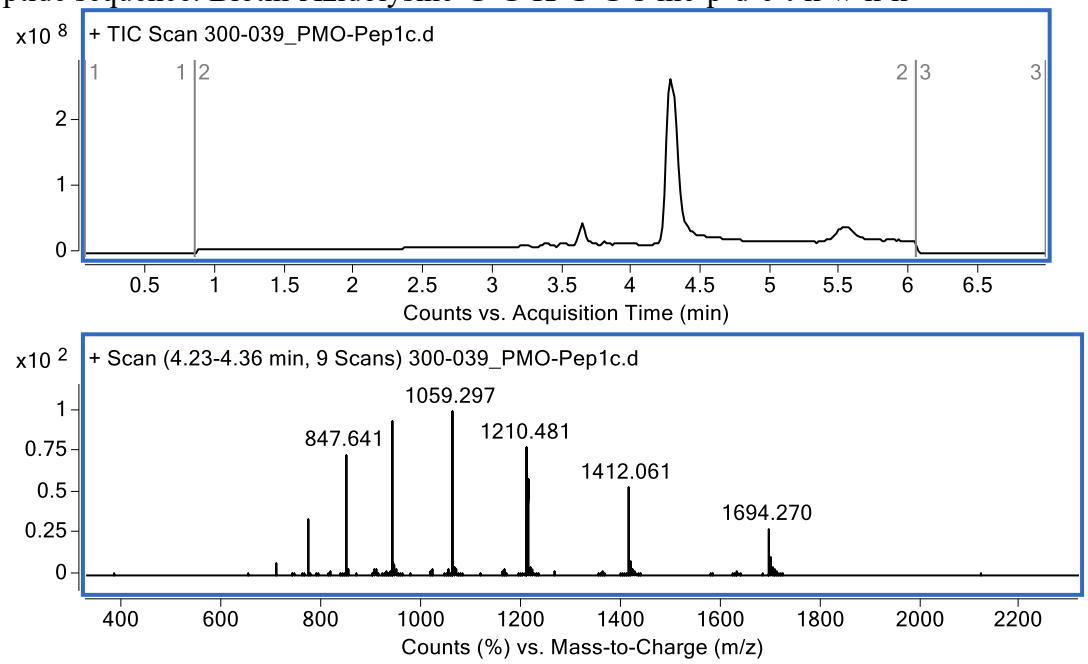
PMO-Pep1a
Mass Expected: 8617.3
Mass Observed: 8617.5
Peptide sequence: Biotin-Azidolysine-G-G-K-G-G-s-βal-r-r-abu-dab-h-k-w-k-k



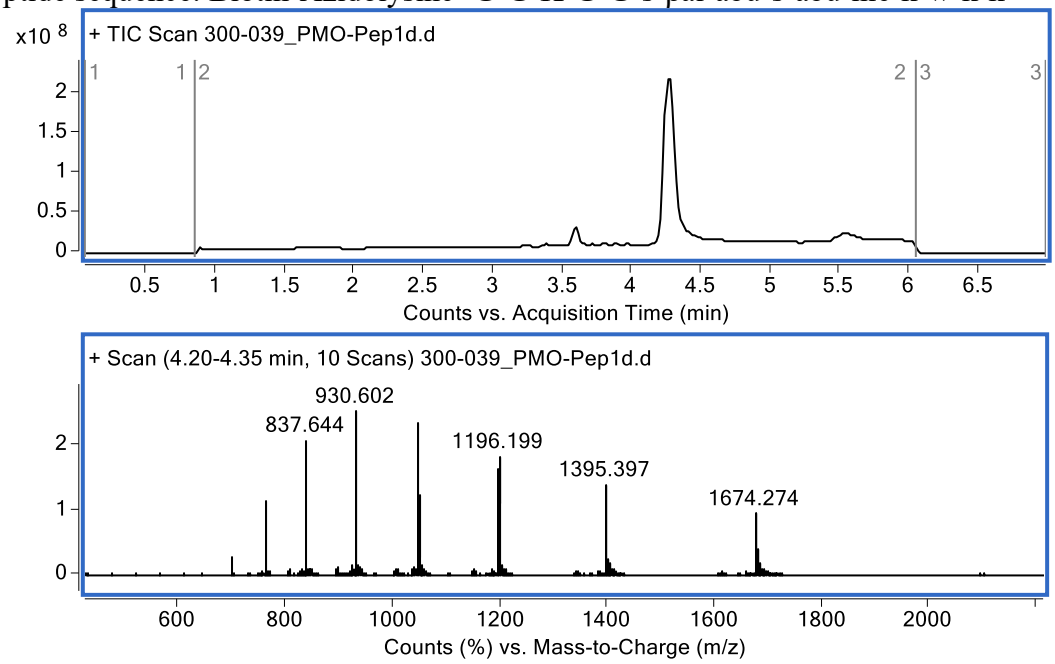
PMO-Pep1b
Mass Expected: 8532.4
Mass Observed: 8532.4
Peptide sequence: Biotin-Azidolysine-G-G-K-G-G-s-abu-G-n-nle-n-h-k-w-k-k



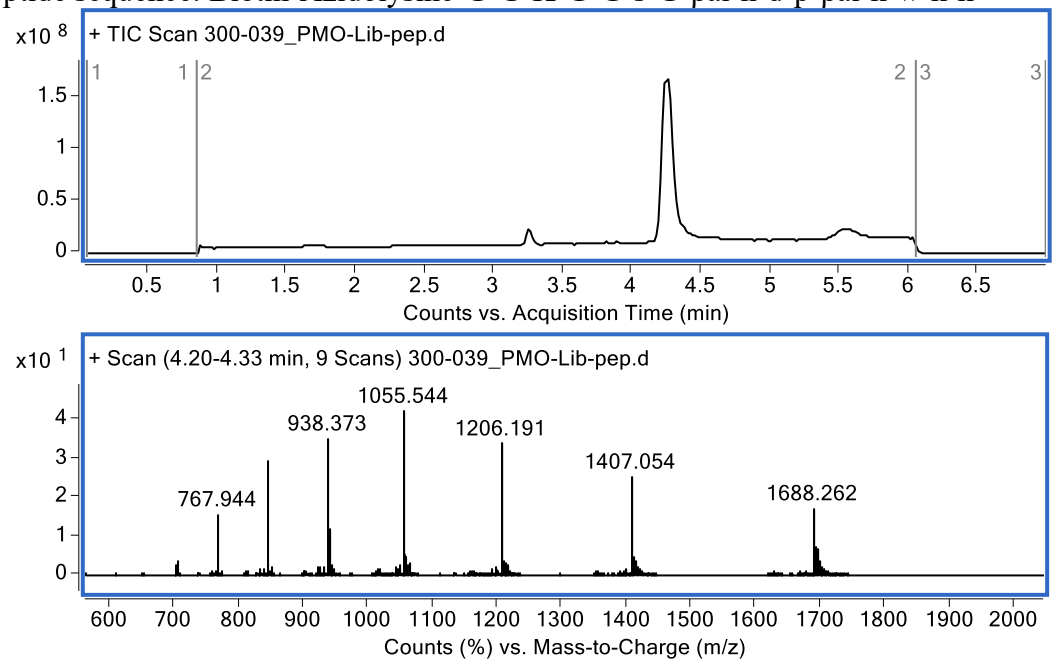
PMO-Pep1c
Mass Expected: 8467.3
Mass Observed: 8467.3
Peptide sequence: Biotin-Azidolysine-G-G-K-G-G-s-nle-p-d-e-t-k-w-k-k



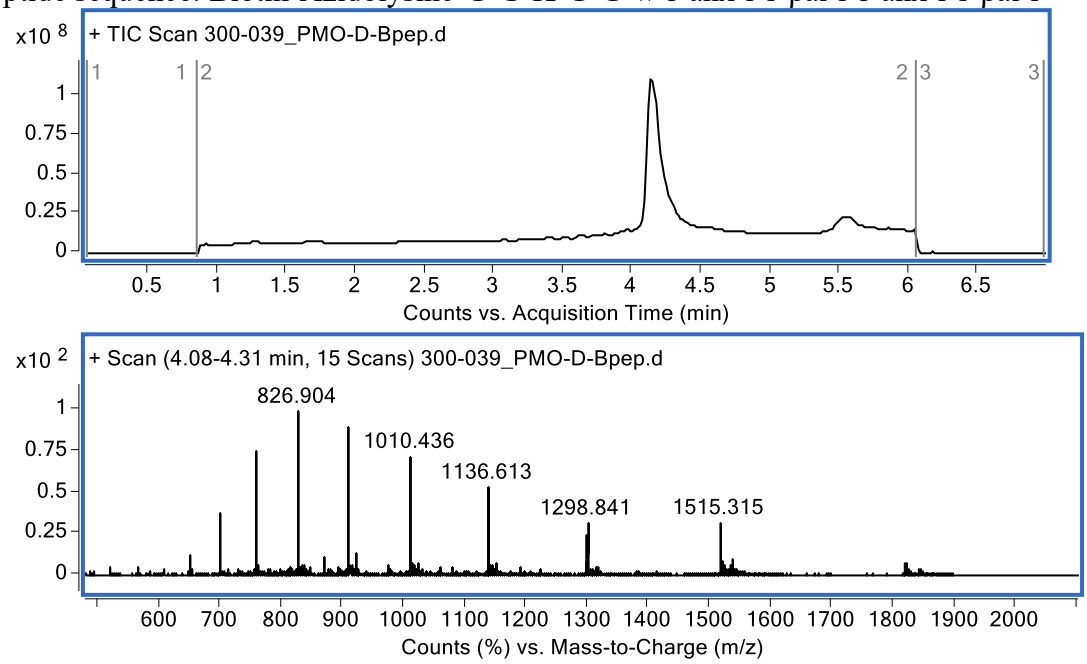
PMO-Pep1d
Mass Expected: 8367.2
Mass Observed: 8367.2
Peptide sequence: Biotin-Azidolysine- G-G-K-G-G-s-βal-abu-s-abu-hle-k-w-k-k



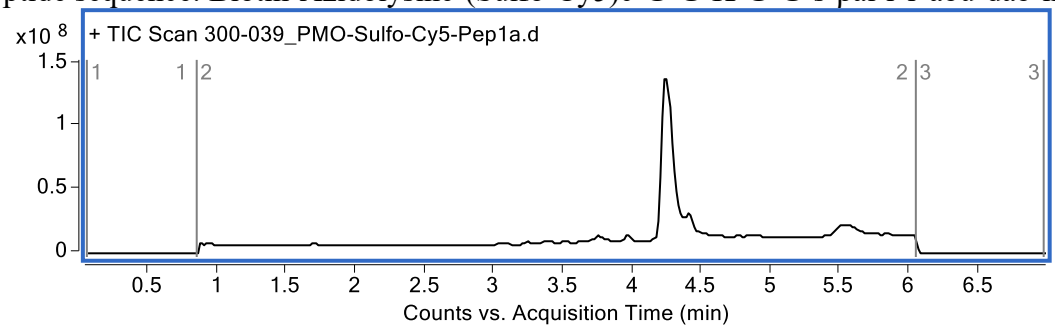
PMO-library-peptide
Mass Expected: 8437.2
Mass Observed: 8437.1
Peptide sequence: Biotin-Azidolysine-G-G-K-G-G-s-G-βal-n-d-p-βal-k-w-k-k

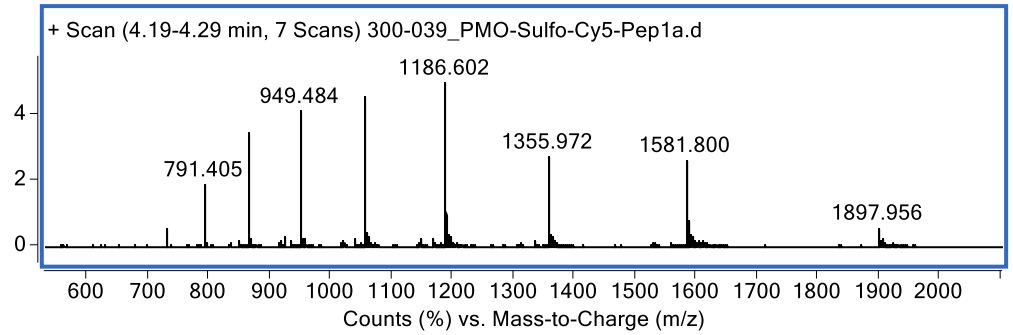


PMO-D-Bpep
Mass Expected: 9086.0 Da
Mass Observed: 9085.7 Da
Peptide sequence: Biotin-Azidolysine-G-G-K-G-G-w-r-ahx-r-r-βal-r-r-ahx-r-r-βal-r

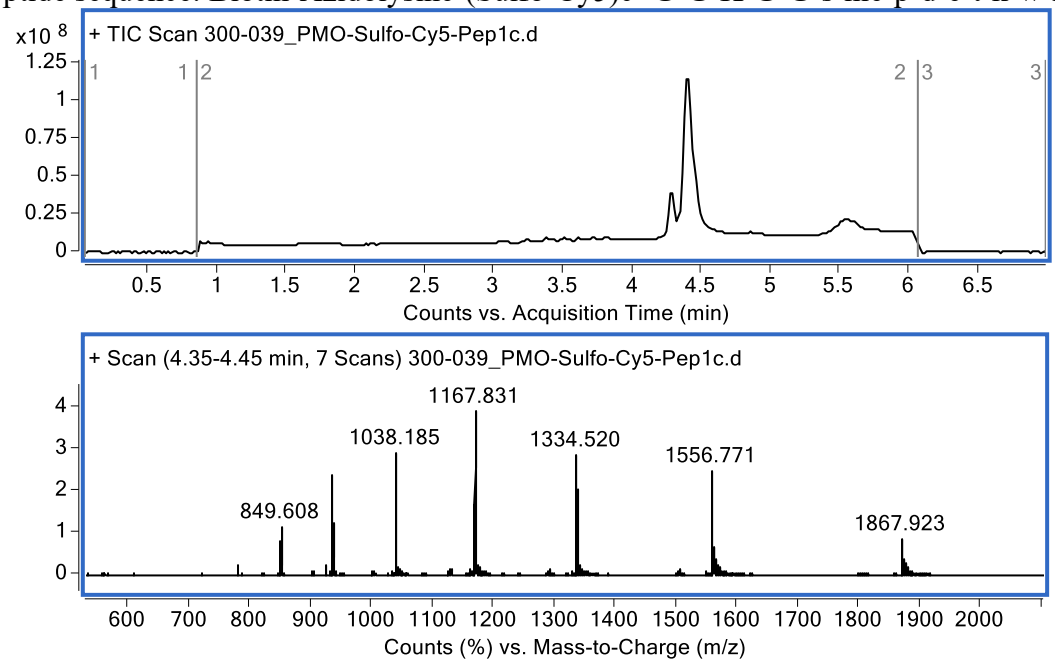


PMO-Sulfo-Cy5-Pep1a
Mass Expected: 9485.6
Mass Observed: 9485.7
Peptide sequence: Biotin-Azidolysine-(Sulfo-Cy5)c-G-G-K-G-G-s-βal-r-r-abu-dab-h-k-w-k-k

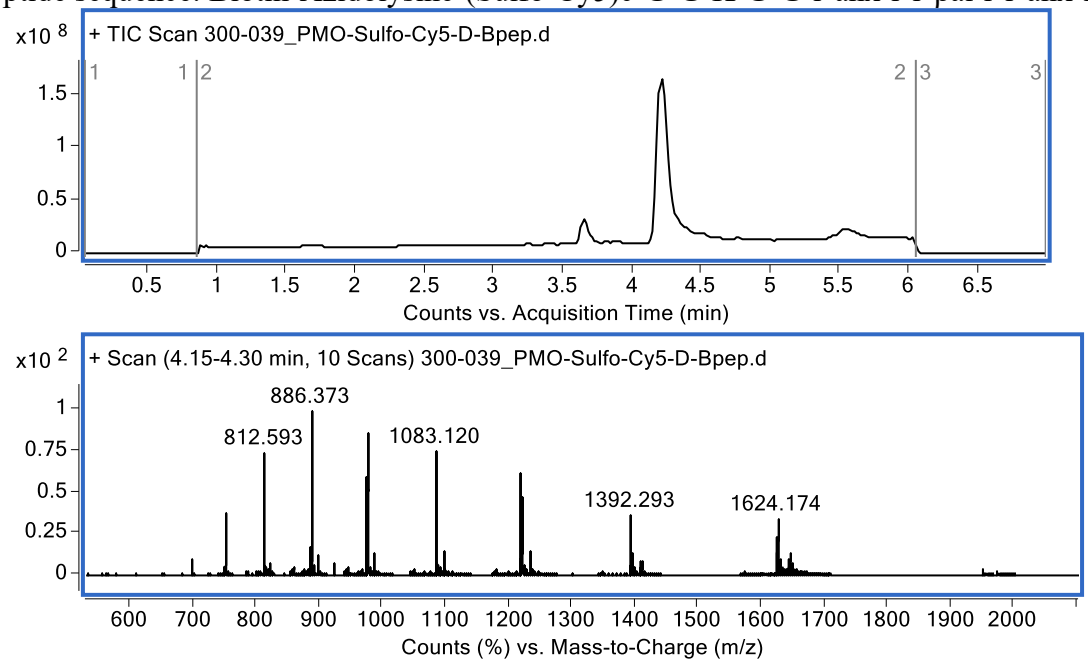




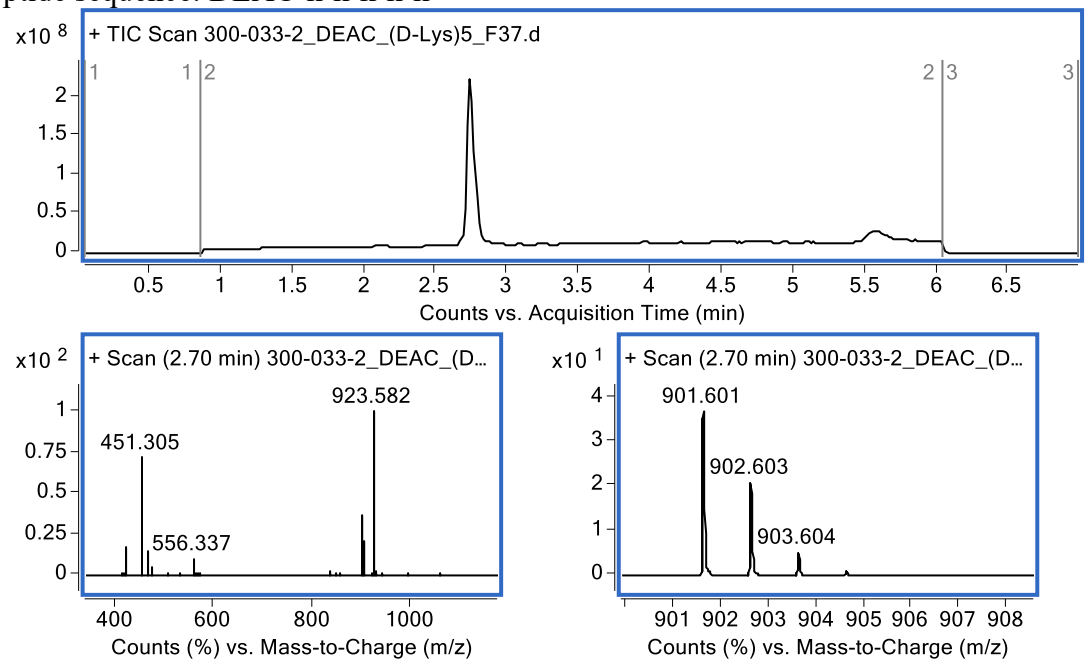
PMO-Sulfo-Cy5-Pep1c
Mass Expected:9335.3
Mass Observed: 9335.5
Peptide sequence: Biotin-Azidolysine-(Sulfo-Cy5)c- G-G-K-G-G-s-nle-p-d-e-t-k-w-k-k



PMO-Sulfo-Cy5-D-Bpep
Mass Expected: 9739.9
Mass Observed: 9739.8
Peptide sequence: Biotin-Azidolysine-(Sulfo-Cy5)c-G-G-K-G-G-r-ahx-r-r-βal-r-r-ahx-r-r-βal-r

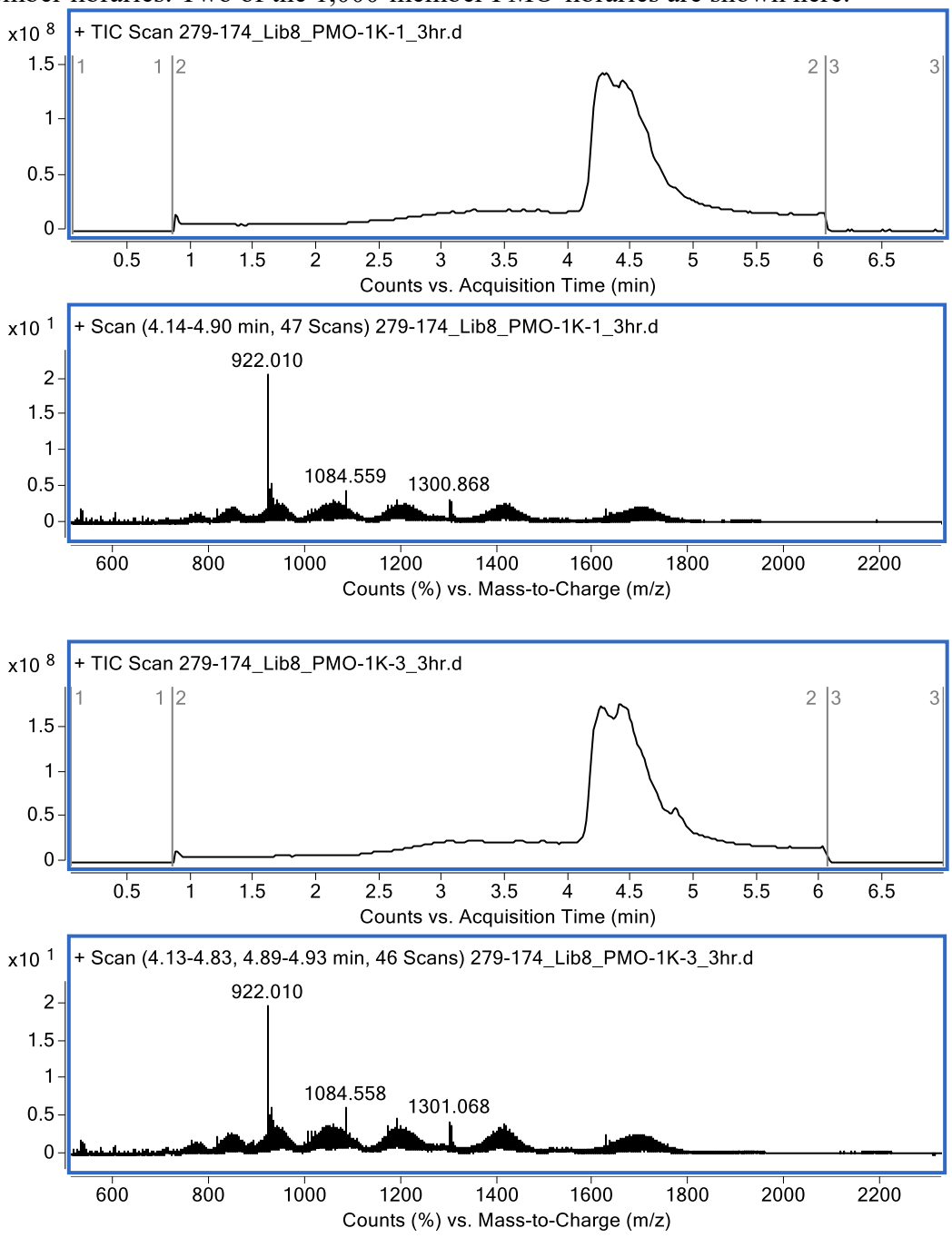


DEAC-k$_5$
Mass Expected: 900.591
Mass Observed: 900.593
Peptide sequence: DEAC-k-k-k-k-k



PMO-Library (1,000 member)

The libraries shown with greater than 1,000 members were produced by combining 1,000 member libraries. Two of the 1,000-member PMO-libraries are shown here.



5.9 Appendix III: Table of all peptides found in whole-library sequencing

All unique peptides found through sequencing a 500-member portion of Library 8 prior to appending the isoseramox and linker. D-residues are given their one-letter abbreviation, while non-canonical residues are given the three-letter abbreviation listed in Fig. 1.1.

Legend:

Monomer code	Monomer name
A	Beta-alanine
D	D-Asp
E	D-Glu
F	D-Phe
G	Gly
H	D-His
K	D-Lys
L	Norleucine
N	D-Asn
P	D-Pro
Q	D-Gln
R	D-Arg
S	D-Ser
T	D-Thr
V	D-Val
a	Dab
b	Pip
c	Gba
d	Amf
e	Nap
f	Dpa
g	Hle
h	Abu
i	Oxidized His
j	Oxidize Trp
k	Oxidized Nap
l	Oxidized Dpa
m	Oxidized Amf
X or Z	C-terminal amide

Scan	Peptide	ALC (%)	m/z	z	RT	Mass	ppm
7553	gbDeDfKWKKZ	99	373.7082	4	35.06	1490.807	-2.4
7566	gbDeDfKWKKZ	99	373.7083	4	35.11	1490.807	-2.2
7582	gbDeDfKWKKZ	99	373.7081	4	35.18	1490.807	-2.7
7598	gbDeDfKWKKZ	99	373.7082	4	35.24	1490.807	-2.4

5323	fFKFNTKWKKZ	98	362.9596	4	26.04	1447.813	-2.3
5420	eLhPLdKWKKZ	98	343.2176	4	26.42	1368.843	-1.6
5482	fNTRfNKWKKZ	98	380.7128	4	26.68	1518.825	-1.8
6373	eTFLGEKWKKZ	98	444.9195	3	30.31	1331.739	-1.7
6516	fFKeSDKWKKZ	98	495.9386	3	30.88	1484.797	-2
7595	GfAfFQKWKKZ	98	479.9317	3	35.23	1436.776	-1.6
8322	DTALffKWKKZ	98	478.9348	3	38.22	1433.786	-2.3
9306	ePLEegKWKKZ	98	362.9661	4	42.47	1447.838	-1.9
7543	gbDeDfKWKKZ	98	373.7083	4	35.02	1490.807	-2.2
5251	gFgNTQKWKKZ	97	444.942	3	25.75	1331.808	-2.9
5515	NLhLgQKWKKZ	97	423.6107	3	26.82	1267.813	-2.2
5991	DfdDfKKWKKZ	97	523.6174	3	28.76	1567.834	-2.2
6100	SaeDfQKWKKZ	97	360.4453	4	29.21	1437.756	-2.4
6159	FSfGgRKWKKZ	97	347.2096	4	29.45	1384.813	-2.9
6587	GfLHfGKWKKZ	97	466.9314	3	31.16	1397.776	-2.6
6771	SEfTeGKWKKZ	97	461.579	3	31.9	1381.718	-2.1
7083	gTEgfPKWKKZ	97	348.9647	4	33.15	1391.833	-2.4
7216	GFPLEfKWKKZ	97	452.2594	3	33.67	1353.76	-2.4
7805	egTQeEKWKKZ	97	489.9422	3	36.08	1466.807	-1.7
8065	egTQeEKWKKZ	97	367.7082	4	37.12	1466.807	-2.5
8706	DTALffKWKKZ	97	478.935	3	39.88	1433.786	-1.9
8757	fegQbFKWKKZ	97	512.9662	3	40.11	1535.881	-2.4
5336	fFKFNTKWKKZ	97	483.6109	3	26.09	1447.813	-1.4
5349	fFKFNTKWKKZ	97	483.6106	3	26.14	1447.813	-2
5365	fFKFNTKWKKZ	97	483.6103	3	26.2	1447.813	-2.6
5394	fFKFNTKWKKZ	97	483.6103	3	26.32	1447.813	-2.6
5404	fFKFNTKWKKZ	97	483.6104	3	26.36	1447.813	-2.4
5417	fFKFNTKWKKZ	97	483.6109	3	26.41	1447.813	-1.4
5528	fNTRfNKWKKZ	97	380.7125	4	26.87	1518.825	-2.4
5570	fNTRfNKWKKZ	97	380.7126	4	27.04	1518.825	-2.1
5998	DfdDfKKWKKZ	97	523.6172	3	28.79	1567.834	-2.6
6014	DfdDfKKWKKZ	97	523.6177	3	28.85	1567.834	-1.7
6027	DfdDfKKWKKZ	97	523.6177	3	28.91	1567.834	-1.7
6040	DfdDfKKWKKZ	97	523.6174	3	28.96	1567.834	-2.4
6053	DfdDfKKWKKZ	97	523.6176	3	29.01	1567.834	-2
6066	DfdDfKKWKKZ	97	523.6173	3	29.07	1567.834	-2.5
6079	DfdDfKKWKKZ	97	523.6174	3	29.12	1567.834	-2.2
6088	DfdDfKKWKKZ	97	523.6179	3	29.16	1567.834	-1.4
6103	SaeDfQKWKKZ	97	480.258	3	29.22	1437.756	-2.4
6115	SaeDfQKWKKZ	97	480.258	3	29.28	1437.756	-2.5
6143	SaeDfQKWKKZ	97	480.2581	3	29.39	1437.756	-2.2
6172	FSfGgRKWKKZ	97	347.2097	4	29.5	1384.813	-2.5
6185	FSfGgRKWKKZ	97	347.2097	4	29.56	1384.813	-2.5

6198	FSfGgRKWKKZ	97	347.2097	4	29.61	1384.813	-2.7
6532	fFKeSDKWKKZ	97	495.9384	3	30.94	1484.797	-2.4
6558	fFKeSDKWKKZ	97	495.9389	3	31.05	1484.797	-1.4
6593	GfLHfGKWKKZ	97	350.4503	4	31.19	1397.776	-2.8
6781	SEfTeGKWKKZ	97	461.579	3	31.94	1381.718	-2.1
6794	SEfTeGKWKKZ	97	461.579	3	31.99	1381.718	-2.2
6807	SEfTeGKWKKZ	97	461.579	3	32.04	1381.718	-2.1
6820	SEfTeGKWKKZ	97	461.5791	3	32.09	1381.718	-1.9
6833	SEfTeGKWKKZ	97	461.5791	3	32.14	1381.718	-2
6846	SEfTeGKWKKZ	97	461.579	3	32.19	1381.718	-2.3
6859	SEfTeGKWKKZ	97	461.579	3	32.24	1381.718	-2.2
6878	SEfTeGKWKKZ	97	461.579	3	32.32	1381.718	-2.3
6894	SEfTeGKWKKZ	97	461.5793	3	32.39	1381.718	-1.6
7585	gbDeDfKWKKZ	97	497.9421	3	35.19	1490.807	-2
7815	egTQeEKWKKZ	97	489.9422	3	36.12	1466.807	-1.8
7828	egTQeEKWKKZ	97	489.9422	3	36.17	1466.807	-1.8
7841	egTQeEKWKKZ	97	489.9419	3	36.23	1466.807	-2.5
7851	egTQeEKWKKZ	97	367.7082	4	36.27	1466.807	-2.4
7864	egTQeEKWKKZ	97	489.9418	3	36.32	1466.807	-2.6
7867	egTQeEKWKKZ	97	367.7081	4	36.33	1466.807	-2.9
7880	egTQeEKWKKZ	97	489.942	3	36.38	1466.807	-2.3
7883	egTQeEKWKKZ	97	367.7082	4	36.39	1466.807	-2.6
7893	egTQeEKWKKZ	97	489.942	3	36.43	1466.807	-2.2
7903	egTQeEKWKKZ	97	489.9421	3	36.47	1466.807	-2
7909	egTQeEKWKKZ	97	367.7083	4	36.5	1466.807	-2.4
7916	egTQeEKWKKZ	97	489.942	3	36.52	1466.807	-2.2
7922	egTQeEKWKKZ	97	367.7083	4	36.55	1466.807	-2.4
7926	egTQeEKWKKZ	97	489.9422	3	36.56	1466.807	-1.8
7935	egTQeEKWKKZ	97	367.7083	4	36.6	1466.807	-2.4
7942	egTQeEKWKKZ	97	489.942	3	36.63	1466.807	-2.1
7948	egTQeEKWKKZ	97	367.7082	4	36.65	1466.807	-2.6
7952	egTQeEKWKKZ	97	489.9421	3	36.66	1466.807	-1.9
7961	egTQeEKWKKZ	97	367.7083	4	36.7	1466.807	-2.3
7965	egTQeEKWKKZ	97	489.9421	3	36.71	1466.807	-2
7974	egTQeEKWKKZ	97	367.7083	4	36.75	1466.807	-2.4
7978	egTQeEKWKKZ	97	489.9424	3	36.77	1466.807	-1.3
7984	egTQeEKWKKZ	97	367.7085	4	36.79	1466.807	-1.8
7991	egTQeEKWKKZ	97	489.9423	3	36.82	1466.807	-1.5
7997	egTQeEKWKKZ	97	367.7084	4	36.85	1466.807	-2
8004	egTQeEKWKKZ	97	489.942	3	36.87	1466.807	-2.1
8020	egTQeEKWKKZ	97	489.9423	3	36.94	1466.807	-1.5
8033	egTQeEKWKKZ	97	489.942	3	36.99	1466.807	-2.3
8046	egTQeEKWKKZ	97	489.9418	3	37.04	1466.807	-2.6

8049	egTQeEKWKKZ	97	367.7082	4	37.05	1466.807	-2.6
8087	egTQeEKWKKZ	97	489.9422	3	37.21	1466.807	-1.8
8334	DTALffKWKKZ	97	478.9348	3	38.28	1433.786	-2.2
8346	DTALffKWKKZ	97	478.9348	3	38.33	1433.786	-2.3
8409	DTALffKWKKZ	97	359.4525	4	38.6	1433.786	-3.4
8469	DTALffKWKKZ	97	478.9349	3	38.87	1433.786	-2.1
8769	fegQbFKWKKZ	97	512.9662	3	40.17	1535.881	-2.4
8802	fegQbFKWKKZ	97	512.9664	3	40.31	1535.881	-2.1
8817	fegQbFKWKKZ	97	512.9662	3	40.38	1535.881	-2.4
8841	fegQbFKWKKZ	97	512.9663	3	40.48	1535.881	-2.3
8853	fegQbFKWKKZ	97	512.9662	3	40.54	1535.881	-2.5
8865	fegQbFKWKKZ	97	512.9659	3	40.59	1535.881	-3
8886	fegQbFKWKKZ	97	512.9662	3	40.68	1535.881	-2.5
8898	fegQbFKWKKZ	97	512.9667	3	40.74	1535.881	-1.5
8913	fegQbFKWKKZ	97	512.9667	3	40.8	1535.881	-1.6
9333	ePLEegKWKKZ	97	362.9661	4	42.58	1447.838	-1.8
5014	PKLDFLKWKKZ	96	434.6068	3	24.8	1300.802	-2.5
5186	gaFQRfKWKKZ	96	368.2266	4	25.49	1468.882	-3.1
5914	ffKHgTKWKKZ	96	382.7294	4	28.44	1526.891	-1.9
6438	SNfDfbKWKKZ	96	492.9297	3	30.57	1475.772	-2.9
6628	TNFLGfKWKKZ	96	336.6952	4	31.34	1342.755	-2.4
5261	gFgNTQKWKKZ	96	333.9584	4	25.79	1331.808	-2.5
5268	gFgNTQKWKKZ	96	444.942	3	25.81	1331.808	-2.7
5281	FggNTQKWKKZ	96	444.9425	3	25.86	1331.808	-1.6
5284	gFgNTQKWKKZ	96	333.9586	4	25.88	1331.808	-1.9
5333	fFKFNTKWKKZ	96	362.9598	4	26.07	1447.813	-1.7
5375	fFKFNTKWKKZ	96	362.9595	4	26.24	1447.813	-2.6
5388	fFKFNTKWKKZ	96	362.9596	4	26.29	1447.813	-2.5
5401	fFKFNTKWKKZ	96	362.9596	4	26.35	1447.813	-2.4
5440	fFKFNTKWKKZ	96	362.9598	4	26.51	1447.813	-1.9
5499	fNTRfNKWKKZ	96	380.7127	4	26.75	1518.825	-1.9
5541	fNTRfNKWKKZ	96	380.7127	4	26.92	1518.825	-2
5554	fNTRfNKWKKZ	96	380.7127	4	26.98	1518.825	-2
5993	DfdDfKKWKKZ	96	523.6174	3	28.77	1567.834	-2.2
6029	DfdDfKKWKKZ	96	523.6177	3	28.91	1567.834	-1.7
6055	DfdDfKKWKKZ	96	523.6176	3	29.02	1567.834	-2
6081	DfdDfKKWKKZ	96	523.6174	3	29.13	1567.834	-2.2
6188	FSfGgRKWKKZ	96	462.6107	3	29.57	1384.813	-2.2
6211	FSfGgRKWKKZ	96	347.2097	4	29.66	1384.813	-2.4
6529	fFKeSDKWKKZ	96	372.2056	4	30.93	1484.797	-2.5
6542	fFKeSDKWKKZ	96	372.2056	4	30.98	1484.797	-2.3
6638	TNFLGfKWKKZ	96	448.5913	3	31.37	1342.755	-2.1
6641	TNFLGfKWKKZ	96	336.6952	4	31.39	1342.755	-2.4

6654	TNFLGfKWKKZ	96	336.6952	4	31.44	1342.755	-2.5
6686	TNFLGfKWKKZ	96	336.6952	4	31.57	1342.755	-2.3
7106	gTEgfPKWKKZ	96	464.9504	3	33.24	1391.833	-2.6
7154	GFPLEfKWKKZ	96	452.2594	3	33.43	1353.76	-2.6
7158	gTEgfPKWKKZ	96	464.9506	3	33.44	1391.833	-2.1
7164	GFPLEfKWKKZ	96	452.2596	3	33.47	1353.76	-2.2
7203	GFPLEfKWKKZ	96	452.2595	3	33.62	1353.76	-2.2
7226	gTEgfPKWKKZ	96	464.9505	3	33.71	1391.833	-2.5
7304	GFPLEfKWKKZ	96	452.2596	3	34.02	1353.76	-2.2
7854	egTQeEKWKKZ	96	489.942	3	36.28	1466.807	-2.2
7896	egTQeEKWKKZ	96	367.7082	4	36.45	1466.807	-2.5
8010	egTQeEKWKKZ	96	367.7082	4	36.9	1466.807	-2.4
8036	egTQeEKWKKZ	96	367.7082	4	37	1466.807	-2.5
8059	egTQeEKWKKZ	96	489.942	3	37.1	1466.807	-2.1
8072	egTQeEKWKKZ	96	489.942	3	37.15	1466.807	-2.1
8424	DTALffKWKKZ	96	359.4531	4	38.67	1433.786	-1.8
8589	DTALffKWKKZ	96	478.9347	3	39.38	1433.786	-2.5
8694	DTALffKWKKZ	96	478.9349	3	39.82	1433.786	-2.2
8874	fegQbFKWKKZ	96	512.9663	3	40.63	1535.881	-2.3
9189	ePLEegKWKKZ	96	362.9659	4	41.97	1447.838	-2.2
9213	ePLEegKWKKZ	96	362.966	4	42.07	1447.838	-2.1
9225	ePLEegKWKKZ	96	362.9659	4	42.12	1447.838	-2.2
9258	ePLEegKWKKZ	96	362.9659	4	42.26	1447.838	-2.2
3470	KFSLEaKWKKZ	95	323.9509	4	18.33	1291.776	-1.6
4126	AKQfFNKWKKZ	95	467.2703	3	21.15	1398.793	-2.4
4386	bgRdGfKWKKZ	95	364.2281	4	22.25	1452.887	-2.6
5644	ffKHgTKWKKZ	95	382.7293	4	27.35	1526.891	-2.1
5941	fGENDLKWKKZ	95	447.2422	3	28.55	1338.709	-2.7
6127	FfQEhLKWKKZ	95	471.9383	3	29.32	1412.797	-2.6
6539	eAFfdEKWKKZ	95	511.2785	3	30.97	1530.817	-2.4
8601	DTALefKWKKZ	95	470.2629	3	39.43	1407.77	-2.4
4150	AKQfFNKWKKZ	95	467.2706	3	21.26	1398.793	-1.7
4396	bgRdGfKWKKZ	95	364.2281	4	22.29	1452.887	-2.6
4406	bgRdGfKWKKZ	95	364.2282	4	22.33	1452.887	-2.3
4416	bgRdGfKWKKZ	95	364.2282	4	22.36	1452.887	-2.1
4429	bgRdGfKWKKZ	95	364.2284	4	22.41	1452.887	-1.6
4442	bgRdGfKWKKZ	95	364.2282	4	22.47	1452.887	-2.3
4455	bgRdGfKWKKZ	95	364.2283	4	22.52	1452.887	-2.1
4468	bgRdGfKWKKZ	95	364.2281	4	22.57	1452.887	-2.4
4481	bgRdGfKWKKZ	95	364.2281	4	22.62	1452.887	-2.6
4494	bgRdGfKWKKZ	95	364.2281	4	22.67	1452.887	-2.5
4520	bgRdGfKWKKZ	95	364.2281	4	22.78	1452.887	-2.5
4533	bgRdGfKWKKZ	95	364.2282	4	22.83	1452.887	-2.1

4546	bgRdGfKWKKZ	95	364.2283	4	22.88	1452.887	-2.1
4565	bgRdGfKWKKZ	95	364.2283	4	22.96	1452.887	-2.1
4575	bgRdGfKWKKZ	95	364.2282	4	23.01	1452.887	-2.2
4594	bgRdGfKWKKZ	95	364.2282	4	23.08	1452.887	-2.3
5258	gFgNTQKWKKZ	95	444.9422	3	25.78	1331.808	-2.3
5294	gFgNTQKWKKZ	95	444.9422	3	25.92	1331.808	-2.4
5307	gFgNTQKWKKZ	95	444.9422	3	25.97	1331.808	-2.4
5346	fFKFNTKWKKZ	95	362.9598	4	26.12	1447.813	-1.9
5359	fFKFNTKWKKZ	95	362.9596	4	26.18	1447.813	-2.3
5414	fFKFNTKWKKZ	95	362.9598	4	26.4	1447.813	-1.8
5427	fFKFNTKWKKZ	95	362.96	4	26.45	1447.813	-1.4
5486	fNTRfNKWKKZ	95	380.7127	4	26.7	1518.825	-2
5502	fNTRfNKWKKZ	95	507.2813	3	26.76	1518.825	-1.7
5512	fNTRfNKWKKZ	95	380.7126	4	26.8	1518.825	-2.1
5978	DfdDfKKWKKZ	95	392.9649	4	28.7	1567.834	-2.2
6000	DfdDfKKWKKZ	95	523.6172	3	28.8	1567.834	-2.6
6001	DfdDfKKWKKZ	95	392.9647	4	28.8	1567.834	-2.8
6011	DfdDfKKWKKZ	95	392.9652	4	28.84	1567.834	-1.5
6016	DfdDfKKWKKZ	95	523.6177	3	28.86	1567.834	-1.7
6037	DfdDfKKWKKZ	95	392.9648	4	28.95	1567.834	-2.6
6042	DfdDfKKWKKZ	95	523.6174	3	28.97	1567.834	-2.4
6063	DfdDfKKWKKZ	95	392.9647	4	29.06	1567.834	-2.7
6068	DfdDfKKWKKZ	95	523.6173	3	29.07	1567.834	-2.5
6090	DfdDfKKWKKZ	95	523.6179	3	29.17	1567.834	-1.4
6124	SaeDfQKWKKZ	95	360.445	4	29.31	1437.756	-3.3
6130	SaeDfQKWKKZ	95	480.258	3	29.34	1437.756	-2.3
6140	SaeDfQKWKKZ	95	360.4452	4	29.38	1437.756	-2.7
6153	SaeDfQKWKKZ	95	480.2581	3	29.43	1437.756	-2.2
6375	eTFLGEKWKKZ	95	444.9195	3	30.32	1331.739	-1.7
6518	fFKeSDKWKKZ	95	495.9386	3	30.89	1484.797	-2
7096	gTEgfPKWKKZ	95	348.9647	4	33.2	1391.833	-2.3
7145	gTEgfPKWKKZ	95	464.9505	3	33.39	1391.833	-2.5
7187	FGPLEfKWKKZ	95	452.2595	3	33.56	1353.76	-2.2
7229	FGPLEfKWKKZ	95	452.2595	3	33.72	1353.76	-2.4
7252	gTEgfPKWKKZ	95	464.9501	3	33.81	1391.833	-3.2
7281	GFPLEfKWKKZ	95	452.2596	3	33.93	1353.76	-2.2
7346	FGPLEfKWKKZ	95	452.2596	3	34.19	1353.76	-2.2
7572	fGAfFQKWKKZ	95	479.9315	3	35.14	1436.776	-2.2
7587	gbDeDfKWKKZ	95	497.9421	3	35.2	1490.807	-2
8023	egTQeEKWKKZ	95	367.7084	4	36.95	1466.807	-2
8160	egTQeEKWKKZ	95	489.942	3	37.52	1466.807	-2.1
8184	egTQeEKWKKZ	95	489.9419	3	37.62	1466.807	-2.3
8244	egTQeEKWKKZ	95	489.9423	3	37.88	1466.807	-1.6

8376	DTALffKWKKZ	95	359.4526	4	38.46	1433.786	-3.1
8394	DTALffKWKKZ	95	478.9345	3	38.53	1433.786	-2.9
8434	DTALffKWKKZ	95	359.4531	4	38.71	1433.786	-1.7
8613	DTALefKWKKZ	95	470.2629	3	39.48	1407.77	-2.5
8685	DTALffKWKKZ	95	478.9349	3	39.79	1433.786	-2.2
8745	DTALffKWKKZ	95	478.9352	3	40.06	1433.786	-1.6
8804	fegQbFKWKKZ	95	512.9664	3	40.32	1535.881	-2.1
9270	ePLEegKWKKZ	95	362.9661	4	42.31	1447.838	-1.7
9282	ePLEegKWKKZ	95	362.9661	4	42.36	1447.838	-1.9
4659	bgRdGfKWKKZ	94	364.2283	4	23.35	1452.887	-2
4965	FKeNFKKWKKZ	94	363.2178	4	24.6	1448.845	-1.6
5027	NFDSfdKWKKZ	94	484.2582	3	24.85	1449.756	-2
5595	PSFfEAKWKKZ	94	336.4374	4	27.15	1341.723	-2.2
8292	PLLEegKWKKZ	94	455.6191	3	38.09	1363.838	-1.8
8483	hfFgSgKWKKZ	94	346.9674	4	38.93	1383.843	-2.1
9801	GfLfGKWWKKZ	94	483.2719	3	44.64	1446.796	-1.8
3446	KFSLEaKWKKZ	94	323.9506	4	18.22	1291.776	-2.3
3452	KFSLEaKWKKZ	94	323.9507	4	18.25	1291.776	-1.9
3482	KFSLEaKWKKZ	94	323.9506	4	18.38	1291.776	-2.5
4507	bgRdGfKWKKZ	94	364.2282	4	22.72	1452.887	-2.2
4607	bgRdGfKWKKZ	94	364.2281	4	23.14	1452.887	-2.6
4975	PKLDFLKWKKZ	94	326.207	4	24.64	1300.802	-2.5
4982	PKLDFLKWKKZ	94	326.2068	4	24.67	1300.802	-2.9
5008	PKLDFLKWKKZ	94	326.207	4	24.77	1300.802	-2.4
5016	PKLDFLKWKKZ	94	434.6068	3	24.8	1300.802	-2.5
5021	PKLDFLKWKKZ	94	326.2069	4	24.82	1300.802	-2.6
5037	PKLDFLKWKKZ	94	326.2072	4	24.89	1300.802	-1.9
5271	gFgNTQKWKKZ	94	333.9584	4	25.83	1331.808	-2.4
5300	gFgNTQKWKKZ	94	333.9586	4	25.94	1331.808	-2.1
5338	fFKFNTKWKKZ	94	483.6109	3	26.09	1447.813	-1.4
5351	fFKFNTKWKKZ	94	483.6106	3	26.15	1447.813	-2
5367	fFKFNTKWKKZ	94	483.6103	3	26.21	1447.813	-2.6
5396	fFKFNTKWKKZ	94	483.6103	3	26.32	1447.813	-2.6
5406	fFKFNTKWKKZ	94	483.6104	3	26.36	1447.813	-2.4
5419	fFKFNTKWKKZ	94	483.6109	3	26.42	1447.813	-1.4
5618	PSFfEAKWKKZ	94	336.4376	4	27.24	1341.723	-1.6
5660	ffKHgTKWKKZ	94	382.7293	4	27.41	1526.891	-2.1
5673	ffKHgTKWKKZ	94	382.7292	4	27.46	1526.891	-2.4
5689	ffKHgTKWKKZ	94	382.7295	4	27.53	1526.891	-1.7
5696	ffKHgTKWKKZ	94	382.7293	4	27.56	1526.891	-2
5988	DfdDfKKWKKZ	94	392.9648	4	28.75	1567.834	-2.5
6021	DfdDfKKWKKZ	94	392.9651	4	28.88	1567.834	-1.6
6076	DfdDfKKWKKZ	94	392.9649	4	29.11	1567.834	-2.2

6091	DfdDfKKWKKZ	94	392.9651	4	29.17	1567.834	-1.8
6117	SaeDfQKWKKZ	94	480.258	3	29.28	1437.756	-2.5
6145	SaeDfQKWKKZ	94	480.2581	3	29.4	1437.756	-2.2
6513	fFKeSDKWKKZ	94	372.2057	4	30.87	1484.797	-2
6534	fFKeSDKWKKZ	94	495.9384	3	30.95	1484.797	-2.4
6560	fFKeSDKWKKZ	94	495.9389	3	31.05	1484.797	-1.4
6565	eAFfdEKWKKZ	94	511.2789	3	31.08	1530.817	-1.7
6581	eAFfdEKWKKZ	94	511.2785	3	31.14	1530.817	-2.5
6589	GfLHfGKWKKZ	94	466.9314	3	31.18	1397.776	-2.6
6677	TNFLGfKWKKZ	94	448.5912	3	31.53	1342.755	-2.2
6722	TNFLGfKWKKZ	94	448.5912	3	31.71	1342.755	-2.3
7050	gTEgfPKWKKZ	94	348.9648	4	33.02	1391.833	-2.2
7057	gTEgfPKWKKZ	94	464.9505	3	33.05	1391.833	-2.4
7076	gTEgfPKWKKZ	94	348.9647	4	33.12	1391.833	-2.5
7167	gTEgfPKWKKZ	94	348.9647	4	33.48	1391.833	-2.4
7197	gTEgfPKWKKZ	94	464.9505	3	33.6	1391.833	-2.5
7255	FGPLEfKWKKZ	94	452.2591	3	33.83	1353.76	-3.3
7271	FGPLEfKWKKZ	94	452.2593	3	33.89	1353.76	-2.6
7330	FGPLEfKWKKZ	94	452.2595	3	34.13	1353.76	-2.2
7359	FGPLEfKWKKZ	94	452.2594	3	34.25	1353.76	-2.6
7636	fGAfFQKWKKZ	94	479.9315	3	35.4	1436.776	-2
7779	GfAfFQKWKKZ	94	479.9315	3	35.97	1436.776	-2.2
8304	PLLEegKWKKZ	94	455.619	3	38.15	1363.838	-2
8310	DTALffKWKKZ	94	478.9347	3	38.17	1433.786	-2.5
8511	DTALffKWKKZ	94	478.9348	3	39.04	1433.786	-2.3
8523	DTALffKWKKZ	94	478.9349	3	39.09	1433.786	-2.1
8571	DTALffKWKKZ	94	478.9347	3	39.3	1433.786	-2.5
8625	DTALefKWKKZ	94	470.2631	3	39.53	1407.77	-2.1
8771	fegQbFKWKKZ	94	512.9662	3	40.17	1535.881	-2.4
8819	fegQbFKWKKZ	94	512.9662	3	40.39	1535.881	-2.4
8889	fegQbFKWKKZ	94	384.9765	4	40.7	1535.881	-2.4
8900	fegQbFKWKKZ	94	512.9667	3	40.74	1535.881	-1.5
9174	ePLEegKWKKZ	94	362.9661	4	41.9	1447.838	-1.9
9471	ePLEegKWKKZ	94	483.6192	3	43.18	1447.838	-1.5
9807	GfLfGKWWKKZ	94	483.2719	3	44.66	1446.796	-1.8
9819	GfLfGKWWKKZ	94	483.2717	3	44.72	1446.796	-2.2
9843	GfLfGKWWKKZ	94	483.2719	3	44.82	1446.796	-1.7
4147	AgTLGNKWKKZ	93	293.6877	4	21.24	1170.724	-1.8
4246	KfGQdLKWKKZ	93	354.2176	4	21.67	1412.845	-2.2
4445	fhNRLTKWKKZ	93	345.9611	4	22.48	1379.819	-2.8
4766	EAfaLGKWKKZ	93	427.9195	3	23.8	1280.739	-2
4793	TRfQAFKWKKZ	93	354.4572	4	23.91	1413.803	-2.4
5134	TKfhKfKWKKZ	93	492.9661	3	25.28	1475.881	-2.8

6735	afNfgAKWKKZ	93	362.4647	4	31.76	1445.834	-2.6
8514	EgTeQeKWKKZ	93	489.9418	3	39.06	1466.807	-2.5
9549	ePLEegKWKKZ	93	483.6192	3	43.52	1447.838	-1.5
4135	AKQfFNKWKKZ	93	467.2702	3	21.19	1398.793	-2.7
4156	AgTLGNKWKKZ	93	293.6875	4	21.28	1170.724	-2.2
4192	AgTLGNKWKKZ	93	293.6874	4	21.43	1170.724	-2.5
4195	AgTLGNKWKKZ	93	391.2474	3	21.45	1170.724	-2.8
4497	fhNRLTKWKKZ	93	345.9611	4	22.69	1379.819	-2.8
4781	hDfaLGKWKKZ	93	427.9194	3	23.86	1280.739	-2.2
4978	FKeNFKKWKKZ	93	363.2176	4	24.66	1448.845	-2.3
4995	PKLDFLKWKKZ	93	326.207	4	24.72	1300.802	-2.4
5001	NFDSfdKWKKZ	93	363.4454	4	24.75	1449.756	-2.2
5011	NFDSfdKWKKZ	93	484.2582	3	24.79	1449.756	-2
5531	fNTRfNKWKKZ	93	507.2809	3	26.88	1518.825	-2.5
5608	PSFfEAKWKKZ	93	336.4376	4	27.2	1341.723	-1.7
6109	SaeDfQKWKKZ	93	360.4451	4	29.25	1437.756	-2.9
6137	FfQEhLKWKKZ	93	471.9384	3	29.37	1412.797	-2.5
6150	FfQEhLKWKKZ	93	471.9384	3	29.42	1412.797	-2.6
6165	FSfGgRKWKKZ	93	462.6103	3	29.48	1384.813	-3
6168	FfQEhLKWKKZ	93	471.938	3	29.49	1412.797	-3.4
6175	FSfGgRKWKKZ	93	462.6104	3	29.52	1384.813	-2.9
6227	FSfGgRKWKKZ	93	347.2098	4	29.73	1384.813	-2.2
6440	SNfDfbKWKKZ	93	492.9297	3	30.58	1475.772	-2.9
6536	eAFfdEKWKKZ	93	383.7108	4	30.96	1530.817	-2.3
6607	GfLHfGKWKKZ	93	466.9313	3	31.25	1397.776	-2.8
6625	TNFLGfKWKKZ	93	448.5913	3	31.32	1342.755	-2.1
6651	TNFLGfKWKKZ	93	448.5912	3	31.43	1342.755	-2.3
6664	TNFLGfKWKKZ	93	448.5914	3	31.48	1342.755	-1.9
6690	TNFLGfKWKKZ	93	448.5912	3	31.58	1342.755	-2.3
6706	TNFLGfKWKKZ	93	448.5911	3	31.65	1342.755	-2.6
6773	SEfTeGKWKKZ	93	461.579	3	31.91	1381.718	-2.1
6783	SEfTeGKWKKZ	93	461.579	3	31.95	1381.718	-2.1
6796	SEfTeGKWKKZ	93	461.579	3	32	1381.718	-2.2
6809	SEfTeGKWKKZ	93	461.579	3	32.05	1381.718	-2.1
6822	SEfTeGKWKKZ	93	461.5791	3	32.1	1381.718	-1.9
6835	SEfTeGKWKKZ	93	461.5791	3	32.15	1381.718	-2
6848	SEfTeGKWKKZ	93	461.579	3	32.2	1381.718	-2.3
6861	SEfTeGKWKKZ	93	461.579	3	32.26	1381.718	-2.2
6880	SEfTeGKWKKZ	93	461.579	3	32.33	1381.718	-2.3
6896	SEfTeGKWKKZ	93	461.5793	3	32.39	1381.718	-1.6
7108	gTEgfPKWKKZ	93	464.9504	3	33.25	1391.833	-2.6
7109	gTEgfPKWKKZ	93	348.9646	4	33.26	1391.833	-2.8
7138	gTEgfPKWKKZ	93	348.9646	4	33.37	1391.833	-2.6

7151	gTEgfPKWKKZ	93	348.9646	4	33.42	1391.833	-2.8
7171	gTEgfPKWKKZ	93	464.9505	3	33.49	1391.833	-2.4
7241	gTEgfPKWKKZ	93	464.9507	3	33.77	1391.833	-2.1
7265	gTEgfPKWKKZ	93	464.9505	3	33.87	1391.833	-2.5
7320	FGPLEfKWKKZ	93	452.2596	3	34.09	1353.76	-2.1
7649	fGAfFQKWKKZ	93	479.9317	3	35.45	1436.776	-1.7
7766	fGAfFQKWKKZ	93	479.9316	3	35.92	1436.776	-1.8
7807	egTQeEKWKKZ	93	489.9422	3	36.09	1466.807	-1.7
7980	egTQeEKWKKZ	93	489.9424	3	36.78	1466.807	-1.3
8022	egTQeEKWKKZ	93	489.9423	3	36.94	1466.807	-1.5
8301	DTALffKWKKZ	93	478.9351	3	38.13	1433.786	-1.8
8358	DTALffKWKKZ	93	478.9349	3	38.38	1433.786	-2.1
8370	DTALffKWKKZ	93	478.9346	3	38.43	1433.786	-2.7
8382	DTALffKWKKZ	93	478.9346	3	38.48	1433.786	-2.7
8428	DTALffKWKKZ	93	478.9348	3	38.68	1433.786	-2.3
8450	DTALffKWKKZ	93	478.9348	3	38.78	1433.786	-2.2
8457	DTALffKWKKZ	93	478.9348	3	38.81	1433.786	-2.3
8480	DTALffKWKKZ	93	478.9348	3	38.91	1433.786	-2.3
8499	DTALffKWKKZ	93	478.9349	3	38.99	1433.786	-2.2
8535	DTALffKWKKZ	93	478.9348	3	39.14	1433.786	-2.3
8547	DTALffKWKKZ	93	478.935	3	39.19	1433.786	-1.9
8559	DTALffKWKKZ	93	478.9347	3	39.25	1433.786	-2.5
8759	fegQbFKWKKZ	93	512.9662	3	40.12	1535.881	-2.4
8793	fegQbFKWKKZ	93	384.9766	4	40.27	1535.881	-2.3
8843	fegQbFKWKKZ	93	512.9663	3	40.49	1535.881	-2.3
8855	fegQbFKWKKZ	93	512.9662	3	40.54	1535.881	-2.5
8867	fegQbFKWKKZ	93	512.9659	3	40.6	1535.881	-3
8888	fegQbFKWKKZ	93	512.9662	3	40.69	1535.881	-2.5
8901	fegQbFKWKKZ	93	384.9769	4	40.75	1535.881	-1.5
8915	fegQbFKWKKZ	93	512.9667	3	40.81	1535.881	-1.6
9150	ePLEegKWKKZ	93	483.6188	3	41.8	1447.838	-2.3
9159	ePLEegKWKKZ	93	483.6191	3	41.83	1447.838	-1.6
9171	ePLEegKWKKZ	93	483.6188	3	41.88	1447.838	-2.2
9183	ePLEegKWKKZ	93	483.6188	3	41.94	1447.838	-2.3
9207	ePLEegKWKKZ	93	483.6188	3	42.04	1447.838	-2.3
9219	ePLEegKWKKZ	93	483.6188	3	42.09	1447.838	-2.3
9231	ePLEegKWKKZ	93	483.619	3	42.14	1447.838	-1.8
9234	ePLEegKWKKZ	93	362.966	4	42.16	1447.838	-2
9243	ePLEegKWKKZ	93	483.6187	3	42.19	1447.838	-2.5
9255	ePLEegKWKKZ	93	483.6188	3	42.24	1447.838	-2.3
9279	ePLEegKWKKZ	93	483.6191	3	42.35	1447.838	-1.6
9330	ePLEegKWKKZ	93	483.6191	3	42.57	1447.838	-1.6
9363	ePLEegKWKKZ	93	483.6191	3	42.71	1447.838	-1.7

9375	ePLEegKWKKZ	93	483.619	3	42.76	1447.838	-1.8
9387	ePLEegKWKKZ	93	483.6191	3	42.81	1447.838	-1.7
9402	ePLEegKWKKZ	93	483.6191	3	42.88	1447.838	-1.6
9411	ePLEegKWKKZ	93	483.619	3	42.92	1447.838	-1.8
9423	ePLEegKWKKZ	93	483.6187	3	42.97	1447.838	-2.5
9435	ePLEegKWKKZ	93	483.6191	3	43.02	1447.838	-1.7
9450	ePLEegKWKKZ	93	483.619	3	43.09	1447.838	-1.8
9459	ePLEegKWKKZ	93	483.6194	3	43.12	1447.838	-1
9483	ePLEegKWKKZ	93	483.6192	3	43.23	1447.838	-1.4
9501	ePLEegKWKKZ	93	483.6192	3	43.31	1447.838	-1.5
9513	ePLEegKWKKZ	93	483.6193	3	43.36	1447.838	-1.3
9522	ePLEegKWKKZ	93	483.6192	3	43.4	1447.838	-1.5
9534	ePLEegKWKKZ	93	483.6192	3	43.46	1447.838	-1.5
9561	ePLEegKWKKZ	93	483.6192	3	43.57	1447.838	-1.4
9573	ePLEegKWKKZ	93	483.6191	3	43.63	1447.838	-1.6
9858	GfLfGKWWKKZ	93	483.272	3	44.89	1446.796	-1.6
5193	gKfGTPKWKKZ	92	441.2754	3	25.52	1320.807	-2.2
5430	QeLbdgKWKKZ	92	364.7295	4	26.47	1454.891	-1.7
8583	NGfTeGKWKKZ	92	446.5716	3	39.35	1336.708	-11.3
4168	AgTLGNKWKKZ	92	293.6875	4	21.33	1170.724	-2.3
4171	AgTLGNKWKKZ	92	391.2475	3	21.35	1170.724	-2.5
4180	AgTLGNKWKKZ	92	293.6876	4	21.38	1170.724	-2
4751	hDfaLGKWKKZ	92	427.9192	3	23.73	1280.739	-2.7
4796	hDfaLGKWKKZ	92	427.9194	3	23.92	1280.739	-2.2
5283	gFgNTQKWKKZ	92	444.9425	3	25.87	1331.808	-1.6
5518	fNTRfNKWKKZ	92	507.2813	3	26.83	1518.825	-1.7
5611	PSFfEAKWKKZ	92	448.2477	3	27.21	1341.723	-1.6
5621	PSFfEAKWKKZ	92	448.2477	3	27.25	1341.723	-1.5
5646	ffKHgTKWKKZ	92	382.7293	4	27.35	1526.891	-2.1
5709	ffKHgTKWKKZ	92	382.7293	4	27.61	1526.891	-2.2
6050	DfdDfKKWKKZ	92	392.9648	4	29	1567.834	-2.5
6105	SaeDfQKWKKZ	92	480.258	3	29.23	1437.756	-2.4
6178	FfQEhLKWKKZ	92	354.2057	4	29.53	1412.797	-2.4
6519	eAFfdEKWKKZ	92	383.7109	4	30.89	1530.817	-1.9
6523	eAFfdEKWKKZ	92	383.7107	4	30.91	1530.817	-2.4
6526	eAFfdEKWKKZ	92	511.2785	3	30.92	1530.817	-2.4
6549	eAFfdEKWKKZ	92	383.711	4	31.01	1530.817	-1.5
6575	eAFfdEKWKKZ	92	383.7107	4	31.11	1530.817	-2.4
6631	afNfgAKWKKZ	92	362.4647	4	31.35	1445.834	-2.7
6640	TNFLGfKWKKZ	92	448.5913	3	31.38	1342.755	-2.1
6679	TNFLGfKWKKZ	92	448.5912	3	31.54	1342.755	-2.2
6724	TNFLGfKWKKZ	92	448.5912	3	31.72	1342.755	-2.3
6891	afNfgAKWKKZ	92	362.4649	4	32.37	1445.834	-2.1

7069	gTEgfPKWKKZ	92	464.9505	3	33.09	1391.833	-2.5
7095	gTEgfPKWKKZ	92	464.9506	3	33.2	1391.833	-2.3
7121	gTEgfPKWKKZ	92	464.9504	3	33.3	1391.833	-2.6
7147	gTEgfPKWKKZ	92	464.9505	3	33.4	1391.833	-2.5
7160	gTEgfPKWKKZ	92	464.9506	3	33.45	1391.833	-2.1
7177	GFPLEfKWKKZ	92	452.2594	3	33.52	1353.76	-2.4
7212	gTEgfPKWKKZ	92	464.9505	3	33.66	1391.833	-2.4
7218	GFPLEfKWKKZ	92	452.2594	3	33.68	1353.76	-2.4
7228	gTEgfPKWKKZ	92	464.9505	3	33.72	1391.833	-2.5
7555	gbDeDfKWKKZ	92	373.7082	4	35.07	1490.807	-2.4
7568	gbDeDfKWKKZ	92	373.7083	4	35.12	1490.807	-2.2
7817	egTQeEKWKKZ	92	489.9422	3	36.13	1466.807	-1.8
7830	egTQeEKWKKZ	92	489.9422	3	36.18	1466.807	-1.8
7866	egTQeEKWKKZ	92	489.9418	3	36.32	1466.807	-2.6
7905	egTQeEKWKKZ	92	489.9421	3	36.48	1466.807	-2
7918	egTQeEKWKKZ	92	489.942	3	36.53	1466.807	-2.2
7944	egTQeEKWKKZ	92	489.942	3	36.63	1466.807	-2.1
7954	egTQeEKWKKZ	92	489.9421	3	36.67	1466.807	-1.9
7967	egTQeEKWKKZ	92	489.9421	3	36.73	1466.807	-2
7993	egTQeEKWKKZ	92	489.9423	3	36.83	1466.807	-1.5
8006	egTQeEKWKKZ	92	489.942	3	36.88	1466.807	-2.1
8035	egTQeEKWKKZ	92	489.942	3	37	1466.807	-2.3
8089	egTQeEKWKKZ	92	489.9422	3	37.22	1466.807	-1.8
8406	DTALffKWKKZ	92	478.9348	3	38.58	1433.786	-2.3
8418	DTALffKWKKZ	92	478.9348	3	38.64	1433.786	-2.2
8440	DTALffKWKKZ	92	478.9348	3	38.73	1433.786	-2.2
8487	DTALffKWKKZ	92	478.9348	3	38.94	1433.786	-2.3
8619	GNfTeGKWKKZ	92	446.5721	3	39.5	1336.708	-10
8634	NGfTeGKWKKZ	92	446.5716	3	39.57	1336.708	-11.3
8646	NGfTeGKWKKZ	92	446.5716	3	39.61	1336.708	-11.3
8687	DTALffKWKKZ	92	478.9349	3	39.79	1433.786	-2.2
9195	ePLEegKWKKZ	92	483.619	3	41.99	1447.838	-1.9
9201	ePLEegKWKKZ	92	362.9661	4	42.02	1447.838	-1.9
9267	ePLEegKWKKZ	92	483.6192	3	42.29	1447.838	-1.5
9291	ePLEegKWKKZ	92	483.6191	3	42.4	1447.838	-1.6
9303	ePLEegKWKKZ	92	483.619	3	42.45	1447.838	-2
9315	ePLEegKWKKZ	92	483.6192	3	42.5	1447.838	-1.5
9339	ePLEegKWKKZ	92	483.6191	3	42.6	1447.838	-1.7
9351	ePLEegKWKKZ	92	483.619	3	42.66	1447.838	-1.8
9425	ePLEegKWKKZ	92	483.6187	3	42.98	1447.838	-2.5
9524	ePLEegKWKKZ	92	483.6192	3	43.41	1447.838	-1.5
9609	ePLEegKWKKZ	92	483.6188	3	43.79	1447.838	-2.2
9645	ePLEegKWKKZ	92	483.6187	3	43.95	1447.838	-2.5

9803	GfLfGKWWKKZ	92	483.2719	3	44.65	1446.796	-1.8
9809	GfLfGKWWKKZ	92	483.2719	3	44.67	1446.796	-1.8
9821	GfLfGKWWKKZ	92	483.2717	3	44.73	1446.796	-2.2
9845	GfLfGKWWKKZ	92	483.2719	3	44.83	1446.796	-1.7
3139	SFEDGGKWKKZ	91	394.2076	3	16.87	1179.604	-2.2
5290	FFTFThKWKKZ	91	439.5879	3	25.9	1315.744	-1.7
7128	PfgghPKWKKZ	91	336.9684	4	33.33	1343.848	-2.8
8950	NhEKLeKWKKZ	91	452.2706	3	40.96	1353.792	-1.5
9474	FffbFfKWKKZ	91	420.2331	4	43.19	1676.906	-1.6
3157	SFEDGGKWKKZ	91	394.2078	3	16.95	1179.604	-1.7
4183	AgTLGNKWKKZ	91	391.2477	3	21.4	1170.724	-2
4438	fhNRLTKWKKZ	91	345.9613	4	22.45	1379.819	-2.1
4458	fhNRLTKWKKZ	91	345.9612	4	22.53	1379.819	-2.3
4471	fhNRLTKWKKZ	91	345.9612	4	22.58	1379.819	-2.6
4484	fhNRLTKWKKZ	91	345.9611	4	22.63	1379.819	-2.7
5089	TKfhKfKWKKZ	91	492.9663	3	25.1	1475.881	-2.3
5105	TKfhKfKWKKZ	91	492.9666	3	25.17	1475.881	-1.8
5222	KgfGTPKWKKZ	91	441.2751	3	25.63	1320.807	-2.8
5235	gKfGTPKWKKZ	91	441.2754	3	25.68	1320.807	-2
5253	gFgNTQKWKKZ	91	444.942	3	25.75	1331.808	-2.9
5260	gFgNTQKWKKZ	91	444.9422	3	25.78	1331.808	-2.3
5270	gFgNTQKWKKZ	91	444.942	3	25.82	1331.808	-2.7
5313	FFTFThKWKKZ	91	439.5876	3	25.99	1315.744	-2.3
5504	fNTRfNKWKKZ	91	507.2813	3	26.77	1518.825	-1.7
5538	NLhLgQKWKKZ	91	423.6108	3	26.91	1267.813	-1.9
5547	fNTRfNKWKKZ	91	507.2814	3	26.95	1518.825	-1.6
5551	NLhLgQKWKKZ	91	423.6108	3	26.96	1267.813	-2
5577	NLhLgQKWKKZ	91	423.6109	3	27.07	1267.813	-1.6
6545	fFKeSDKWKKZ	91	495.9384	3	31	1484.797	-2.4
6555	fFKeSDKWKKZ	91	372.2065	4	31.04	1484.797	0
6562	eAFfdEKWKKZ	91	383.711	4	31.06	1530.817	-1.7
6567	eAFfdEKWKKZ	91	511.2789	3	31.08	1530.817	-1.7
6583	eAFfdEKWKKZ	91	511.2785	3	31.15	1530.817	-2.5
6627	TNFLGfKWKKZ	91	448.5913	3	31.33	1342.755	-2.1
6644	afNfgAKWKKZ	91	362.4648	4	31.4	1445.834	-2.4
6666	TNFLGfKWKKZ	91	448.5914	3	31.49	1342.755	-1.9
6692	TNFLGfKWKKZ	91	448.5912	3	31.59	1342.755	-2.3
6839	afNfgAKWKKZ	91	362.4648	4	32.17	1445.834	-2.2
7059	gTEgfPKWKKZ	91	464.9505	3	33.05	1391.833	-2.4
7082	gTEgfPKWKKZ	91	464.9506	3	33.15	1391.833	-2.2
7134	gTEgfPKWKKZ	91	464.9506	3	33.35	1391.833	-2.3
7156	GFPLEfKWKKZ	91	452.2594	3	33.44	1353.76	-2.6
7189	GFPLEfKWKKZ	91	452.2595	3	33.57	1353.76	-2.2

7199	gTEgfPKWKKZ	91	464.9505	3	33.61	1391.833	-2.5
7231	GFPLEfKWKKZ	91	452.2595	3	33.73	1353.76	-2.4
7242	FGPLEfKWKKZ	91	452.2596	3	33.78	1353.76	-2.2
7294	GFPLEfKWKKZ	91	452.2595	3	33.98	1353.76	-2.2
7323	gTEgPfKWKKZ	91	464.9506	3	34.1	1391.833	-2.1
7348	GFPLEfKWKKZ	91	452.2596	3	34.2	1353.76	-2.2
7361	GFPLEfKWKKZ	91	452.2594	3	34.25	1353.76	-2.6
7545	gbDeDfKWKKZ	91	373.7083	4	35.03	1490.807	-2.2
7584	gbDeDfKWKKZ	91	373.7081	4	35.18	1490.807	-2.7
7597	GfAfFQKWKKZ	91	479.9317	3	35.24	1436.776	-1.6
7600	gbDeDfKWKKZ	91	373.7082	4	35.25	1490.807	-2.4
7843	egTQeEKWKKZ	91	489.9419	3	36.23	1466.807	-2.5
7856	egTQeEKWKKZ	91	489.942	3	36.28	1466.807	-2.2
7882	egTQeEKWKKZ	91	489.942	3	36.39	1466.807	-2.3
7895	egTQeEKWKKZ	91	489.942	3	36.44	1466.807	-2.2
7928	egTQeEKWKKZ	91	489.9422	3	36.57	1466.807	-1.8
8048	egTQeEKWKKZ	91	489.9418	3	37.05	1466.807	-2.6
8061	egTQeEKWKKZ	91	489.942	3	37.1	1466.807	-2.1
8324	DTALffKWKKZ	91	478.9348	3	38.23	1433.786	-2.3
8336	DTALffKWKKZ	91	478.9348	3	38.29	1433.786	-2.2
8501	DTALffKWKKZ	91	478.9349	3	39	1433.786	-2.2
8513	DTALffKWKKZ	91	478.9348	3	39.05	1433.786	-2.3
8525	DTALffKWKKZ	91	478.9349	3	39.1	1433.786	-2.1
8573	DTALffKWKKZ	91	478.9347	3	39.31	1433.786	-2.5
8963	NhEKLeKWKKZ	91	452.2701	3	41.01	1353.792	-2.6
8989	NhEKLeKWKKZ	91	452.2705	3	41.12	1353.792	-1.6
9002	NhEKLeKWKKZ	91	452.2705	3	41.17	1353.792	-1.7
9015	NhEKLeKWKKZ	91	452.2701	3	41.22	1353.792	-2.6
9028	NhEKLeKWKKZ	91	452.2699	3	41.28	1353.792	-3
9041	NhEKLeKWKKZ	91	452.2702	3	41.33	1353.792	-2.4
9054	NhEKLeKWKKZ	91	452.2701	3	41.39	1353.792	-2.6
9066	NhEKLeKWKKZ	91	452.2701	3	41.44	1353.792	-2.6
9079	NhEKLeKWKKZ	91	452.2704	3	41.49	1353.792	-2
9091	NhEKLeKWKKZ	91	452.2703	3	41.54	1353.792	-2.1
9117	NhEKLeKWKKZ	91	452.2708	3	41.65	1353.792	-1.1
9185	ePLEegKWKKZ	91	483.6188	3	41.95	1447.838	-2.3
9221	ePLEegKWKKZ	91	483.6188	3	42.1	1447.838	-2.3
9233	ePLEegKWKKZ	91	483.619	3	42.15	1447.838	-1.8
9245	ePLEegKWKKZ	91	483.6187	3	42.2	1447.838	-2.5
9257	ePLEegKWKKZ	91	483.6188	3	42.25	1447.838	-2.3
9281	ePLEegKWKKZ	91	483.6191	3	42.36	1447.838	-1.6
9377	ePLEegKWKKZ	91	483.619	3	42.77	1447.838	-1.8
9389	ePLEegKWKKZ	91	483.6191	3	42.82	1447.838	-1.7

9437	ePLEegKWKKZ	91	483.6191	3	43.03	1447.838	-1.7
9452	ePLEegKWKKZ	91	483.619	3	43.1	1447.838	-1.8
9461	ePLEegKWKKZ	91	483.6194	3	43.13	1447.838	-1
9473	ePLEegKWKKZ	91	483.6192	3	43.19	1447.838	-1.5
9485	ePLEegKWKKZ	91	483.6192	3	43.24	1447.838	-1.4
9503	ePLEegKWKKZ	91	483.6192	3	43.32	1447.838	-1.5
9515	ePLEegKWKKZ	91	483.6193	3	43.37	1447.838	-1.3
9551	ePLEegKWKKZ	91	483.6192	3	43.53	1447.838	-1.5
9563	ePLEegKWKKZ	91	483.6192	3	43.58	1447.838	-1.4
9657	ePLEegKWKKZ	91	483.6191	3	44	1447.838	-1.6
9860	GfLfGKWWKKZ	91	483.272	3	44.9	1446.796	-1.6
3931	TaDLKeKWKKZ	90	336.4546	4	20.31	1341.792	-2.1
6764	EFeNedKWKKZ	90	516.9419	3	31.87	1547.808	-2.4
7024	afNfgAKWKKZ	90	362.4648	4	32.91	1445.834	-2.5
7753	APPeeNKWKKZ	90	454.5879	3	35.87	1360.744	-1.9
3431	KFSLEaKWKKZ	90	323.9508	4	18.16	1291.776	-1.6
4529	fhNRLTKWKKZ	90	345.9611	4	22.82	1379.819	-2.8
5013	NFDSfdKWKKZ	90	484.2582	3	24.79	1449.756	-2
5023	PKLDFLKWKKZ	90	326.2069	4	24.83	1300.802	-2.6
5029	NFDSfdKWKKZ	90	484.2582	3	24.86	1449.756	-2
5039	PKLDFLKWKKZ	90	326.2072	4	24.9	1300.802	-1.9
5121	KTfhKfKWKKZ	90	492.9664	3	25.23	1475.881	-2.2
5274	FFTFThKWKKZ	90	439.5876	3	25.84	1315.744	-2.3
5296	gFgNTQKWKKZ	90	444.9422	3	25.93	1331.808	-2.4
5326	FFTFThKWKKZ	90	439.5877	3	26.05	1315.744	-2.2
5484	fNTRfNKWKKZ	90	380.7128	4	26.69	1518.825	-1.8
5525	NLhLgQKWKKZ	90	423.6104	3	26.85	1267.813	-2.8
5549	fNTRfNKWKKZ	90	507.2814	3	26.95	1518.825	-1.6
5564	NLhLgQKWKKZ	90	423.6106	3	27.02	1267.813	-2.3
5589	NLhLgQKWKKZ	90	423.6105	3	27.12	1267.813	-2.5
5675	ffKHgTKWKKZ	90	382.7292	4	27.47	1526.891	-2.4
5980	DfdDfKKWKKZ	90	392.9649	4	28.71	1567.834	-2.2
6132	SaeDfQKWKKZ	90	480.258	3	29.34	1437.756	-2.3
6155	SaeDfQKWKKZ	90	480.2581	3	29.44	1437.756	-2.2
6541	eAFfdEKWKKZ	90	511.2785	3	30.98	1530.817	-2.4
6653	TNFLGfKWKKZ	90	448.5912	3	31.44	1342.755	-2.3
6680	afNfgAKWKKZ	90	362.4648	4	31.54	1445.834	-2.5
6708	TNFLGfKWKKZ	90	448.5911	3	31.65	1342.755	-2.6
6766	EFeNedKWKKZ	90	516.9419	3	31.88	1547.808	-2.4
7122	gTEgfPKWKKZ	90	348.9646	4	33.31	1391.833	-2.8
7166	GFPLEfKWKKZ	90	452.2596	3	33.48	1353.76	-2.2
7173	gTEgfPKWKKZ	90	464.9505	3	33.5	1391.833	-2.4
7179	GFPLEfKWKKZ	90	452.2594	3	33.53	1353.76	-2.4

7186	gTEgfPKWKKZ	90	464.9506	3	33.55	1391.833	-2.2
7205	GFPLEfKWKKZ	90	452.2595	3	33.63	1353.76	-2.2
7254	gTEgfPKWKKZ	90	464.9501	3	33.82	1391.833	-3.2
7257	GFPLEfKWKKZ	90	452.2591	3	33.83	1353.76	-3.3
7267	gTEgfPKWKKZ	90	464.9505	3	33.87	1391.833	-2.5
7284	gTEgfPKWKKZ	90	464.9505	3	33.94	1391.833	-2.4
7296	GFPLEfKWKKZ	90	452.2595	3	33.99	1353.76	-2.2
7306	GFPLEfKWKKZ	90	452.2596	3	34.03	1353.76	-2.2
7332	GFPLEfKWKKZ	90	452.2595	3	34.13	1353.76	-2.2
8074	egTQeEKWKKZ	90	489.942	3	37.16	1466.807	-2.1
8294	PLLEegKWKKZ	90	455.6191	3	38.1	1363.838	-1.8
8303	DTALffKWKKZ	90	478.9351	3	38.14	1433.786	-1.8
8306	PLLEegKWKKZ	90	455.619	3	38.16	1363.838	-2
8312	DTALffKWKKZ	90	478.9347	3	38.18	1433.786	-2.5
8360	DTALffKWKKZ	90	478.9349	3	38.39	1433.786	-2.1
8372	DTALffKWKKZ	90	478.9346	3	38.44	1433.786	-2.7
8452	DTALffKWKKZ	90	478.9348	3	38.79	1433.786	-2.2
8459	DTALffKWKKZ	90	478.9348	3	38.82	1433.786	-2.3
8493	EgTeQeKWKKZ	90	489.9419	3	38.97	1466.807	-2.5
8537	DTALffKWKKZ	90	478.9348	3	39.15	1433.786	-2.3
8549	DTALffKWKKZ	90	478.935	3	39.2	1433.786	-1.9
8591	DTALffKWKKZ	90	478.9347	3	39.39	1433.786	-2.5
8595	NGfTeGKWKKZ	90	446.5718	3	39.4	1336.708	-10.7
8607	GNfTeGKWKKZ	90	446.5715	3	39.45	1336.708	-11.4
8615	DTALefKWKKZ	90	470.2629	3	39.49	1407.77	-2.5
8627	DTALefKWKKZ	90	470.2631	3	39.54	1407.77	-2.1
8670	NGfTeGKWKKZ	90	446.5722	3	39.72	1336.708	-10
8708	DTALffKWKKZ	90	478.935	3	39.89	1433.786	-1.9
8876	fegQbFKWKKZ	90	512.9663	3	40.64	1535.881	-2.3
8937	NhEKLeKWKKZ	90	452.2706	3	40.9	1353.792	-1.4
8976	NhEKLeKWKKZ	90	452.2706	3	41.06	1353.792	-1.5
9101	NhEKLeKWKKZ	90	452.2704	3	41.58	1353.792	-2
9152	ePLEegKWKKZ	90	483.6188	3	41.8	1447.838	-2.3
9161	ePLEegKWKKZ	90	483.6191	3	41.84	1447.838	-1.6
9173	ePLEegKWKKZ	90	483.6188	3	41.89	1447.838	-2.2
9197	ePLEegKWKKZ	90	483.619	3	42	1447.838	-1.9
9209	ePLEegKWKKZ	90	483.6188	3	42.05	1447.838	-2.3
9269	ePLEegKWKKZ	90	483.6192	3	42.3	1447.838	-1.5
9293	ePLEegKWKKZ	90	483.6191	3	42.41	1447.838	-1.6
9305	ePLEegKWKKZ	90	483.619	3	42.46	1447.838	-2
9353	ePLEegKWKKZ	90	483.619	3	42.67	1447.838	-1.8
9365	ePLEegKWKKZ	90	483.6191	3	42.72	1447.838	-1.7
9404	ePLEegKWKKZ	90	483.6191	3	42.89	1447.838	-1.6

9413	ePLEegKWKKZ	90	483.619	3	42.93	1447.838	-1.8
9489	FffbFfKWKKZ	90	420.2332	4	43.26	1676.906	-1.5
9498	FffbFfKWKKZ	90	420.2332	4	43.3	1676.906	-1.4
9536	ePLEegKWKKZ	90	483.6192	3	43.46	1447.838	-1.5
9831	GfLfGKWWKKZ	90	483.2715	3	44.77	1446.796	-2.5
9833	GfLfGKWWKKZ	90	483.2715	3	44.78	1446.796	-2.5
9879	GfLfGKWWKKZ	90	483.272	3	44.98	1446.796	-1.6
3832	GKEgFdKWKKZ	89	338.9601	4	19.88	1351.813	-1.2
4174	KaTfQLKWKKZ	89	346.2163	4	21.36	1380.839	-2.3
4933	FLAhaeKWKKZ	89	326.2018	4	24.48	1300.781	-2.1
5708	fgAGbFKWKKZ	89	447.2719	3	27.6	1338.797	-2
5786	fQhhbfKWKKZ	89	365.465	4	27.92	1457.834	-2
3145	SFEDGGKWKKZ	89	394.2074	3	16.89	1179.604	-2.7
3472	KFSLEaKWKKZ	89	323.9509	4	18.34	1291.776	-1.6
3847	KGEgFdKWKKZ	89	338.9597	4	19.95	1351.813	-2.1
4132	AKQfFNKWKKZ	89	350.7045	4	21.18	1398.793	-2.6
4977	PKLDFLKWKKZ	89	326.207	4	24.65	1300.802	-2.5
4984	PKLDFLKWKKZ	89	326.2068	4	24.68	1300.802	-2.9
4997	PKLDFLKWKKZ	89	326.207	4	24.73	1300.802	-2.4
5160	gKfGTPKWKKZ	89	331.2084	4	25.39	1320.807	-2
5173	gKfGTPKWKKZ	89	331.2082	4	25.44	1320.807	-2.7
5183	gKfGTPKWKKZ	89	441.2751	3	25.48	1320.807	-2.9
5195	gKfGTPKWKKZ	89	441.2754	3	25.53	1320.807	-2.2
5209	gKfGTPKWKKZ	89	441.2751	3	25.58	1320.807	-2.8
5245	gKfGTPKWKKZ	89	331.2082	4	25.72	1320.807	-2.5
5297	FFTFThKWKKZ	89	439.5878	3	25.93	1315.744	-1.9
5309	gFgNTQKWKKZ	89	444.9422	3	25.98	1331.808	-2.4
5325	fFKFNTKWKKZ	89	362.9596	4	26.04	1447.813	-2.3
5377	fFKFNTKWKKZ	89	362.9595	4	26.25	1447.813	-2.6
5533	fNTRfNKWKKZ	89	507.2809	3	26.89	1518.825	-2.5
5613	PSFfEAKWKKZ	89	448.2477	3	27.22	1341.723	-1.6
5662	ffKHgTKWKKZ	89	382.7293	4	27.42	1526.891	-2.1
5796	fQhhbfKWKKZ	89	365.4647	4	27.95	1457.834	-2.8
5847	fQhhbfKWKKZ	89	365.4651	4	28.16	1457.834	-1.7
5943	fGENDLKWKKZ	89	447.2422	3	28.56	1338.709	-2.7
6129	FfQEhLKWKKZ	89	471.9383	3	29.33	1412.797	-2.6
6167	FSfGgRKWKKZ	89	462.6103	3	29.49	1384.813	-3
6528	eAFfdEKWKKZ	89	511.2785	3	30.93	1530.817	-2.4
6538	eAFfdEKWKKZ	89	383.7108	4	30.97	1530.817	-2.3
6571	fFKeSDKWKKZ	89	372.2076	4	31.1	1484.797	2.8
6657	afNfgAKWKKZ	89	362.4648	4	31.45	1445.834	-2.3
6693	afNfgAKWKKZ	89	362.4648	4	31.6	1445.834	-2.5
6797	afNfgAKWKKZ	89	362.4647	4	32	1445.834	-2.7

6813	afNfgAKWKKZ	89	362.4648	4	32.07	1445.834	-2.5
6901	afNfgAKWKKZ	89	362.4647	4	32.42	1445.834	-2.6
6998	afNfgAKWKKZ	89	362.4648	4	32.81	1445.834	-2.4
7244	GFPLEfKWKKZ	89	452.2596	3	33.78	1353.76	-2.2
7273	GFPLEfKWKKZ	89	452.2593	3	33.9	1353.76	-2.6
7283	GFPLEfKWKKZ	89	452.2596	3	33.94	1353.76	-2.2
7291	gTEgfPKWKKZ	89	464.9506	3	33.97	1391.833	-2.3
7310	gTEgfPKWKKZ	89	464.9504	3	34.05	1391.833	-2.5
7356	PfgghPKWKKZ	89	336.9685	4	34.23	1343.848	-2.6
7638	GfAfFQKWKKZ	89	479.9315	3	35.41	1436.776	-2
7768	GfAfFQKWKKZ	89	479.9316	3	35.93	1436.776	-1.8
8162	egTQeEKWKKZ	89	489.942	3	37.52	1466.807	-2.1
8186	egTQeEKWKKZ	89	489.9419	3	37.63	1466.807	-2.3
8196	DTALffKWKKZ	89	478.9348	3	37.67	1433.786	-2.4
8246	egTQeEKWKKZ	89	489.9423	3	37.89	1466.807	-1.6
8348	DTALffKWKKZ	89	478.9348	3	38.34	1433.786	-2.3
8384	DTALffKWKKZ	89	478.9346	3	38.49	1433.786	-2.7
8396	DTALffKWKKZ	89	478.9345	3	38.54	1433.786	-2.9
8408	DTALffKWKKZ	89	478.9348	3	38.59	1433.786	-2.3
8430	DTALffKWKKZ	89	478.9348	3	38.69	1433.786	-2.3
8442	DTALffKWKKZ	89	478.9348	3	38.74	1433.786	-2.2
8471	DTALffKWKKZ	89	478.9349	3	38.87	1433.786	-2.1
8482	DTALffKWKKZ	89	478.9348	3	38.92	1433.786	-2.3
8489	DTALffKWKKZ	89	478.9348	3	38.95	1433.786	-2.3
8561	DTALffKWKKZ	89	478.9347	3	39.26	1433.786	-2.5
8585	NGfTeGKWKKZ	89	446.5716	3	39.36	1336.708	-11.3
8603	DTALefKWKKZ	89	470.2629	3	39.44	1407.77	-2.4
8658	GNfTeGKWKKZ	89	446.5717	3	39.66	1336.708	-11.1
8747	DTALffKWKKZ	89	478.9352	3	40.07	1433.786	-1.6
9246	FFLEegKWKKZ	89	362.9659	4	42.21	1447.838	-2.5
9317	ePLEegKWKKZ	89	483.6192	3	42.51	1447.838	-1.5
9332	ePLEegKWKKZ	89	483.6191	3	42.58	1447.838	-1.6
9341	ePLEegKWKKZ	89	483.6191	3	42.62	1447.838	-1.7
9575	ePLEegKWKKZ	89	483.6191	3	43.63	1447.838	-1.6
9611	ePLEegKWKKZ	89	483.6188	3	43.79	1447.838	-2.2
9647	ePLEegKWKKZ	89	483.6187	3	43.96	1447.838	-2.5
6590	SNFLGfKWKKZ	88	443.9192	3	31.18	1328.739	-2.6
8778	FFNeAeKWKKZ	88	366.2002	4	40.21	1460.776	-2.7
3862	GKEgFdKWKKZ	88	338.9598	4	20.01	1351.813	-1.9
4102	AKQfFNKWKKZ	88	350.7045	4	21.05	1398.793	-2.6
4138	KaTfQLKWKKZ	88	346.2164	4	21.21	1380.839	-2
4159	AKQfFNKWKKZ	88	350.7046	4	21.3	1398.793	-2.3
4753	EAfaLGKWKKZ	88	427.9192	3	23.74	1280.739	-2.7

4783	DhfaLGKWKKZ	88	427.9194	3	23.87	1280.739	-2.2
4798	EAfaLGKWKKZ	88	427.9194	3	23.93	1280.739	-2.2
5010	PKLDFLKWKKZ	88	326.207	4	24.78	1300.802	-2.4
5190	gKfGTPKWKKZ	88	331.2083	4	25.5	1320.807	-2.3
5232	gKfGTPKWKKZ	88	331.2084	4	25.67	1320.807	-2
5237	gKfGTPKWKKZ	88	441.2754	3	25.69	1320.807	-2
5292	FFTFThKWKKZ	88	439.5879	3	25.91	1315.744	-1.7
5335	fFKFNTKWKKZ	88	362.9598	4	26.08	1447.813	-1.7
5361	fFKFNTKWKKZ	88	362.9596	4	26.19	1447.813	-2.3
5442	fFKFNTKWKKZ	88	362.9598	4	26.51	1447.813	-1.9
5488	fNTRfNKWKKZ	88	380.7127	4	26.71	1518.825	-2
5514	fNTRfNKWKKZ	88	380.7126	4	26.81	1518.825	-2.1
5517	NLhLgQKWKKZ	88	423.6107	3	26.82	1267.813	-2.2
5605	NLhLgQKWKKZ	88	423.6109	3	27.19	1267.813	-1.6
5623	PSFfEAKWKKZ	88	448.2477	3	27.26	1341.723	-1.5
5631	PSFfEAKWKKZ	88	336.4375	4	27.29	1341.723	-1.9
5691	ffKHgTKWKKZ	88	382.7295	4	27.54	1526.891	-1.7
5698	ffKHgTKWKKZ	88	382.7293	4	27.56	1526.891	-2
6003	DfdDfKKWKKZ	88	392.9647	4	28.81	1567.834	-2.8
6065	DfdDfKKWKKZ	88	392.9647	4	29.06	1567.834	-2.7
6093	DfdDfKKWKKZ	88	392.9651	4	29.18	1567.834	-1.8
6139	FfQEhLKWKKZ	88	471.9384	3	29.37	1412.797	-2.5
6142	SaeDfQKWKKZ	88	360.4452	4	29.38	1437.756	-2.7
6162	FfQEhLKWKKZ	88	354.2052	4	29.47	1412.797	-3.7
6177	FSfGgRKWKKZ	88	462.6104	3	29.53	1384.813	-2.9
6190	FSfGgRKWKKZ	88	462.6107	3	29.58	1384.813	-2.2
6203	FSfGgRKWKKZ	88	462.6104	3	29.63	1384.813	-2.9
6521	eAFfdEKWKKZ	88	383.7109	4	30.9	1530.817	-1.9
6551	eAFfdEKWKKZ	88	383.711	4	31.02	1530.817	-1.5
6554	eAFfdEKWKKZ	88	511.279	3	31.03	1530.817	-1.5
6667	afNfgAKWKKZ	88	362.4649	4	31.49	1445.834	-2.1
6774	afNfgAKWKKZ	88	362.4648	4	31.91	1445.834	-2.4
6855	afNfgAKWKKZ	88	362.4647	4	32.23	1445.834	-2.6
6881	afNfgAKWKKZ	88	362.4648	4	32.33	1445.834	-2.3
6927	afNfgAKWKKZ	88	362.4647	4	32.52	1445.834	-2.7
6953	afNfgAKWKKZ	88	362.4648	4	32.63	1445.834	-2.5
7200	PfgghPKWKKZ	88	336.9686	4	33.61	1343.848	-2.4
7322	GFPLEfKWKKZ	88	452.2596	3	34.09	1353.76	-2.1
7574	GfAfFQKWKKZ	88	479.9315	3	35.14	1436.776	-2.2
7724	APPeeNKWKKZ	88	454.5879	3	35.75	1360.744	-1.8
7755	APPeeNKWKKZ	88	454.5879	3	35.88	1360.744	-1.9
8420	DTALffKWKKZ	88	478.9348	3	38.65	1433.786	-2.2
8516	EgTeQeKWKKZ	88	489.9418	3	39.07	1466.807	-2.5

8696	DTALffKWKKZ	88	478.9349	3	39.83	1433.786	-2.2
8903	fegQbFKWKKZ	88	384.9769	4	40.76	1535.881	-1.5
8952	NhEKLeKWKKZ	88	452.2706	3	40.96	1353.792	-1.5
8965	NhEKLeKWKKZ	88	452.2701	3	41.02	1353.792	-2.6
8978	NhEKLeKWKKZ	88	452.2706	3	41.07	1353.792	-1.5
8991	NhEKLeKWKKZ	88	452.2705	3	41.12	1353.792	-1.6
9004	NhEKLeKWKKZ	88	452.2705	3	41.18	1353.792	-1.7
9017	NhEKLeKWKKZ	88	452.2701	3	41.23	1353.792	-2.6
9030	NhEKLeKWKKZ	88	452.2699	3	41.29	1353.792	-3
9068	NhEKLeKWKKZ	88	452.2701	3	41.44	1353.792	-2.6
9081	NhEKLeKWKKZ	88	452.2704	3	41.5	1353.792	-2
9294	PeLEegKWKKZ	88	362.9661	4	42.41	1447.838	-1.9
9465	FffbFfKWKKZ	88	420.2333	4	43.15	1676.906	-1.2
9659	ePLEegKWKKZ	88	483.6191	3	44.01	1447.838	-1.6
3883	AAKDLFKWKKZ	87	309.1916	4	20.1	1232.739	-1.7
6389	FdDDfFKWKKZ	87	378.7004	4	30.38	1510.776	-2.5
8703	LNTQeEKWKKZ	87	457.5862	3	39.86	1369.751	-10
11244	eLhPLdKWKKZ	87	457.2876	3	50.98	1368.843	-1.7
3172	FSEDGGKWKKZ	87	394.2078	3	17.01	1179.604	-1.6
3448	KFSLEaKWKKZ	87	323.9506	4	18.23	1291.776	-2.3
3484	KFSLEaKWKKZ	87	323.9506	4	18.39	1291.776	-2.5
4111	AKQfFNKWKKZ	87	350.7044	4	21.09	1398.793	-2.7
4120	AKQfFNKWKKZ	87	350.7045	4	21.12	1398.793	-2.5
4144	AKQfFNKWKKZ	87	350.7047	4	21.23	1398.793	-2.1
4173	AgTLGNKWKKZ	87	391.2475	3	21.35	1170.724	-2.5
5224	gKfGTPKWKKZ	87	441.2751	3	25.64	1320.807	-2.8
5315	FFTFThKWKKZ	87	439.5876	3	26	1315.744	-2.3
5328	FFTFThKWKKZ	87	439.5877	3	26.05	1315.744	-2.2
5403	fFKFNTKWKKZ	87	362.9596	4	26.35	1447.813	-2.4
5416	fFKFNTKWKKZ	87	362.9598	4	26.41	1447.813	-1.8
5501	fNTRfNKWKKZ	87	380.7127	4	26.76	1518.825	-1.9
5540	NLhLgQKWKKZ	87	423.6108	3	26.92	1267.813	-1.9
5809	fQhhbfKWKKZ	87	365.4648	4	28.01	1457.834	-2.6
5860	fQhhbfKWKKZ	87	365.4646	4	28.21	1457.834	-3
5902	fQhhbfKWKKZ	87	365.4648	4	28.39	1457.834	-2.5
5990	DfdDfKKWKKZ	87	392.9648	4	28.75	1567.834	-2.5
6023	DfdDfKKWKKZ	87	392.9651	4	28.89	1567.834	-1.6
6078	DfdDfKKWKKZ	87	392.9649	4	29.12	1567.834	-2.2
6102	SaeDfQKWKKZ	87	360.4453	4	29.22	1437.756	-2.4
6111	SaeDfQKWKKZ	87	360.4451	4	29.26	1437.756	-2.9
6121	FfQEhLKWKKZ	87	354.2056	4	29.3	1412.797	-2.6
6126	SaeDfQKWKKZ	87	360.445	4	29.32	1437.756	-3.3
6134	FfQEhLKWKKZ	87	354.2056	4	29.35	1412.797	-2.5

6147	FfQEhLKWKKZ	87	354.2056	4	29.4	1412.797	-2.5
6152	FfQEhLKWKKZ	87	471.9384	3	29.43	1412.797	-2.6
6170	FfQEhLKWKKZ	87	471.938	3	29.5	1412.797	-3.4
6515	fFKeSDKWKKZ	87	372.2057	4	30.88	1484.797	-2
6525	eAFfdEKWKKZ	87	383.7107	4	30.92	1530.817	-2.4
6577	eAFfdEKWKKZ	87	383.7107	4	31.13	1530.817	-2.4
6673	afNfgAKWKKZ	87	482.9508	3	31.52	1445.834	-1.9
6683	afNfgAKWKKZ	87	482.9507	3	31.56	1445.834	-2.2
6748	afNfgAKWKKZ	87	362.4648	4	31.81	1445.834	-2.4
6761	afNfgAKWKKZ	87	362.4647	4	31.86	1445.834	-2.6
6787	afNfgAKWKKZ	87	362.4648	4	31.96	1445.834	-2.4
7651	GfAfFQKWKKZ	87	479.9317	3	35.46	1436.776	-1.7
8198	DTALffKWKKZ	87	478.9348	3	37.68	1433.786	-2.4
8787	FFNeAeKWKKZ	87	366.2003	4	40.24	1460.776	-2.4
9043	NhEKLeKWKKZ	87	452.2702	3	41.34	1353.792	-2.4
9056	NhEKLeKWKKZ	87	452.2701	3	41.4	1353.792	-2.6
9093	NhEKLeKWKKZ	87	452.2703	3	41.55	1353.792	-2.1
9119	NhEKLeKWKKZ	87	452.2708	3	41.66	1353.792	-1.1
9284	ePLEegKWKKZ	87	362.9661	4	42.37	1447.838	-1.9
9881	GfLfGKWWKKZ	87	483.272	3	44.99	1446.796	-1.6
11246	eLhPLdKWKKZ	87	457.2876	3	50.99	1368.843	-1.7
11292	eLhPLdKWKKZ	87	457.2873	3	51.19	1368.843	-2.4
4913	dAQgfGKWKKZ	86	343.4569	4	24.4	1369.802	-2.6
4964	GffGGaKWKKZ	86	435.9126	3	24.6	1304.718	-1.8
5650	hfFgSbKWKKZ	86	346.7124	4	27.37	1382.823	-1.7
3454	KFSLEaKWKKZ	86	323.9507	4	18.26	1291.776	-1.9
3802	KGEgFdKWKKZ	86	338.9598	4	19.75	1351.813	-1.9
3874	AAKDLFKWKKZ	86	309.1915	4	20.06	1232.739	-1.9
3895	AAKDLFKWKKZ	86	309.1914	4	20.15	1232.739	-2.1
4128	AKQfFNKWKKZ	86	467.2703	3	21.16	1398.793	-2.4
4185	AgTLGNKWKKZ	86	391.2477	3	21.41	1170.724	-2
4768	hDfaLGKWKKZ	86	427.9195	3	23.8	1280.739	-2
4967	FKeNFKKWKKZ	86	363.2178	4	24.61	1448.845	-1.6
5073	KTfhKfKWKKZ	86	369.9769	4	25.04	1475.881	-1.5
5112	KTfhKfKWKKZ	86	369.9766	4	25.19	1475.881	-2.3
5125	KTfhKfKWKKZ	86	369.9763	4	25.25	1475.881	-2.9
5136	TKfhKfKWKKZ	86	492.9661	3	25.29	1475.881	-2.8
5177	gKfGTPKWKKZ	86	331.2082	4	25.45	1320.807	-2.6
5185	gKfGTPKWKKZ	86	441.2751	3	25.49	1320.807	-2.9
5203	KgfGTPKWKKZ	86	331.2082	4	25.56	1320.807	-2.7
5216	KgfGTPKWKKZ	86	331.2082	4	25.61	1320.807	-2.5
5276	FFTFThKWKKZ	86	439.5876	3	25.85	1315.744	-2.3
5299	FFTFThKWKKZ	86	439.5878	3	25.94	1315.744	-1.9

5348	fFKFNTKWKKZ	86	362.9598	4	26.13	1447.813	-1.9
5390	fFKFNTKWKKZ	86	362.9596	4	26.3	1447.813	-2.5
5520	fNTRfNKWKKZ	86	507.2813	3	26.83	1518.825	-1.7
5527	NLhLgQKWKKZ	86	423.6104	3	26.86	1267.813	-2.8
5543	fNTRfNKWKKZ	86	380.7127	4	26.93	1518.825	-2
5553	NLhLgQKWKKZ	86	423.6108	3	26.97	1267.813	-2
5579	NLhLgQKWKKZ	86	423.6109	3	27.08	1267.813	-1.6
5586	PSFfEAKKWKZ	86	336.4375	4	27.11	1341.723	-1.8
5634	PSFfEAKWKKZ	86	448.2477	3	27.31	1341.723	-1.5
5695	fgAGbFKWKKZ	86	447.272	3	27.55	1338.797	-1.8
5835	fQhhbfKWKKZ	86	365.4647	4	28.11	1457.834	-2.7
5872	fQhhbfKWKKZ	86	365.4648	4	28.26	1457.834	-2.6
6094	SaeDfQKWKKZ	86	360.4456	4	29.19	1437.756	-1.7
6647	TNFLGfKWKKZ	86	672.3831	2	31.41	1342.755	-2.4
6738	afNfgAKWKKZ	86	482.9507	3	31.77	1445.834	-2.3
6868	afNfgAKWKKZ	86	362.4648	4	32.28	1445.834	-2.4
7037	afNfgAKWKKZ	86	362.4647	4	32.96	1445.834	-2.7
7135	PfgghPKWKKZ	86	336.9685	4	33.36	1343.848	-2.6
7726	APPeeNKWKKZ	86	454.5879	3	35.76	1360.744	-1.8
7781	GfAfFQKWKKZ	86	479.9315	3	35.98	1436.776	-2.2
8597	GNfTeGKWKKZ	86	446.5718	3	39.41	1336.708	-10.7
8621	NGfTeGKWKKZ	86	446.5721	3	39.51	1336.708	-10
8643	TbTQeEKWKKZ	86	457.5858	3	39.6	1369.751	-11.2
8891	fegQbFKWKKZ	86	384.9765	4	40.7	1535.881	-2.4
8939	NhEKLeKWKKZ	86	452.2706	3	40.91	1353.792	-1.4
9103	NhEKLeKWKKZ	86	452.2704	3	41.59	1353.792	-2
9215	ePLEegKWKKZ	86	362.966	4	42.07	1447.838	-2.1
9308	ePLEegKWKKZ	86	362.9661	4	42.47	1447.838	-1.9
9476	FffbFfKWKKZ	86	420.2331	4	43.2	1676.906	-1.6
11184	eLhPLdKWKKZ	86	457.2875	3	50.72	1368.843	-1.9
11294	eLhPLdKWKKZ	86	457.2873	3	51.2	1368.843	-2.4
3958	RKLgKLKWKKZ	85	339.2438	4	20.42	1352.95	-2.7
4050	GGaLgLKWKKZ	85	385.928	3	20.82	1154.765	-2.5
4234	NhEKLeKWKKZ	85	339.4546	4	21.62	1353.792	-1.9
10275	QHNfgAKWKKZ	85	463.5999	3	46.72	1387.788	-7.1
3117	SFEDGGKWKKZ	85	394.2078	3	16.77	1179.604	-1.7
3141	SFEDGGKWKKZ	85	394.2076	3	16.88	1179.604	-2.2
3159	SFEDGGKWKKZ	85	394.2078	3	16.96	1179.604	-1.7
3811	GKEgFdKWKKZ	85	338.9597	4	19.79	1351.813	-2.2
3871	GKEgFdKWKKZ	85	338.9599	4	20.05	1351.813	-1.7
3922	AAKDLFKWKKZ	85	309.191	4	20.27	1232.739	-3.4
4027	RKLgKLKWKKZ	85	339.2438	4	20.72	1352.95	-2.7
5151	TKfhKfKWKKZ	85	369.9766	4	25.35	1475.881	-2.2

5167	TKfhKfKWKKZ	85	369.9765	4	25.42	1475.881	-2.6
5211	gKfGTPKWKKZ	85	441.2751	3	25.59	1320.807	-2.8
5302	gFgNTQKWKKZ	85	333.9586	4	25.95	1331.808	-2.1
5429	fFKFNTKWKKZ	85	362.96	4	26.46	1447.813	-1.4
5432	QeLbdgKWKKZ	85	364.7295	4	26.47	1454.891	-1.7
5572	fNTRfNKWKKZ	85	380.7126	4	27.05	1518.825	-2.1
5591	NLhLgQKWKKZ	85	423.6105	3	27.13	1267.813	-2.5
5669	fgAGbFKWKKZ	85	447.2719	3	27.45	1338.797	-2
5822	fQhhbfKWKKZ	85	365.4651	4	28.06	1457.834	-1.8
6013	DfdDfKKWKKZ	85	392.9652	4	28.85	1567.834	-1.5
6039	DfdDfKKWKKZ	85	392.9648	4	28.95	1567.834	-2.6
6161	FSfGgRKWKKZ	85	347.2096	4	29.46	1384.813	-2.9
6174	FSfGgRKWKKZ	85	347.2097	4	29.51	1384.813	-2.5
6187	FSfGgRKWKKZ	85	347.2097	4	29.57	1384.813	-2.5
6200	FSfGgRKWKKZ	85	347.2097	4	29.62	1384.813	-2.7
6531	fFKeSDKWKKZ	85	372.2056	4	30.94	1484.797	-2.5
6660	TNFLGfKWKKZ	85	672.3832	2	31.46	1342.755	-2.2
6803	afNfgAKWKKZ	85	482.9505	3	32.03	1445.834	-2.6
6940	afNfgAKWKKZ	85	362.4648	4	32.57	1445.834	-2.4
7008	afNfgAKWKKZ	85	362.4645	4	32.85	1445.834	-3.2
8364	DTALffKWKKZ	85	717.8981	2	38.41	1433.786	-2.9
8431	DTALffKWKKZ	85	717.8983	2	38.7	1433.786	-2.7
8609	GNfTeGKWKKZ	85	446.5715	3	39.46	1336.708	-11.4
8631	TbTQeEKWKKZ	85	457.5858	3	39.55	1369.751	-11
8648	NGfTeGKWKKZ	85	446.5716	3	39.62	1336.708	-11.3
8660	NGfTeGKWKKZ	85	446.5717	3	39.67	1336.708	-11.1
8667	TbTQeEKWKKZ	85	457.586	3	39.7	1369.751	-10.7
8691	TbTQeEKWKKZ	85	457.586	3	39.81	1369.751	-10.7
8715	LNTQeEKWKKZ	85	457.5858	3	39.92	1369.751	-10.8
9176	ePLEegKWKKZ	85	362.9661	4	41.91	1447.838	-1.9
9191	fALEegKWKKZ	85	362.9659	4	41.97	1447.838	-2.2
9360	ePLEegKWKKZ	85	362.966	4	42.7	1447.838	-2
9525	FffbFfKWKKZ	85	420.233	4	43.42	1676.906	-1.8
11172	eLhPLdKWKKZ	85	457.2879	3	50.67	1368.843	-1.1
11186	eLhPLdKWKKZ	85	457.2875	3	50.73	1368.843	-1.9
11268	eLhPLdKWKKZ	85	457.2875	3	51.09	1368.843	-1.9
11280	eLhPLdKWKKZ	85	457.288	3	51.14	1368.843	-0.7
11304	eLhPLdKWKKZ	85	457.2878	3	51.24	1368.843	-1.1
4744	EFGGedKWKKZ	84	451.2473	3	23.7	1350.724	-2.7
5495	FFNRhfKWKKZ	84	365.9623	4	26.73	1459.824	-2.5
8538	NEHTeGKWKKZ	84	441.9	3	39.16	1322.688	-7.7
4074	GGaLgLKWKKZ	84	385.9279	3	20.93	1154.765	-2.7
4137	AKQfFNKWKKZ	84	467.2702	3	21.2	1398.793	-2.7

4149	AgTLGNKWKKZ	84	293.6877	4	21.25	1170.724	-1.8
4152	AKQfFNKWKKZ	84	467.2706	3	21.26	1398.793	-1.7
4194	AgTLGNKWKKZ	84	293.6874	4	21.44	1170.724	-2.5
4197	AgTLGNKWKKZ	84	391.2474	3	21.46	1170.724	-2.8
4248	KfGQdLKWKKZ	84	354.2176	4	21.67	1412.845	-2.2
4398	bgRdGfKWKKZ	84	364.2281	4	22.29	1452.887	-2.6
4418	bgRdGfKWKKZ	84	364.2282	4	22.37	1452.887	-2.1
4522	bgRdGfKWKKZ	84	364.2281	4	22.79	1452.887	-2.5
4548	bgRdGfKWKKZ	84	364.2283	4	22.89	1452.887	-2.1
4855	RTfQAFKWKKZ	84	354.4575	4	24.16	1413.803	-1.8
5086	TKfhKfKWKKZ	84	369.9765	4	25.09	1475.881	-2.4
5091	TKfhKfKWKKZ	84	492.9663	3	25.11	1475.881	-2.3
5099	KTfhKfKWKKZ	84	369.9767	4	25.14	1475.881	-2.1
5138	TKfhKfKWKKZ	84	369.9765	4	25.3	1475.881	-2.5
5263	gFgNTQKWKKZ	84	333.9584	4	25.79	1331.808	-2.5
5286	gFgNTQKWKKZ	84	333.9586	4	25.89	1331.808	-1.9
5433	eLhPLdKWKKZ	84	457.2878	3	26.48	1368.843	-1.3
5530	fNTRfNKWKKZ	84	380.7125	4	26.88	1518.825	-2.4
5566	NLhLgQKWKKZ	84	423.6106	3	27.03	1267.813	-2.3
5659	fgAGbFKWKKZ	84	447.272	3	27.41	1338.797	-1.9
5720	fgAGbFKWKKZ	84	447.2714	3	27.65	1338.797	-3.1
6052	DfdDfKKWKKZ	84	392.9648	4	29.01	1567.834	-2.5
6630	TNFLGfKWKKZ	84	336.6952	4	31.34	1342.755	-2.4
6688	TNFLGfKWKKZ	84	336.6952	4	31.57	1342.755	-2.3
6703	afNfgAKWKKZ	84	362.4647	4	31.63	1445.834	-2.6
6972	afNfgAKWKKZ	84	362.465	4	32.7	1445.834	-1.8
7278	PfgghPKWKKZ	84	448.9557	3	33.92	1343.848	-2.4
7286	gTEgfPKWKKZ	84	464.9505	3	33.95	1391.833	-2.4
7325	gTEgPfKWKKZ	84	464.9506	3	34.1	1391.833	-2.1
7853	egTQeEKWKKZ	84	367.7082	4	36.27	1466.807	-2.4
8378	DTALffKWKKZ	84	359.4526	4	38.47	1433.786	-3.1
8426	DTALffKWKKZ	84	359.4531	4	38.67	1433.786	-1.8
8436	DTALffKWKKZ	84	359.4531	4	38.72	1433.786	-1.7
8443	DTALffKWKKZ	84	717.8982	2	38.75	1433.786	-2.8
8460	DTALffKWKKZ	84	717.8975	2	38.83	1433.786	-3.8
8495	EgTeQeKWKKZ	84	489.9419	3	38.98	1466.807	-2.5
8636	GNfTeGKWKKZ	84	446.5716	3	39.57	1336.708	-11.3
8795	fegQbFKWKKZ	84	384.9766	4	40.28	1535.881	-2.3
9227	ePLEegKWKKZ	84	362.9659	4	42.13	1447.838	-2.2
9236	ePLEegKWKKZ	84	362.966	4	42.16	1447.838	-2
9248	FFLEegKWKKZ	84	362.9659	4	42.21	1447.838	-2.5
9260	ePLEegKWKKZ	84	362.9659	4	42.27	1447.838	-2.2
9272	ePLEegKWKKZ	84	362.9661	4	42.32	1447.838	-1.7

9467	FffbFfKWKKZ	84	420.2333	4	43.16	1676.906	-1.2
10287	HQNfgAKWKKZ	84	463.5997	3	46.77	1387.788	-7.5
10299	QHNfgAKWKKZ	84	463.6002	3	46.82	1387.788	-6.6
11157	eLhPLdKWKKZ	84	457.2874	3	50.61	1368.843	-2.1
11174	eLhPLdKWKKZ	84	457.2879	3	50.68	1368.843	-1.1
11232	eLhPLdKWKKZ	84	457.2878	3	50.93	1368.843	-1.3
11256	eLhPLdKWKKZ	84	457.2877	3	51.03	1368.843	-1.5
11270	eLhPLdKWKKZ	84	457.2875	3	51.09	1368.843	-1.9
11282	eLhPLdKWKKZ	84	457.288	3	51.15	1368.843	-0.7
11306	eLhPLdKWKKZ	84	457.2878	3	51.25	1368.843	-1.1
3407	hFEEADKWKKZ	83	316.9222	4	18.05	1263.661	-1.1
4359	AhAfgGKWKKZ	83	408.2563	3	22.14	1221.739	6.9
10120	NHNfgAKWKKZ	83	458.928	3	46.06	1373.772	-7
3823	GKEgFdKWKKZ	83	338.9598	4	19.84	1351.813	-1.9
3849	KGEgFdKWKKZ	83	338.9597	4	19.95	1351.813	-2.1
3933	TaDLKeKWKKZ	83	336.4546	4	20.31	1341.792	-2.1
3943	RKLgKLKWKKZ	83	339.2439	4	20.36	1352.95	-2.5
3964	RKLgKLKWKKZ	83	339.2439	4	20.44	1352.95	-2.3
4032	GGaLgLKWKKZ	83	385.9281	3	20.74	1154.765	-2.3
4098	GGaLgLKWKKZ	83	385.928	3	21.03	1154.765	-2.5
4158	AgTLGNKWKKZ	83	293.6875	4	21.29	1170.724	-2.2
4444	bgRdGfKWKKZ	83	364.2282	4	22.48	1452.887	-2.3
4447	fhNRLTKWKKZ	83	345.9611	4	22.49	1379.819	-2.8
4470	bgRdGfKWKKZ	83	364.2281	4	22.58	1452.887	-2.4
4483	bgRdGfKWKKZ	83	364.2281	4	22.63	1452.887	-2.6
4499	fhNRLTKWKKZ	83	345.9611	4	22.69	1379.819	-2.8
4577	bgRdGfKWKKZ	83	364.2282	4	23.01	1452.887	-2.2
4596	bgRdGfKWKKZ	83	364.2282	4	23.09	1452.887	-2.3
4990	GffGGaKWKKZ	83	435.9122	3	24.7	1304.718	-2.7
5000	GffGGaKWKKZ	83	435.9124	3	24.74	1304.718	-2.3
5107	TKfhKfKWKKZ	83	492.9666	3	25.18	1475.881	-1.8
5123	TKfhKfKWKKZ	83	492.9664	3	25.24	1475.881	-2.2
5337	fFKFNTKWKKZ	83	483.6109	3	26.09	1447.813	-1.4
5418	fFKFNTKWKKZ	83	483.6109	3	26.41	1447.813	-1.4
5435	eLhPLdKWKKZ	83	457.2878	3	26.48	1368.843	-1.3
5556	fNTRfNKWKKZ	83	380.7127	4	26.98	1518.825	-2
5607	NLhLgQKWKKZ	83	423.6109	3	27.2	1267.813	-1.6
5682	fgAGbFKWKKZ	83	447.272	3	27.5	1338.797	-1.9
5711	ffKHgTKWKKZ	83	382.7293	4	27.62	1526.891	-2.2
5887	fQhhbfKWKKZ	83	365.4648	4	28.33	1457.834	-2.5
6649	TNFLGfKWKKZ	83	672.3831	2	31.42	1342.755	-2.4
6696	afNfgAKWKKZ	83	482.9506	3	31.61	1445.834	-2.5
6719	afNfgAKWKKZ	83	362.4648	4	31.7	1445.834	-2.5

7148	PfgghPKWKKZ	83	336.9685	4	33.41	1343.848	-2.6
7174	PfgghPKWKKZ	83	336.9686	4	33.51	1343.848	-2.4
7268	PfgghPKWKKZ	83	448.9556	3	33.88	1343.848	-2.5
7293	gTEgfPKWKKZ	83	464.9506	3	33.98	1391.833	-2.3
7297	PfgghPKWKKZ	83	448.9557	3	33.99	1343.848	-2.2
7307	PfgghPKWKKZ	83	448.9557	3	34.03	1343.848	-2.4
7986	egTQeEKWKKZ	83	367.7085	4	36.8	1466.807	-1.8
8051	egTQeEKWKKZ	83	367.7082	4	37.06	1466.807	-2.6
8366	DTALffKWKKZ	83	717.8981	2	38.41	1433.786	-2.9
8385	DTALffKWKKZ	83	717.8979	2	38.5	1433.786	-3.2
8411	DTALffKWKKZ	83	359.4525	4	38.61	1433.786	-3.4
8433	DTALffKWKKZ	83	717.8983	2	38.7	1433.786	-2.7
8453	DTALffKWKKZ	83	717.8981	2	38.8	1433.786	-2.9
8622	TbTQeEKWKKZ	83	457.5863	3	39.52	1369.751	-10
8705	TbTQeEKWKKZ	83	457.5862	3	39.87	1369.751	-10.2
9296	PeLEegKWKKZ	83	362.9661	4	42.42	1447.838	-1.9
9335	ePLEegKWKKZ	83	362.9661	4	42.59	1447.838	-1.8
9441	FffbFfKWKKZ	83	420.233	4	43.05	1676.906	-1.8
9491	FffbFfKWKKZ	83	420.2332	4	43.27	1676.906	-1.5
11234	eLhPLdKWKKZ	83	457.2878	3	50.94	1368.843	-1.3
11258	eLhPLdKWKKZ	83	457.2877	3	51.04	1368.843	-1.5
5266	hPPASEKWKKZ	82	385.5613	3	25.81	1153.661	1.1
3174	FSEDGGKWKKZ	82	394.2078	3	17.02	1179.604	-1.6
3398	hFEEADKWKKZ	82	316.9218	4	18.02	1263.661	-2.5
4038	GGaLgLKWKKZ	82	385.928	3	20.77	1154.765	-2.6
4062	GGaLgLKWKKZ	82	385.9279	3	20.87	1154.765	-2.7
4086	GGaLgLKWKKZ	82	385.9281	3	20.98	1154.765	-2.3
4170	AgTLGNKWKKZ	82	293.6875	4	21.34	1170.724	-2.3
4182	AgTLGNKWKKZ	82	293.6876	4	21.39	1170.724	-2
4375	AhAfgGKWKKZ	82	408.2563	3	22.2	1221.739	6.8
4388	bgRdGfKWKKZ	82	364.2281	4	22.25	1452.887	-2.6
4440	fhNRLTKWKKZ	82	345.9613	4	22.46	1379.819	-2.1
4457	bgRdGfKWKKZ	82	364.2283	4	22.53	1452.887	-2.1
4460	fhNRLTKWKKZ	82	345.9612	4	22.54	1379.819	-2.3
4473	fhNRLTKWKKZ	82	345.9612	4	22.59	1379.819	-2.6
4486	fhNRLTKWKKZ	82	345.9611	4	22.64	1379.819	-2.7
4496	bgRdGfKWKKZ	82	364.2281	4	22.68	1452.887	-2.5
4535	bgRdGfKWKKZ	82	364.2282	4	22.84	1452.887	-2.1
4756	EFGGedKWKKZ	82	451.2474	3	23.75	1350.724	-2.4
4835	TRfhNFKWKKZ	82	472.2737	3	24.08	1413.804	-2.9
4974	GffGGaKWKKZ	82	435.9124	3	24.64	1304.718	-2.3
4980	FKeNFKKWKKZ	82	363.2176	4	24.66	1448.845	-2.3
5279	hPPASEKWKKZ	82	385.5613	3	25.86	1153.661	1

5289	hPPASEKWKKZ	82	385.5615	3	25.9	1153.661	1.6
5350	fFKFNTKWKKZ	82	483.6106	3	26.14	1447.813	-2
5366	fFKFNTKWKKZ	82	483.6103	3	26.2	1447.813	-2.6
5395	fFKFNTKWKKZ	82	483.6103	3	26.32	1447.813	-2.6
5405	fFKFNTKWKKZ	82	483.6104	3	26.36	1447.813	-2.4
5446	fFKFNTKWKKZ	82	483.6105	3	26.53	1447.813	-2
5447	fFKFNTKWKKZ	82	483.6105	3	26.53	1447.813	-2
5798	fQhhbfKWKKZ	82	365.4647	4	27.96	1457.834	-2.8
5849	fQhhbfKWKKZ	82	365.4651	4	28.17	1457.834	-1.7
6213	FSfGgRKWKKZ	82	347.2097	4	29.67	1384.813	-2.4
6217	FSfGgRKWKKZ	82	462.6107	3	29.69	1384.813	-2.2
6391	FdDDfFKWKKZ	82	378.7004	4	30.38	1510.776	-2.5
6514	fFKeSDKWKKZ	82	372.2057	4	30.87	1484.797	-2
6643	TNFLGfKWKKZ	82	336.6952	4	31.39	1342.755	-2.4
6656	TNFLGfKWKKZ	82	336.6952	4	31.45	1342.755	-2.5
7161	PfgghPKWKKZ	82	336.9686	4	33.46	1343.848	-2.2
7190	PfgghPKWKKZ	82	336.9686	4	33.57	1343.848	-2.4
7317	PfgghPKWKKZ	82	448.9557	3	34.07	1343.848	-2.2
7583	gbDeDfKWKKZ	82	373.7081	4	35.18	1490.807	-2.7
7885	egTQeEKWKKZ	82	367.7082	4	36.4	1466.807	-2.6
7911	egTQeEKWKKZ	82	367.7083	4	36.5	1466.807	-2.4
7950	egTQeEKWKKZ	82	367.7082	4	36.66	1466.807	-2.6
8387	DTALffKWKKZ	82	717.8979	2	38.5	1433.786	-3.2
8462	DTALffKWKKZ	82	717.8975	2	38.84	1433.786	-3.8
8473	EgTeQeKWKKZ	82	367.7085	4	38.88	1466.807	-1.8
8529	NEHTeGKWKKZ	82	441.8999	3	39.12	1322.688	-7.8
8550	NEHTeGKWKKZ	82	441.9	3	39.21	1322.688	-7.7
8586	NEHTeGKWKKZ	82	441.8998	3	39.37	1322.688	-8.1
8672	GNfTeGKWKKZ	82	446.5722	3	39.72	1336.708	-10
8681	TbTQeEKWKKZ	82	457.5862	3	39.77	1369.751	-10.2
8693	TbTQeEKWKKZ	82	457.586	3	39.82	1369.751	-10.7
8733	TbTQeEKWKKZ	82	457.5859	3	40	1369.751	-10.9
9456	ePLEegKWKKZ	82	362.966	4	43.12	1447.838	-2
9500	FffbFfKWKKZ	82	420.2332	4	43.3	1676.906	-1.4
11159	eLhPLdKWKKZ	82	457.2874	3	50.61	1368.843	-2.1
4206	PhPGGHKWKKZ	81	373.558	3	21.5	1117.651	1.2
4991	begDTbKWKKZ	81	345.9583	4	24.71	1379.808	-2.9
6466	AGPhfeKWKKZ	81	440.2523	3	30.68	1317.739	-2.7
3129	FSEDGGKWKKZ	81	394.2075	3	16.83	1179.604	-2.4
3497	KFSLEaKWKKZ	81	323.9505	4	18.45	1291.776	-2.8
3498	KFSLEaKWKKZ	81	323.9505	4	18.45	1291.776	-2.8
3834	GKEgFdKWKKZ	81	338.9601	4	19.89	1351.813	-1.2
3864	GKEgFdKWKKZ	81	338.9598	4	20.02	1351.813	-1.9

4110	GGaLgLKWKKZ	81	385.928	3	21.08	1154.765	-2.6
4408	bgRdGfKWKKZ	81	364.2282	4	22.33	1452.887	-2.3
4431	bgRdGfKWKKZ	81	364.2284	4	22.42	1452.887	-1.6
4735	EFGGedKWKKZ	81	451.2475	3	23.67	1350.724	-2.2
4800	TRfQAFKWKKZ	81	354.4572	4	23.93	1413.803	-2.4
4813	TRfQAFKWKKZ	81	354.4574	4	23.99	1413.803	-1.8
4826	TRfQAFKWKKZ	81	354.4572	4	24.04	1413.803	-2.5
4943	FLAhaeKWKKZ	81	326.202	4	24.52	1300.781	-1.4
5004	begDTbKWKKZ	81	345.9584	4	24.76	1379.808	-2.4
5449	FFNRhfKWKKZ	81	365.9624	4	26.54	1459.824	-2.3
5476	FFNRhfKWKKZ	81	365.9625	4	26.66	1459.824	-2.2
5483	fNTRfNKWKKZ	81	380.7128	4	26.68	1518.825	-1.8
5597	PSFfEAKWKKZ	81	336.4374	4	27.16	1341.723	-2.2
5610	PSFfEAKWKKZ	81	336.4376	4	27.21	1341.723	-1.7
5620	PSFfEAKWKKZ	81	336.4376	4	27.25	1341.723	-1.6
5904	fQhhbfKWKKZ	81	365.4648	4	28.4	1457.834	-2.5
6118	FfQEhLKWKKZ	81	354.2055	4	29.29	1412.797	-2.7
6517	fFKeSDKWKKZ	81	495.9386	3	30.88	1484.797	-2
6530	fFKeSDKWKKZ	81	372.2056	4	30.93	1484.797	-2.5
6533	fFKeSDKWKKZ	81	495.9384	3	30.95	1484.797	-2.4
6543	fFKeSDKWKKZ	81	372.2056	4	30.99	1484.797	-2.3
6546	fFKeSDKWKKZ	81	495.9384	3	31	1484.797	-2.4
6662	TNFLGfKWKKZ	81	672.3832	2	31.47	1342.755	-2.2
6672	TNFLGfKWKKZ	81	336.6954	4	31.51	1342.755	-1.9
7078	gTEgfPKWKKZ	81	348.9647	4	33.13	1391.833	-2.5
7098	gTEgfPKWKKZ	81	348.9647	4	33.21	1391.833	-2.3
7111	gTEgfPKWKKZ	81	348.9646	4	33.26	1391.833	-2.8
7124	gTEgfPKWKKZ	81	348.9646	4	33.31	1391.833	-2.8
7312	gTEgfPKWKKZ	81	464.9504	3	34.05	1391.833	-2.5
7333	PfgghPKWKKZ	81	448.9557	3	34.14	1343.848	-2.4
7554	gbDeDfKWKKZ	81	373.7082	4	35.06	1490.807	-2.4
7567	gbDeDfKWKKZ	81	373.7083	4	35.11	1490.807	-2.2
7586	gbDeDfKWKKZ	81	497.9421	3	35.19	1490.807	-2
7599	gbDeDfKWKKZ	81	373.7082	4	35.24	1490.807	-2.4
7924	egTQeEKWKKZ	81	367.7083	4	36.55	1466.807	-2.4
7927	egTQeEKWKKZ	81	489.9422	3	36.57	1466.807	-1.8
7937	egTQeEKWKKZ	81	367.7083	4	36.61	1466.807	-2.4
7943	egTQeEKWKKZ	81	489.942	3	36.63	1466.807	-2.1
7963	egTQeEKWKKZ	81	367.7083	4	36.71	1466.807	-2.3
7999	egTQeEKWKKZ	81	367.7084	4	36.85	1466.807	-2
8067	egTQeEKWKKZ	81	367.7082	4	37.13	1466.807	-2.5
8421	DTALffKWKKZ	81	717.8982	2	38.65	1433.786	-2.8
8445	DTALffKWKKZ	81	717.8982	2	38.76	1433.786	-2.8

8562	NHETeGKWKKZ	81	441.8998	3	39.26	1322.688	-8
8624	TbTQeEKWKKZ	81	457.5863	3	39.52	1369.751	-10
8645	TbTQeEKWKKZ	81	457.5858	3	39.61	1369.751	-11.2
8811	FFNeAeKWKKZ	81	366.2002	4	40.35	1460.776	-2.6
8829	FFNeAeKWKKZ	81	366.2001	4	40.43	1460.776	-2.9
10311	HQNfgAKWKKZ	81	463.5998	3	46.87	1387.788	-7.5
11145	eLhPLdKWKKZ	81	457.2876	3	50.55	1368.843	-1.7
2893	LAhSNDKWKKZ	80	294.1734	4	15.79	1172.667	-1.8
3781	gKLESEKWKKZ	80	434.6019	3	19.66	1300.787	-2
4261	dTKLQfKWKKZ	80	365.2239	4	21.73	1456.871	-2.8
7491	fAfFAGKWKKZ	80	460.9241	3	34.8	1379.754	-2.7
8490	ASSSggKWKKZ	80	392.2487	3	38.96	1173.723	0.6
9198	ESiSfGKKKKZ	80	422.9051	3	42	1265.688	4.2
9914	aGiEegKWKKZ	80	451.2632	3	45.13	1350.756	8.5
3419	hFEEADKWKKZ	80	316.9219	4	18.11	1263.661	-2
3471	KFSLEaKWKKZ	80	323.9509	4	18.33	1291.776	-1.6
3782	gKLESEKWKKZ	80	434.6019	3	19.66	1300.787	-2
4140	KaTfQLKWKKZ	80	346.2164	4	21.21	1380.839	-2
4305	AhAfgGKWKKZ	80	408.2562	3	21.91	1221.739	6.5
4330	AhAfgGKWKKZ	80	408.2564	3	22.02	1221.739	7
4346	AhAfgGKWKKZ	80	408.2564	3	22.08	1221.739	7
4509	bgRdGfKWKKZ	80	364.2282	4	22.73	1452.887	-2.2
4531	fhNRLTKWKKZ	80	345.9611	4	22.82	1379.819	-2.8
4609	bgRdGfKWKKZ	80	364.2281	4	23.14	1452.887	-2.6
4822	TRfQAFKWKKZ	80	472.2743	3	24.03	1413.803	-1.5
4839	TRfQAFKWKKZ	80	354.4574	4	24.09	1413.803	-1.9
4848	TRfQAFKWKKZ	80	472.274	3	24.13	1413.803	-2.1
4861	TRfQAFKWKKZ	80	472.2744	3	24.19	1413.803	-1.3
4871	TRfQAFKWKKZ	80	354.4573	4	24.23	1413.803	-2.1
5162	KgfGTPKWKKZ	80	331.2084	4	25.4	1320.807	-2
5389	fFKFNTKWKKZ	80	362.9596	4	26.3	1447.813	-2.5
5428	fFKFNTKWKKZ	80	362.96	4	26.45	1447.813	-1.4
5636	PSFfEAKWKKZ	80	448.2477	3	27.31	1341.723	-1.5
5862	fQhhbfKWKKZ	80	365.4646	4	28.22	1457.834	-3
5915	ffKHgTKWKKZ	80	382.7294	4	28.44	1526.891	-1.9
6116	SaeDfQKWKKZ	80	480.258	3	29.28	1437.756	-2.5
6489	AGPhfeKWKKZ	80	440.2528	3	30.77	1317.739	-1.5
6544	fFKeSDKWKKZ	80	372.2056	4	30.99	1484.797	-2.3
6559	fFKeSDKWKKZ	80	495.9389	3	31.05	1484.797	-1.4
7169	gTEgfPKWKKZ	80	348.9647	4	33.49	1391.833	-2.4
7280	PfgghPKWKKZ	80	448.9557	3	33.92	1343.848	-2.4
7309	PfgghPKWKKZ	80	448.9557	3	34.04	1343.848	-2.4
7544	gbDeDfKWKKZ	80	373.7083	4	35.02	1490.807	-2.2

7816	egTQeEKWKKZ	80	489.9422	3	36.12	1466.807	-1.8
7842	egTQeEKWKKZ	80	489.9419	3	36.23	1466.807	-2.5
7855	egTQeEKWKKZ	80	489.942	3	36.28	1466.807	-2.2
7869	egTQeEKWKKZ	80	367.7081	4	36.34	1466.807	-2.9
7881	egTQeEKWKKZ	80	489.942	3	36.38	1466.807	-2.3
7894	egTQeEKWKKZ	80	489.942	3	36.44	1466.807	-2.2
7898	egTQeEKWKKZ	80	367.7082	4	36.45	1466.807	-2.5
7904	egTQeEKWKKZ	80	489.9421	3	36.47	1466.807	-2
7917	egTQeEKWKKZ	80	489.942	3	36.53	1466.807	-2.2
7953	egTQeEKWKKZ	80	489.9421	3	36.67	1466.807	-1.9
7966	egTQeEKWKKZ	80	489.9421	3	36.72	1466.807	-2
7975	egTQeEKWKKZ	80	367.7083	4	36.75	1466.807	-2.4
7976	egTQeEKWKKZ	80	367.7083	4	36.76	1466.807	-2.4
8005	egTQeEKWKKZ	80	489.942	3	36.87	1466.807	-2.1
8038	egTQeEKWKKZ	80	367.7082	4	37.01	1466.807	-2.5
8455	DTALffKWKKZ	80	717.8981	2	38.8	1433.786	-2.9
8485	hfFgSgKWKKZ	80	346.9674	4	38.94	1383.843	-2.1
8540	NEHTeGKWKKZ	80	441.9	3	39.17	1322.688	-7.7
8633	TbTQeEKWKKZ	80	457.5858	3	39.56	1369.751	-11
8655	TbTQeEKWKKZ	80	457.586	3	39.65	1369.751	-10.7
8657	TbTQeEKWKKZ	80	457.586	3	39.66	1369.751	-10.7
8669	TbTQeEKWKKZ	80	457.586	3	39.71	1369.751	-10.7
8717	LNTQeEKWKKZ	80	457.5858	3	39.93	1369.751	-10.8
8799	FFNeAeKWKKZ	80	366.2003	4	40.3	1460.776	-2.5
9451	fALEegKWKKZ	80	483.619	3	43.09	1447.838	-1.8
2851	AhLSQQKWKKZ	79	400.9111	3	15.62	1199.714	-2
2929	hLASNDKWKKZ	79	294.1733	4	15.95	1172.667	-1.9
3886	fEPAHGKWKKZ	79	326.4325	4	20.11	1301.703	-1.8
4538	GGgRLFKWKKZ	79	415.9354	3	22.85	1244.787	-2.3
4736	fFKFDaKWKKZ	79	362.9596	4	23.67	1447.813	-2.4
2869	AhLSQQKWKKZ	79	300.9353	4	15.69	1199.714	-1.4
3147	SFEDGGKWKKZ	79	394.2074	3	16.9	1179.604	-2.7
3453	KFSLEaKWKKZ	79	323.9507	4	18.25	1291.776	-1.9
3483	KFSLEaKWKKZ	79	323.9506	4	18.38	1291.776	-2.5
4262	dTKLQfKWKKZ	79	365.2239	4	21.73	1456.871	-2.8
4317	AhAfgGKWKKZ	79	408.256	3	21.96	1221.739	6.2
4567	bgRdGfKWKKZ	79	364.2283	4	22.97	1452.887	-2.1
4661	bgRdGfKWKKZ	79	364.2283	4	23.36	1452.887	-2
4737	fFKFDaKWKKZ	79	362.9596	4	23.67	1447.813	-2.4
4920	FLAhaeKWKKZ	79	326.2018	4	24.43	1300.781	-2.1
5015	PKLDFLKWKKZ	79	434.6068	3	24.8	1300.802	-2.5
5188	gaFQRfKWKKZ	79	368.2266	4	25.5	1468.882	-3.1
5205	KgfGTPKWKKZ	79	331.2082	4	25.57	1320.807	-2.7

5218	KgfGTPKWKKZ	79	331.2082	4	25.62	1320.807	-2.5
5273	gFgNTQKWKKZ	79	333.9584	4	25.83	1331.808	-2.4
5324	fFKFNTKWKKZ	79	362.9596	4	26.04	1447.813	-2.3
5334	fFKFNTKWKKZ	79	362.9598	4	26.08	1447.813	-1.7
5347	fFKFNTKWKKZ	79	362.9598	4	26.13	1447.813	-1.9
5360	fFKFNTKWKKZ	79	362.9596	4	26.18	1447.813	-2.3
5376	fFKFNTKWKKZ	79	362.9595	4	26.24	1447.813	-2.6
5402	fFKFNTKWKKZ	79	362.9596	4	26.35	1447.813	-2.4
5415	fFKFNTKWKKZ	79	362.9598	4	26.4	1447.813	-1.8
5461	FFNRhfKWKKZ	79	365.9625	4	26.59	1459.824	-2.2
5487	fNTRfNKWKKZ	79	380.7127	4	26.7	1518.825	-2
5500	fNTRfNKWKKZ	79	380.7127	4	26.75	1518.825	-1.9
5503	fNTRfNKWKKZ	79	507.2813	3	26.76	1518.825	-1.7
5513	fNTRfNKWKKZ	79	380.7126	4	26.81	1518.825	-2.1
5529	fNTRfNKWKKZ	79	380.7125	4	26.87	1518.825	-2.4
5633	PSFfEAKWKKZ	79	336.4375	4	27.3	1341.723	-1.9
5811	fQhhbfKWKKZ	79	365.4648	4	28.02	1457.834	-2.6
5837	fQhhbfKWKKZ	79	365.4647	4	28.12	1457.834	-2.7
6106	DfdDfKKWKKZ	79	392.9647	4	29.24	1567.834	-2.7
6144	SaeDfQKWKKZ	79	480.2581	3	29.39	1437.756	-2.2
6556	fFKeSDKWKKZ	79	372.2065	4	31.04	1484.797	0
6592	SNFLGfKWKKZ	79	443.9192	3	31.19	1328.739	-2.6
7085	gTEgfPKWKKZ	79	348.9647	4	33.16	1391.833	-2.4
7270	PfgghPKWKKZ	79	448.9556	3	33.88	1343.848	-2.5
7299	PfgghPKWKKZ	79	448.9557	3	34	1343.848	-2.2
7806	egTQeEKWKKZ	79	489.9422	3	36.08	1466.807	-1.7
7829	egTQeEKWKKZ	79	489.9422	3	36.17	1466.807	-1.8
7852	egTQeEKWKKZ	79	367.7082	4	36.27	1466.807	-2.4
7868	egTQeEKWKKZ	79	367.7081	4	36.33	1466.807	-2.9
7897	egTQeEKWKKZ	79	367.7082	4	36.45	1466.807	-2.5
7936	egTQeEKWKKZ	79	367.7083	4	36.6	1466.807	-2.4
7949	egTQeEKWKKZ	79	367.7082	4	36.65	1466.807	-2.6
7962	egTQeEKWKKZ	79	367.7083	4	36.7	1466.807	-2.3
7979	egTQeEKWKKZ	79	489.9424	3	36.77	1466.807	-1.3
7985	egTQeEKWKKZ	79	367.7085	4	36.8	1466.807	-1.8
7992	egTQeEKWKKZ	79	489.9423	3	36.82	1466.807	-1.5
8034	egTQeEKWKKZ	79	489.942	3	36.99	1466.807	-2.3
8047	egTQeEKWKKZ	79	489.9418	3	37.04	1466.807	-2.6
8050	egTQeEKWKKZ	79	367.7082	4	37.06	1466.807	-2.6
8103	egTQeEKWKKZ	79	489.9421	3	37.28	1466.807	-2
8104	egTQeEKWKKZ	79	489.9421	3	37.28	1466.807	-2
8574	ENHTeGKWKKZ	79	441.8998	3	39.31	1322.688	-8.1
8735	TbTQeEKWKKZ	79	457.5859	3	40.01	1369.751	-10.9

9249	ESiSfGKKKKZ	79	422.9048	3	42.22	1265.688	3.7
9527	FffbFfKWKKZ	79	420.233	4	43.42	1676.906	-1.8
10033	aGiEegKWKKZ	79	451.263	3	45.67	1350.756	8.2
10035	aGiEegKWKKZ	79	451.263	3	45.68	1350.756	8.2
11147	eLhPLdKWKKZ	79	457.2876	3	50.56	1368.843	-1.7
4069	LQREQeKWKKZ	78	360.7112	4	20.91	1438.82	-2.8
4225	LEFTDTKWKKZ	78	432.2424	3	21.58	1293.708	-2
5626	EgTGAeKWKKZ	78	424.2477	3	27.27	1269.723	-1.6
9928	bSPEegKWKKZ	78	451.2633	3	45.21	1350.781	-9.8
9940	QPhEegKWKKZ	78	451.2632	3	45.26	1350.781	-10
3432	KFSLEaKWKKZ	78	323.9508	4	18.16	1291.776	-1.6
3447	KFSLEaKWKKZ	78	323.9506	4	18.23	1291.776	-2.3
3804	KGEgFdKWKKZ	78	338.9598	4	19.76	1351.813	-1.9
3897	AAKDLFKWKKZ	78	309.1914	4	20.16	1232.739	-2.1
4176	KaTfQLKWKKZ	78	346.2163	4	21.37	1380.839	-2.3
4226	LEFTDTKWKKZ	78	432.2424	3	21.58	1293.708	-2
4936	dAQgfGKWKKZ	78	343.4571	4	24.49	1369.802	-2.2
4985	dAQgfGKWKKZ	78	343.4569	4	24.68	1369.802	-2.7
5175	KgfGTPKWKKZ	78	331.2082	4	25.45	1320.807	-2.7
5179	gKfGTPKWKKZ	78	331.2082	4	25.46	1320.807	-2.6
5247	KgfGTPKWKKZ	78	331.2082	4	25.73	1320.807	-2.5
5441	fFKFNTKWKKZ	78	362.9598	4	26.51	1447.813	-1.9
5532	fNTRfNKWKKZ	78	507.2809	3	26.88	1518.825	-2.5
5542	fNTRfNKWKKZ	78	380.7127	4	26.92	1518.825	-2
5571	fNTRfNKWKKZ	78	380.7126	4	27.04	1518.825	-2.1
5647	PSFfEAKWKKZ	78	448.2476	3	27.36	1341.723	-1.8
5788	fQhhbfKWKKZ	78	365.465	4	27.92	1457.834	-2
5824	fQhhbfKWKKZ	78	365.4651	4	28.07	1457.834	-1.8
5874	fQhhbfKWKKZ	78	365.4648	4	28.27	1457.834	-2.6
6104	SaeDfQKWKKZ	78	480.258	3	29.23	1437.756	-2.4
6154	SaeDfQKWKKZ	78	480.2581	3	29.43	1437.756	-2.2
6219	FSfGgRKWKKZ	78	462.6107	3	29.69	1384.813	-2.2
6572	fFKeSDKWKKZ	78	372.2076	4	31.1	1484.797	2.8
6685	afNfgAKWKKZ	78	482.9507	3	31.56	1445.834	-2.2
6805	afNfgAKWKKZ	78	482.9505	3	32.03	1445.834	-2.6
7052	gTEgfPKWKKZ	78	348.9648	4	33.02	1391.833	-2.2
7140	gTEgfPKWKKZ	78	348.9646	4	33.37	1391.833	-2.6
7153	gTEgfPKWKKZ	78	348.9646	4	33.42	1391.833	-2.8
7319	PfgghPKWKKZ	78	448.9557	3	34.08	1343.848	-2.2
7371	PfgghPKWKKZ	78	336.9685	4	34.29	1343.848	-2.6
7506	fAfFAGKWKKZ	78	460.9243	3	34.86	1379.754	-2.2
7884	egTQeEKWKKZ	78	367.7082	4	36.4	1466.807	-2.6
7910	egTQeEKWKKZ	78	367.7083	4	36.5	1466.807	-2.4

7923	egTQeEKWKKZ	78	367.7083	4	36.55	1466.807	-2.4
7998	egTQeEKWKKZ	78	367.7084	4	36.85	1466.807	-2
8011	egTQeEKWKKZ	78	367.7082	4	36.9	1466.807	-2.4
8012	egTQeEKWKKZ	78	367.7082	4	36.9	1466.807	-2.4
8037	egTQeEKWKKZ	78	367.7082	4	37	1466.807	-2.5
8066	egTQeEKWKKZ	78	367.7082	4	37.12	1466.807	-2.5
8073	egTQeEKWKKZ	78	489.942	3	37.15	1466.807	-2.1
8552	NEHTeGKWKKZ	78	441.9	3	39.22	1322.688	-7.7
8890	fegQbFKWKKZ	78	384.9765	4	40.7	1535.881	-2.4
9203	ePLEegKWKKZ	78	362.9661	4	42.02	1447.838	-1.9
9324	ePLEegKWKKZ	78	362.9661	4	42.54	1447.838	-1.9
9916	aGiEegKWKKZ	78	451.2632	3	45.15	1350.756	8.5
9975	QPhEegKWKKZ	78	451.263	3	45.42	1350.781	-10.5
3790	SRNfhGKWKKZ	77	328.4417	4	19.7	1309.741	-2.2
4033	fTNETSKWKKZ	77	336.6824	4	20.75	1342.703	-2.1
5469	NgPeASKWKKZ	77	427.9195	3	26.63	1280.739	-2.1
6349	bGGFgfKWKKZ	77	442.6001	3	30.21	1324.781	-1.9
6476	hPAGfeKWKKZ	77	440.2523	3	30.72	1317.739	-2.5
7046	GaSLEfKWKKZ	77	433.2525	3	33	1296.734	1.2
8258	hQhLgQKWKKZ	77	418.935	3	37.94	1253.797	-11.4
3813	GKEgFdKWKKZ	77	338.9597	4	19.8	1351.813	-2.2
3873	GKEgFdKWKKZ	77	338.9599	4	20.06	1351.813	-1.7
3924	AAKDLFKWKKZ	77	309.191	4	20.28	1232.739	-3.4
4127	AKQfFNKWKKZ	77	467.2703	3	21.16	1398.793	-2.4
4314	AhAfgGKWKKZ	77	306.444	4	21.95	1221.739	6.8
4411	AhAfgGKWKKZ	77	306.444	4	22.34	1221.739	6.7
4525	GGgRLFKWKKZ	77	415.9353	3	22.8	1244.787	-2.5
4561	GGgRLFKWKKZ	77	415.9354	3	22.95	1244.787	-2.3
4857	TRfQAFKWKKZ	77	354.4575	4	24.17	1413.803	-1.8
4915	dNhgfGKWKKZ	77	343.4569	4	24.41	1369.802	-2.6
4923	dAQgfGKWKKZ	77	343.4571	4	24.44	1369.802	-2.3
4946	dAQgfGKWKKZ	77	343.4573	4	24.53	1369.802	-1.5
4956	dAQgfGKWKKZ	77	343.4573	4	24.57	1369.802	-1.6
5003	NFDSfdKWKKZ	77	363.4454	4	24.75	1449.756	-2.2
5192	gKfGTPKWKKZ	77	331.2083	4	25.51	1320.807	-2.3
5234	gKfGTPKWKKZ	77	331.2084	4	25.68	1320.807	-2
5521	NgPeASKWKKZ	77	427.9197	3	26.84	1280.739	-1.5
5889	fQhhbfKWKKZ	77	365.4648	4	28.34	1457.834	-2.5
5979	DfdDfKKWKKZ	77	392.9649	4	28.71	1567.834	-2.2
6002	DfdDfKKWKKZ	77	392.9647	4	28.8	1567.834	-2.8
6022	DfdDfKKWKKZ	77	392.9651	4	28.88	1567.834	-1.6
6077	DfdDfKKWKKZ	77	392.9649	4	29.11	1567.834	-2.2
6095	SaeDfQKWKKZ	77	360.4456	4	29.19	1437.756	-1.7

6101	SaeDfQKWKKZ	77	360.4453	4	29.21	1437.756	-2.4
6123	FfQEhLKWKKZ	77	354.2056	4	29.31	1412.797	-2.6
6160	FSfGgRKWKKZ	77	347.2096	4	29.46	1384.813	-2.9
6164	FfQEhLKWKKZ	77	354.2052	4	29.47	1412.797	-3.7
6675	afNfgAKWKKZ	77	482.9508	3	31.52	1445.834	-1.9
6815	afNfgAKWKKZ	77	362.4648	4	32.07	1445.834	-2.5
7865	egTQeEKWKKZ	77	489.9418	3	36.32	1466.807	-2.6
8024	egTQeEKWKKZ	77	367.7084	4	36.95	1466.807	-2
8025	egTQeEKWKKZ	77	367.7084	4	36.96	1466.807	-2
8060	egTQeEKWKKZ	77	489.942	3	37.1	1466.807	-2.1
8088	egTQeEKWKKZ	77	489.9422	3	37.21	1466.807	-1.8
8423	DTALffKWKKZ	77	717.8982	2	38.66	1433.786	-2.8
8492	ASSSggKWKKZ	77	392.2487	3	38.97	1173.723	0.6
8526	ASSSggKWKKZ	77	392.2489	3	39.11	1173.723	1.3
8564	NHETeGKWKKZ	77	441.8998	3	39.27	1322.688	-8
8866	fegQbFKWKKZ	77	512.9659	3	40.59	1535.881	-3
9186	ESiSfGKKKKZ	77	422.905	3	41.95	1265.688	4.1
9261	EiSSfGKKKKZ	77	422.905	3	42.27	1265.688	4
10099	bSPEegKWKKZ	77	451.263	3	45.96	1350.781	-10.5
10313	QHNfgAKWKKZ	77	463.5998	3	46.88	1387.788	-7.5
2968	RNESHFKWKKZ	76	340.4408	4	16.12	1357.737	-2
3292	QQSPLGKWKKZ	76	400.2391	3	17.55	1197.698	-2.4
8361	NGHNAfKWKKZ	76	435.5683	3	38.39	1303.694	-8.2
2956	hLASNDKWKKZ	76	294.1734	4	16.07	1172.667	-1.7
3876	AAKDLFKWKKZ	76	309.1915	4	20.07	1232.739	-1.9
3960	RKLgKLKWKKZ	76	339.2438	4	20.43	1352.95	-2.7
4065	GGaLgLKWKKZ	76	289.6978	4	20.89	1154.765	-2.7
4134	AKQfFNKWKKZ	76	350.7045	4	21.19	1398.793	-2.6
4161	AKQfFNKWKKZ	76	350.7046	4	21.3	1398.793	-2.3
4302	AhAfgGKWKKZ	76	306.444	4	21.9	1221.739	6.8
4340	AhAfgGKWKKZ	76	306.444	4	22.06	1221.739	6.8
4366	AhAfgGKWKKZ	76	306.4441	4	22.16	1221.739	7
4379	AhAfgGKWKKZ	76	306.4438	4	22.22	1221.739	6.2
4387	bgRdGfKWKKZ	76	364.2281	4	22.25	1452.887	-2.6
4395	AhAfgGKWKKZ	76	306.4439	4	22.28	1221.739	6.4
4397	bgRdGfKWKKZ	76	364.2281	4	22.29	1452.887	-2.6
4407	bgRdGfKWKKZ	76	364.2282	4	22.33	1452.887	-2.3
4417	bgRdGfKWKKZ	76	364.2282	4	22.37	1452.887	-2.1
4430	bgRdGfKWKKZ	76	364.2284	4	22.42	1452.887	-1.6
4443	bgRdGfKWKKZ	76	364.2282	4	22.47	1452.887	-2.3
4456	bgRdGfKWKKZ	76	364.2283	4	22.52	1452.887	-2.1
4482	bgRdGfKWKKZ	76	364.2281	4	22.62	1452.887	-2.6
4508	bgRdGfKWKKZ	76	364.2282	4	22.73	1452.887	-2.2

4521	bgRdGfKWKKZ	76	364.2281	4	22.78	1452.887	-2.5
4534	bgRdGfKWKKZ	76	364.2282	4	22.83	1452.887	-2.1
4566	bgRdGfKWKKZ	76	364.2283	4	22.97	1452.887	-2.1
4576	bgRdGfKWKKZ	76	364.2282	4	23.01	1452.887	-2.2
4595	bgRdGfKWKKZ	76	364.2282	4	23.09	1452.887	-2.3
4660	bgRdGfKWKKZ	76	364.2283	4	23.36	1452.887	-2
4749	EAfaLGKWKKZ	76	427.9193	3	23.72	1280.739	-2.5
4752	EAfaLGKWKKZ	76	427.9192	3	23.74	1280.739	-2.7
4767	EAfaLGKWKKZ	76	427.9195	3	23.8	1280.739	-2
4797	EAfaLGKWKKZ	76	427.9194	3	23.92	1280.739	-2.2
4966	FKeNFKKWKKZ	76	363.2178	4	24.61	1448.845	-1.6
4969	dAQgfGKWKKZ	76	343.4571	4	24.62	1369.802	-2.3
5422	eLhPLdKWKKZ	76	343.2176	4	26.43	1368.843	-1.6
5481	NgPeASKWKKZ	76	427.9196	3	26.68	1280.739	-1.9
5494	NgPeASKWKKZ	76	427.9196	3	26.73	1280.739	-1.9
5555	fNTRfNKWKKZ	76	380.7127	4	26.98	1518.825	-2
5838	fQhhbfKWKKZ	76	486.9506	3	28.12	1457.834	-2.5
5866	fQhhbfKWKKZ	76	486.9503	3	28.24	1457.834	-3.1
6012	DfdDfKKWKKZ	76	392.9652	4	28.84	1567.834	-1.5
6051	DfdDfKKWKKZ	76	392.9648	4	29	1567.834	-2.5
6064	DfdDfKKWKKZ	76	392.9647	4	29.06	1567.834	-2.7
6092	DfdDfKKWKKZ	76	392.9651	4	29.18	1567.834	-1.8
6125	SaeDfQKWKKZ	76	360.445	4	29.31	1437.756	-3.3
6173	FSfGgRKWKKZ	76	347.2097	4	29.51	1384.813	-2.5
6186	FSfGgRKWKKZ	76	347.2097	4	29.56	1384.813	-2.5
6212	FSfGgRKWKKZ	76	347.2097	4	29.67	1384.813	-2.4
6359	bGGFgfKWKKZ	76	332.2018	4	30.26	1324.781	-2.3
6512	hPAGfeKWKKZ	76	440.2526	3	30.87	1317.739	-2
6659	afNfgAKWKKZ	76	362.4648	4	31.46	1445.834	-2.3
6740	afNfgAKWKKZ	76	482.9507	3	31.78	1445.834	-2.3
6776	afNfgAKWKKZ	76	362.4648	4	31.92	1445.834	-2.4
7213	PfgghPKWKKZ	76	336.9685	4	33.66	1343.848	-2.5
7335	PfgghPKWKKZ	76	448.9557	3	34.14	1343.848	-2.4
8021	egTQeEKWKKZ	76	489.9423	3	36.94	1466.807	-1.5
8475	DTALffKWKKZ	76	717.8991	2	38.9	1433.786	-1.5
8477	DTALffKWKKZ	76	717.8991	2	38.9	1433.786	-1.5
8784	fegQbFKWKKZ	76	512.9662	3	40.23	1535.881	-2.4
8785	fegQbFKWKKZ	76	512.9662	3	40.24	1535.881	-2.4
8842	fegQbFKWKKZ	76	512.9663	3	40.48	1535.881	-2.3
8854	fegQbFKWKKZ	76	512.9662	3	40.54	1535.881	-2.5
8875	fegQbFKWKKZ	76	512.9663	3	40.63	1535.881	-2.3
8899	fegQbFKWKKZ	76	512.9667	3	40.74	1535.881	-1.5
9443	FffbFfKWKKZ	76	420.233	4	43.06	1676.906	-1.8

9600	AfLEegKWKKZ	76	483.6192	3	43.75	1447.838	-1.5
9601	AfLEegKWKKZ	76	483.6192	3	43.75	1447.838	-1.5
9646	AfLEegKWKKZ	76	483.6187	3	43.95	1447.838	-2.5
9963	bSPEegKWKKZ	76	451.2633	3	45.36	1350.781	-9.7
9987	bSPEegKWKKZ	76	451.2633	3	45.47	1350.781	-9.8
10011	bSPEegKWKKZ	76	451.2632	3	45.57	1350.781	-9.9
10047	bSPEegKWKKZ	76	451.2632	3	45.73	1350.781	-10
10086	bSPEegKWKKZ	76	451.263	3	45.9	1350.781	-10.4
10158	NHNfgAKWKKZ	76	458.928	3	46.23	1373.772	-7
5643	AGAgfGKWKKZ	75	398.9089	3	27.34	1193.707	-2.1
6071	bHDFRDKWKKZ	75	462.2582	3	29.09	1383.752	0.3
2848	AhLSQQKWKKZ	75	300.9352	4	15.6	1199.714	-1.9
3146	SFEDGGKWKKZ	75	394.2074	3	16.9	1179.604	-2.7
4034	fTNETSKWKKZ	75	336.6824	4	20.75	1342.703	-2.1
4104	AKQfFNKWKKZ	75	350.7045	4	21.06	1398.793	-2.6
4136	AKQfFNKWKKZ	75	467.2702	3	21.19	1398.793	-2.7
4151	AKQfFNKWKKZ	75	467.2706	3	21.26	1398.793	-1.7
4172	AgTLGNKWKKZ	75	391.2475	3	21.35	1170.724	-2.5
4184	AgTLGNKWKKZ	75	391.2477	3	21.4	1170.724	-2
4196	AgTLGNKWKKZ	75	391.2474	3	21.45	1170.724	-2.8
4247	KfGQdLKWKKZ	75	354.2176	4	21.67	1412.845	-2.2
4353	AhAfgGKWKKZ	75	306.4441	4	22.11	1221.739	7
4469	bgRdGfKWKKZ	75	364.2281	4	22.57	1452.887	-2.4
4495	bgRdGfKWKKZ	75	364.2281	4	22.68	1452.887	-2.5
4547	bgRdGfKWKKZ	75	364.2283	4	22.89	1452.887	-2.1
4593	GGgRLFKWKKZ	75	415.9353	3	23.08	1244.787	-2.5
4782	EAfaLGKWKKZ	75	427.9194	3	23.86	1280.739	-2.2
4993	begDTbKWKKZ	75	345.9583	4	24.71	1379.808	-2.9
5127	KTfhKfKWKKZ	75	369.9763	4	25.26	1475.881	-2.9
5690	ffKHgTKWKKZ	75	382.7295	4	27.53	1526.891	-1.7
5697	ffKHgTKWKKZ	75	382.7293	4	27.56	1526.891	-2
5831	fQLGbfKWKKZ	75	486.9511	3	28.1	1457.834	-1.3
5868	fQhhbfKWKKZ	75	486.9503	3	28.25	1457.834	-3.1
6038	DfdDfKKWKKZ	75	392.9648	4	28.95	1567.834	-2.6
6110	SaeDfQKWKKZ	75	360.4451	4	29.25	1437.756	-2.9
6131	SaeDfQKWKKZ	75	480.258	3	29.34	1437.756	-2.3
6136	FfQEhLKWKKZ	75	354.2056	4	29.36	1412.797	-2.5
6141	SaeDfQKWKKZ	75	360.4452	4	29.38	1437.756	-2.7
6149	FfQEhLKWKKZ	75	354.2056	4	29.41	1412.797	-2.5
6374	eTFLGEKWKKZ	75	444.9195	3	30.31	1331.739	-1.7
6499	hPAGfeKWKKZ	75	440.2524	3	30.81	1317.739	-2.4
6799	afNfgAKWKKZ	75	362.4647	4	32.01	1445.834	-2.7
6893	afNfLhKWKKZ	75	362.4649	4	32.38	1445.834	-2.1

7036	GaSLEfKWKKZ	75	433.2524	3	32.96	1296.734	1
7062	GaSLEfKWKKZ	75	433.2523	3	33.06	1296.734	0.6
7075	GaSLEfKWKKZ	75	433.2522	3	33.12	1296.734	0.5
7091	GaSLEfKWKKZ	75	433.2524	3	33.18	1296.734	1
8531	NEHTeGKWKKZ	75	441.8999	3	39.13	1322.688	-7.8
8588	NEHTeGKWKKZ	75	441.8998	3	39.37	1322.688	-8.1
8758	fegQbFKWKKZ	75	512.9662	3	40.11	1535.881	-2.4
8794	fegQbFKWKKZ	75	384.9766	4	40.28	1535.881	-2.3
8796	fegQbFKWKKZ	75	512.9664	3	40.29	1535.881	-2.1
8797	fegQbFKWKKZ	75	512.9664	3	40.29	1535.881	-2.1
8803	fegQbFKWKKZ	75	512.9664	3	40.31	1535.881	-2.1
8818	fegQbFKWKKZ	75	512.9662	3	40.38	1535.881	-2.4
8887	fegQbFKWKKZ	75	512.9662	3	40.68	1535.881	-2.5
8914	fegQbFKWKKZ	75	512.9667	3	40.8	1535.881	-1.6
9460	fALEegKWKKZ	75	483.6194	3	43.13	1447.838	-1
9574	AfLEegKWKKZ	75	483.6191	3	43.63	1447.838	-1.6
9894	GfLfGKWWKKZ	75	483.2721	3	45.05	1446.796	-1.4
9951	bSPEegKWKKZ	75	451.2633	3	45.31	1350.781	-9.7
9999	bSPEegKWKKZ	75	451.2633	3	45.52	1350.781	-9.7
10023	bSPEegKWKKZ	75	451.2632	3	45.62	1350.781	-9.9
10101	bSPEegKWKKZ	75	451.263	3	45.97	1350.781	-10.5
10122	NHNfgAKWKKZ	75	458.928	3	46.07	1373.772	-7
10277	QHNfgAKWKKZ	75	463.5999	3	46.73	1387.788	-7.1
10301	QHNfgAKWKKZ	75	463.6002	3	46.83	1387.788	-6.6
3395	PGNgGNKWKKZ	74	289.4247	4	18	1153.672	-2.3
5052	bGHedaKWKKZ	74	461.2683	3	24.95	1380.793	-7.2
7370	GGieDfKWKKZ	74	348.4282	4	34.29	1389.698	-10.6
7387	PfgghPKWKKZ	74	336.9686	4	34.36	1343.848	-2.2
8276	TAbLgQKWKKZ	74	418.9352	3	38.02	1253.797	-10.7
8388	NHGANfKWKKZ	74	435.5684	3	38.51	1303.694	-8
10809	TFDfgAKWKKZ	74	458.2685	3	49.07	1371.77	9.7
10830	EHPfgAKWKKZ	74	458.2681	3	49.16	1371.782	0.8
2898	hLASNDKWKKZ	74	391.8955	3	15.82	1172.667	-1.6
2943	LAhSNDKWKKZ	74	294.1733	4	16.01	1172.667	-2
4236	NhEKLeKWKKZ	74	339.4546	4	21.62	1353.792	-1.9
4237	KfGQdLKWKKZ	74	354.2176	4	21.63	1412.845	-2.3
4238	KfGQdLKWKKZ	74	354.2176	4	21.63	1412.845	-2.3
4290	hAAfgGKWKKZ	74	306.4441	4	21.85	1221.739	7
4424	AhAfgGKWKKZ	74	306.4441	4	22.4	1221.739	6.9
4828	TRfQAFKWKKZ	74	354.4572	4	24.05	1413.803	-2.5
4976	PKLDFLKWKKZ	74	326.207	4	24.65	1300.802	-2.5
4983	PKLDFLKWKKZ	74	326.2068	4	24.67	1300.802	-2.9
4996	KPLDFLKWKKZ	74	326.207	4	24.72	1300.802	-2.4

5006	begDTbKWKKZ	74	345.9584	4	24.76	1379.808	-2.4
5038	PKLDFLKWKKZ	74	326.2072	4	24.89	1300.802	-1.9
5075	KTfhKfKWKKZ	74	369.9769	4	25.05	1475.881	-1.5
5114	KTfhKfKWKKZ	74	369.9766	4	25.2	1475.881	-2.3
5169	KTfhKfKWKKZ	74	369.9765	4	25.42	1475.881	-2.6
5519	fNTRfNKWKKZ	74	507.2813	3	26.83	1518.825	-1.7
5600	AGAgfGKWKKZ	74	398.9088	3	27.17	1193.707	-2.4
5833	fQLGbfKWKKZ	74	486.9511	3	28.1	1457.834	-1.3
5840	fQhhbfKWKKZ	74	486.9506	3	28.13	1457.834	-2.5
5853	fQhhbfKWKKZ	74	486.9512	3	28.19	1457.834	-1.3
6229	FSfGgRKWKKZ	74	347.2098	4	29.73	1384.813	-2.2
6390	FdDDfFKWKKZ	74	378.7004	4	30.38	1510.776	-2.5
6669	afNfgAKWKKZ	74	362.4649	4	31.5	1445.834	-2.1
6763	afNfgAKWKKZ	74	362.4647	4	31.87	1445.834	-2.6
6789	afNfgAKWKKZ	74	362.4648	4	31.97	1445.834	-2.4
6841	afNfLhKWKKZ	74	362.4648	4	32.18	1445.834	-2.2
6903	afNfgAKWKKZ	74	362.4647	4	32.42	1445.834	-2.6
6974	afNfLhKWKKZ	74	362.465	4	32.71	1445.834	-1.8
7301	PfgghPKWKKZ	74	336.9685	4	34.01	1343.848	-2.6
8127	egTQeEKWKKZ	74	489.9424	3	37.38	1466.807	-1.4
8128	egTQeEKWKKZ	74	489.9424	3	37.38	1466.807	-1.4
8187	egTQeEKWKKZ	74	367.7083	4	37.63	1466.807	-2.3
8293	PLLEegKWKKZ	74	455.6191	3	38.1	1363.838	-1.8
8337	NGHNAfKWKKZ	74	435.5683	3	38.29	1303.694	-8.3
8363	NGHNAfKWKKZ	74	435.5683	3	38.4	1303.694	-8.2
8467	EgTeQeKWKKZ	74	367.7081	4	38.86	1466.807	-2.8
8528	ASSSggKWKKZ	74	392.2489	3	39.12	1173.723	1.3
8770	fegQbFKWKKZ	74	512.9662	3	40.17	1535.881	-2.4
9220	fALEegKWKKZ	74	483.6188	3	42.09	1447.838	-2.3
9403	AfLEegKWKKZ	74	483.6191	3	42.88	1447.838	-1.6
9588	AfLEegKWKKZ	74	483.6191	3	43.69	1447.838	-1.6
9589	AfLEegKWKKZ	74	483.6191	3	43.7	1447.838	-1.6
9624	fALEegKWKKZ	74	483.6191	3	43.85	1447.838	-1.6
9625	fALEegKWKKZ	74	483.6191	3	43.85	1447.838	-1.6
10059	bSPEegKWKKZ	74	451.2629	3	45.78	1350.781	-10.6
10134	NHNfgAKWKKZ	74	458.9277	3	46.12	1373.772	-7.7
10143	NHNfgAKWKKZ	74	458.9277	3	46.16	1373.772	-7.8
10289	QHNfgAKWKKZ	74	463.5997	3	46.78	1387.788	-7.5
3780	gPLShAKWKKZ	73	390.2599	3	19.65	1167.749	7.3
4941	fTAhfNKWKKZ	73	469.2668	3	24.51	1404.771	5.6
9768	hfFgSfKWKKZ	73	494.2876	3	44.5	1479.843	-1.6
10818	HPEfgAKWKKZ	73	458.2684	3	49.11	1371.782	1.4
3409	hFEEADKWKKZ	73	316.9222	4	18.06	1263.661	-1.1

4122	AKQfFNKWKKZ	73	350.7045	4	21.14	1398.793	-2.5
4146	AKQfFNKWKKZ	73	350.7047	4	21.24	1398.793	-2.1
4148	AgTLGNKWKKZ	73	293.6877	4	21.25	1170.724	-1.8
4181	AgTLGNKWKKZ	73	293.6876	4	21.39	1170.724	-2
4213	LhTLGNKWKKZ	73	293.6875	4	21.53	1170.724	-2.3
4214	LhTLGNKWKKZ	73	293.6875	4	21.53	1170.724	-2.3
4608	bgRdGfKWKKZ	73	364.2281	4	23.14	1452.887	-2.6
4795	TRfQAFKWKKZ	73	354.4572	4	23.91	1413.803	-2.4
4802	TRfQAFKWKKZ	73	354.4572	4	23.94	1413.803	-2.4
4815	TRfQAFKWKKZ	73	354.4574	4	24	1413.803	-1.8
4824	TRfQAFKWKKZ	73	472.2743	3	24.03	1413.803	-1.5
4837	TRfhNFKWKKZ	73	472.2737	3	24.09	1413.804	-2.9
4841	TRfQAFKWKKZ	73	354.4574	4	24.1	1413.803	-1.9
4863	TRfQAFKWKKZ	73	472.2744	3	24.19	1413.803	-1.3
4979	FKeNFKKWKKZ	73	363.2176	4	24.66	1448.845	-2.3
5009	PKLDFLKWKKZ	73	326.207	4	24.78	1300.802	-2.4
5022	PKLDFLKWKKZ	73	326.2069	4	24.83	1300.802	-2.6
5042	bGHedaKWKKZ	73	461.2685	3	24.91	1380.793	-6.8
5088	KTfhKfKWKKZ	73	369.9765	4	25.1	1475.881	-2.4
5187	gaFQRfKWKKZ	73	368.2266	4	25.49	1468.882	-3.1
5652	hfFgSbKWKKZ	73	346.7124	4	27.38	1382.823	-1.7
5688	fgAGbFKWKKZ	73	335.7057	4	27.52	1338.797	-2
5701	fgAGbFKWKKZ	73	335.7059	4	27.58	1338.797	-1.7
5855	fQhhbfKWKKZ	73	486.9512	3	28.19	1457.834	-1.3
5989	DfdDfKKWKKZ	73	392.9648	4	28.75	1567.834	-2.5
6107	DfdDfKKWKKZ	73	392.9647	4	29.24	1567.834	-2.7
6199	FSfGgRKWKKZ	73	347.2097	4	29.62	1384.813	-2.7
6527	eAFfdEKWKKZ	73	511.2785	3	30.92	1530.817	-2.4
6582	AeFfdEKWKKZ	73	511.2785	3	31.14	1530.817	-2.5
6594	fGLHfGKWKKZ	73	350.4503	4	31.2	1397.776	-2.8
6633	afNfgAKWKKZ	73	362.4647	4	31.35	1445.834	-2.7
6646	afNfgAKWKKZ	73	362.4648	4	31.41	1445.834	-2.4
6698	afNfgAKWKKZ	73	482.9506	3	31.61	1445.834	-2.5
6705	afNfLhKWKKZ	73	362.4647	4	31.64	1445.834	-2.6
6737	afNfLhKWKKZ	73	362.4647	4	31.77	1445.834	-2.6
6750	afNfgAKWKKZ	73	362.4648	4	31.82	1445.834	-2.4
6870	afNfgAKWKKZ	73	362.4648	4	32.29	1445.834	-2.4
6883	afNfgAKWKKZ	73	362.4648	4	32.34	1445.834	-2.3
6929	afNfgAKWKKZ	73	362.4647	4	32.53	1445.834	-2.7
6942	afNfLhKWKKZ	73	362.4648	4	32.58	1445.834	-2.4
6955	afNfgAKWKKZ	73	362.4648	4	32.63	1445.834	-2.5
6984	GaSLEfKWKKZ	73	433.2524	3	32.75	1296.734	0.9
7020	GaSLEfKWKKZ	73	433.2522	3	32.9	1296.734	0.5

7223	PfgghPKWKKZ	73	336.9685	4	33.7	1343.848	-2.6
7327	PfgghPKWKKZ	73	336.9686	4	34.11	1343.848	-2.2
7376	GGieDfKWKKZ	73	464.2352	3	34.31	1389.698	-10.5
8188	egTQeEKWKKZ	73	367.7083	4	37.63	1466.807	-2.3
8259	egTQeEKWKKZ	73	489.9419	3	37.95	1466.807	-2.5
8260	egTQeEKWKKZ	73	489.9419	3	37.95	1466.807	-2.5
8270	bShLgQKWKKZ	73	418.9351	3	38	1253.797	-11.1
8339	NGHNAfKWKKZ	73	435.5683	3	38.3	1303.694	-8.3
8349	NGHNAfKWKKZ	73	435.5683	3	38.34	1303.694	-8.3
8373	NGHNAfKWKKZ	73	435.5683	3	38.45	1303.694	-8.2
8400	NHGANfKWKKZ	73	435.5687	3	38.56	1303.694	-7.4
8519	ASSSggKWKKZ	73	392.2486	3	39.08	1173.723	0.4
8576	NEHTeGKWKKZ	73	441.8998	3	39.32	1322.688	-8.1
8604	DhSLffKWKKZ	73	478.9348	3	39.44	1433.786	-2.4
8724	DhSLffKWKKZ	73	478.9347	3	39.96	1433.786	-2.5
8780	FFNeAeKWKKZ	73	366.2002	4	40.21	1460.776	-2.7
9151	FFLEegKWKKZ	73	483.6188	3	41.8	1447.838	-2.3
9160	AfLEegKWKKZ	73	483.6191	3	41.83	1447.838	-1.6
9172	AfLEegKWKKZ	73	483.6188	3	41.89	1447.838	-2.2
9184	AfLEegKWKKZ	73	483.6188	3	41.94	1447.838	-2.3
9196	AfLEegKWKKZ	73	483.619	3	41.99	1447.838	-1.9
9208	AfLEegKWKKZ	73	483.6188	3	42.04	1447.838	-2.3
9232	AfLEegKWKKZ	73	483.619	3	42.15	1447.838	-1.8
9244	AfLEegKWKKZ	73	483.6187	3	42.2	1447.838	-2.5
9256	AfLEegKWKKZ	73	483.6188	3	42.25	1447.838	-2.3
9268	AfLEegKWKKZ	73	483.6192	3	42.3	1447.838	-1.5
9280	AfLEegKWKKZ	73	483.6191	3	42.35	1447.838	-1.6
9292	AfLEegKWKKZ	73	483.6191	3	42.4	1447.838	-1.6
9304	AfLEegKWKKZ	73	483.619	3	42.45	1447.838	-2
9316	AfLEegKWKKZ	73	483.6192	3	42.51	1447.838	-1.5
9331	fALEegKWKKZ	73	483.6191	3	42.57	1447.838	-1.6
9340	AfLEegKWKKZ	73	483.6191	3	42.61	1447.838	-1.7
9352	AfLEegKWKKZ	73	483.619	3	42.66	1447.838	-1.8
9364	AfLEegKWKKZ	73	483.6191	3	42.71	1447.838	-1.7
9376	fALEegKWKKZ	73	483.619	3	42.76	1447.838	-1.8
9388	AfLEegKWKKZ	73	483.6191	3	42.82	1447.838	-1.7
9412	AfLEegKWKKZ	73	483.619	3	42.92	1447.838	-1.8
9436	AfLEegKWKKZ	73	483.6191	3	43.02	1447.838	-1.7
9472	AfLEegKWKKZ	73	483.6192	3	43.18	1447.838	-1.5
9484	AfLEegKWKKZ	73	483.6192	3	43.23	1447.838	-1.4
9502	AfLEegKWKKZ	73	483.6192	3	43.31	1447.838	-1.5
9514	FFLEegKWKKZ	73	483.6193	3	43.36	1447.838	-1.3
9523	AfLEegKWKKZ	73	483.6192	3	43.4	1447.838	-1.5

9535	FFLEegKWKKZ	73	483.6192	3	43.46	1447.838	-1.5
9550	AfLEegKWKKZ	73	483.6192	3	43.52	1447.838	-1.5
9562	AfLEegKWKKZ	73	483.6192	3	43.58	1447.838	-1.4
9610	AfLEegKWKKZ	73	483.6188	3	43.79	1447.838	-2.2
10071	QhPEegKWKKZ	73	451.2632	3	45.83	1350.781	-9.9
10110	NHNfgAKWKKZ	73	458.9279	3	46.01	1373.772	-7.2
10811	TFDfgAKWKKZ	73	458.2685	3	49.08	1371.77	9.7
4684	dhAhgfKWKKZ	72	452.619	3	23.46	1354.828	5.3
5149	LhLgAGKWKKZ	72	385.5963	3	25.34	1153.77	-2.6
10005	HGALfGKWKKZ	72	416.2419	3	45.55	1245.713	-7.7
10038	GAHLfGKWKKZ	72	416.2417	3	45.69	1245.713	-8
3116	SFEDGGKWKKZ	72	394.2078	3	16.77	1179.604	-1.7
3128	SFEDGGKWKKZ	72	394.2075	3	16.82	1179.604	-2.4
3140	SFEDGGKWKKZ	72	394.2076	3	16.87	1179.604	-2.2
3400	hFEEADKWKKZ	72	316.9218	4	18.02	1263.661	-2.5
3433	KFSLEaKWKKZ	72	323.9508	4	18.17	1291.776	-1.6
4157	AgTLGNKWKKZ	72	293.6875	4	21.28	1170.724	-2.2
4193	AgTLGNKWKKZ	72	293.6874	4	21.44	1170.724	-2.5
5055	bHGedaKWKKZ	72	346.203	4	24.96	1380.793	-7.2
5182	LhLgAGKWKKZ	72	385.5963	3	25.47	1153.77	-2.6
5252	gFgNTQKWKKZ	72	444.942	3	25.75	1331.808	-2.9
5259	gFgNTQKWKKZ	72	444.9422	3	25.78	1331.808	-2.3
5269	gFgNTQKWKKZ	72	444.942	3	25.82	1331.808	-2.7
5295	gFgNTQKWKKZ	72	444.9422	3	25.92	1331.808	-2.4
5678	fgGAbFKWKKZ	72	335.7057	4	27.48	1338.797	-2.2
6228	FSfGgRKWKKZ	72	347.2098	4	29.73	1384.813	-2.2
6642	TNFLGfKWKKZ	72	336.6952	4	31.39	1342.755	-2.4
6919	afNfgAKWKKZ	72	362.465	4	32.49	1445.834	-1.9
7010	afNfLhKWKKZ	72	362.4645	4	32.85	1445.834	-3.2
7026	afNfgAKWKKZ	72	362.4648	4	32.92	1445.834	-2.5
7275	PfgghPKWKKZ	72	336.9686	4	33.9	1343.848	-2.4
7288	PfgghPKWKKZ	72	336.9686	4	33.95	1343.848	-2.2
8235	egTQeEKWKKZ	72	489.9422	3	37.84	1466.807	-1.8
8236	egTQeEKWKKZ	72	489.9422	3	37.84	1466.807	-1.8
8305	LPLEegKWKKZ	72	455.619	3	38.15	1363.838	-2
8351	NGHNAfKWKKZ	72	435.5683	3	38.35	1303.694	-8.3
8390	NHGANfKWKKZ	72	435.5684	3	38.52	1303.694	-8
8707	DhSLffKWKKZ	72	478.935	3	39.88	1433.786	-1.9
9424	AfLEegKWKKZ	72	483.6187	3	42.97	1447.838	-2.5
10842	TFDfgAKWKKZ	72	458.2684	3	49.22	1371.77	9.6
4020	bALGgHKWKKZ	71	305.6994	4	20.69	1218.772	-2.5
5081	dSGedaKWKKZ	71	461.2686	3	25.07	1380.782	1.7
5445	gPAGGeKWKKZ	71	398.9089	3	26.53	1193.707	-1.9

2889	hLASNDKWKKZ	71	391.8953	3	15.78	1172.667	-2.2
2909	AhLSQQKWKKZ	71	400.9114	3	15.86	1199.714	-1.3
3884	AAKDLFKWKKZ	71	309.1916	4	20.1	1232.739	-1.7
3896	AAKDLFKWKKZ	71	309.1914	4	20.15	1232.739	-2.1
4029	RKLgLKKWKKZ	71	339.2438	4	20.73	1352.95	-2.7
4070	LQREQeKWKKZ	71	360.7112	4	20.91	1438.82	-2.8
4113	AKQfFNKWKKZ	71	350.7044	4	21.1	1398.793	-2.7
4169	AgTLGNKWKKZ	71	293.6875	4	21.34	1170.724	-2.3
4327	AhAfgGKWKKZ	71	306.4441	4	22	1221.739	6.9
4606	LGARLFKWKKZ	71	415.9349	3	23.13	1244.787	-3.3
4850	TRfQAFKWKKZ	71	472.274	3	24.14	1413.803	-2.1
4873	TRfQAFKWKKZ	71	354.4573	4	24.24	1413.803	-2.1
4935	FLAhaeKWKKZ	71	326.2018	4	24.49	1300.781	-2.1
5140	KTfhKfKWKKZ	71	369.9765	4	25.31	1475.881	-2.5
5153	TKfhKfKWKKZ	71	369.9766	4	25.36	1475.881	-2.2
5159	LhLgAGKWKKZ	71	385.5965	3	25.38	1153.77	-2.1
5166	LhLgAGKWKKZ	71	385.5963	3	25.41	1153.77	-2.7
5282	gFgNTQKWKKZ	71	444.9425	3	25.87	1331.808	-1.6
5308	gFgNTQKWKKZ	71	444.9422	3	25.97	1331.808	-2.4
5630	GGhgfGKWKKZ	71	398.9089	3	27.29	1193.707	-2
5645	ffKHgTKWKKZ	71	382.7293	4	27.35	1526.891	-2.1
5661	ffKHgTKWKKZ	71	382.7293	4	27.41	1526.891	-2.1
5674	ffKHgTKWKKZ	71	382.7292	4	27.47	1526.891	-2.4
5942	fGENDLKWKKZ	71	447.2422	3	28.55	1338.709	-2.7
6180	FfQEhLKWKKZ	71	354.2057	4	29.54	1412.797	-2.4
6584	GfLHfGKWKKZ	71	466.9316	3	31.16	1397.776	-2.3
6585	GfLHfGKWKKZ	71	466.9316	3	31.16	1397.776	-2.3
6596	eAFfdEKWKKZ	71	383.7106	4	31.21	1530.817	-2.6
6597	eAFfdEKWKKZ	71	383.7106	4	31.21	1530.817	-2.6
6629	TNFLGfKWKKZ	71	336.6952	4	31.34	1342.755	-2.4
6655	TNFLGfKWKKZ	71	336.6952	4	31.44	1342.755	-2.5
6682	afNfLhKWKKZ	71	362.4648	4	31.55	1445.834	-2.5
6687	TNFLGfKWKKZ	71	336.6952	4	31.57	1342.755	-2.3
6721	afNfLhKWKKZ	71	362.4648	4	31.7	1445.834	-2.5
6857	afNfgAKWKKZ	71	362.4647	4	32.24	1445.834	-2.6
6997	GaSLEfKWKKZ	71	325.1912	4	32.8	1296.734	1.3
7262	PfgghPKWKKZ	71	336.9685	4	33.85	1343.848	-2.7
8245	egTQeEKWKKZ	71	489.9423	3	37.88	1466.807	-1.6
8302	DhSLffKWKKZ	71	478.9351	3	38.14	1433.786	-1.8
8375	NGHNAfKWKKZ	71	435.5683	3	38.45	1303.694	-8.2
8831	FFNeAeKWKKZ	71	366.2001	4	40.44	1460.776	-2.9
9175	FFLEegKWKKZ	71	362.9661	4	41.9	1447.838	-1.9
9190	FFLEegKWKKZ	71	362.9659	4	41.97	1447.838	-2.2

9214	FFLEegKWKKZ	71	362.966	4	42.07	1447.838	-2.1
9226	FFLEegKWKKZ	71	362.9659	4	42.12	1447.838	-2.2
9235	FFLEegKWKKZ	71	362.966	4	42.16	1447.838	-2
9259	FFLEegKWKKZ	71	362.9659	4	42.26	1447.838	-2.2
9283	FFLEegKWKKZ	71	362.9661	4	42.36	1447.838	-1.9
9307	FFLEegKWKKZ	71	362.9661	4	42.47	1447.838	-1.9
9334	FFLEegKWKKZ	71	362.9661	4	42.58	1447.838	-1.8
9361	FFLEegKWKKZ	71	362.966	4	42.7	1447.838	-2
9420	FFLEegKWKKZ	71	362.9661	4	42.96	1447.838	-1.8
9421	FFLEegKWKKZ	71	362.9661	4	42.96	1447.838	-1.8
9457	FFLEegKWKKZ	71	362.966	4	43.12	1447.838	-2
9658	AfLEegKWKKZ	71	483.6191	3	44	1447.838	-1.6
9757	hfFgSfKWKKZ	71	494.2874	3	44.45	1479.843	-1.9
10053	GAHLfGKWKKZ	71	416.2415	3	45.75	1245.713	-8.6
10820	HPEfgAKWKKZ	71	458.2684	3	49.12	1371.782	1.4
10832	EHPfgAKWKKZ	71	458.2681	3	49.17	1371.782	0.8
10844	TFDfgAKWKKZ	71	458.2684	3	49.23	1371.77	9.6
8732	GGHFgSKWKKZ	70	400.9014	3	40	1199.693	-8.8
10640	NFGEPfKWKKZ	70	452.5809	3	48.31	1354.719	1.8
11482	DeGSfGKWKKZ	70	442.2333	3	52.03	1323.676	1.4
11634	TiDQPfKWKKZ	70	469.2522	3	52.73	1404.73	3.2
11719	NeGEPfKWKKZ	70	469.2524	3	53.12	1404.734	0.9
2907	hLASNDKWKKZ	70	391.8954	3	15.86	1172.667	-1.8
2971	LhASNDKWKKZ	70	294.1734	4	16.14	1172.667	-1.8
3158	SFEDGGKWKKZ	70	394.2078	3	16.95	1179.604	-1.7
3792	TcNfhGKWKKZ	70	328.4417	4	19.71	1309.741	-2.2
3978	RKLgLKKWKKZ	70	339.2437	4	20.51	1352.95	-2.9
3990	RKLgLKKWKKZ	70	339.2438	4	20.56	1352.95	-2.6
4005	RKLgLKKWKKZ	70	339.2439	4	20.62	1352.95	-2.3
4512	LGARLFKWKKZ	70	415.9353	3	22.75	1244.787	-2.2
4574	LGARLFKWKKZ	70	415.9355	3	23	1244.787	-1.9
4906	fTAhfNKWKKZ	70	352.202	4	24.37	1404.771	5.7
4919	fTAhfNKWKKZ	70	352.2019	4	24.42	1404.771	5.5
4951	fTAhfNKWKKZ	70	352.2022	4	24.55	1404.771	6.5
5604	GGhgfGKWKKZ	70	398.909	3	27.18	1193.707	-1.8
5617	GGhgfGKWKKZ	70	398.9091	3	27.24	1193.707	-1.5
6588	GfLHfGKWKKZ	70	466.9314	3	31.17	1397.776	-2.6
7000	afNfLhKWKKZ	70	362.4648	4	32.81	1445.834	-2.4
7389	PfgghPKWKKZ	70	336.9686	4	34.37	1343.848	-2.2
8185	egTQeEKWKKZ	70	489.9419	3	37.62	1466.807	-2.3
8300	TbALgQKWKKZ	70	418.935	3	38.13	1253.797	-11.3
8402	NHGANfKWKKZ	70	435.5687	3	38.57	1303.694	-7.4
8481	DhSLffKWKKZ	70	478.9348	3	38.92	1433.786	-2.3

8640	DhSLefKWKKZ	70	470.2625	3	39.59	1407.77	-3.2
8664	DhSLffKWKKZ	70	478.9348	3	39.69	1433.786	-2.2
8746	DhSLffKWKKZ	70	478.9352	3	40.06	1433.786	-1.6
8789	FFNeAeKWKKZ	70	366.2003	4	40.25	1460.776	-2.4
9202	FFLEegKWKKZ	70	362.9661	4	42.02	1447.838	-1.9
9247	FFLEegKWKKZ	70	362.9659	4	42.21	1447.838	-2.5
9295	FFLEegKWKKZ	70	362.9661	4	42.42	1447.838	-1.9
9325	FFLEegKWKKZ	70	362.9661	4	42.55	1447.838	-1.9
9408	FFLEegKWKKZ	70	362.966	4	42.91	1447.838	-2
9409	FFLEegKWKKZ	70	362.966	4	42.91	1447.838	-2
9672	AfLEegKWKKZ	70	483.619	3	44.07	1447.838	-1.8
9673	AfLEegKWKKZ	70	483.619	3	44.07	1447.838	-1.8
9993	AGHLfGKWKKZ	70	416.2419	3	45.5	1245.713	-7.7
10017	AGHLfGKWKKZ	70	416.2418	3	45.6	1245.713	-7.9
10029	GAHLfGKWKKZ	70	416.242	3	45.65	1245.713	-7.3
11636	TiDQPfKWKKZ	70	469.2522	3	52.74	1404.73	3.2
11644	TiDQPfKWKKZ	70	469.2522	3	52.78	1404.73	3.2
11646	TiDQPfKWKKZ	70	469.2522	3	52.79	1404.73	3.2
11666	eGDQPfKWKKZ	70	469.2525	3	52.88	1404.734	1.1
4671	KAFAgfKWKKZ	69	452.619	3	23.4	1354.828	5.3
4713	GFSLLFKWKKZ	69	418.256	3	23.58	1251.749	-2.6
4870	KAiQfNKWKKZ	69	352.202	4	24.22	1404.778	0.8
5665	GGNFFQKWKKZ	69	413.565	3	27.43	1237.672	1
6075	DbHFRDKWKKZ	69	346.9457	4	29.1	1383.752	0.9
6087	iQPFRDKWKKZ	69	346.9457	4	29.16	1383.752	1
2852	AhLSQQKWKKZ	69	400.9111	3	15.62	1199.714	-2
2865	LAhSNDKWKKZ	69	294.1734	4	15.68	1172.667	-1.8
2902	AhLSQQKWKKZ	69	400.9113	3	15.84	1199.714	-1.5
2903	AhLSQQKWKKZ	69	400.9113	3	15.84	1199.714	-1.5
2919	hLASNDKWKKZ	69	294.1733	4	15.91	1172.667	-2.1
2922	hLASNDKWKKZ	69	391.8954	3	15.92	1172.667	-1.7
3421	hFEEADKWKKZ	69	316.9219	4	18.11	1263.661	-2
3875	AAKDLFKWKKZ	69	309.1915	4	20.06	1232.739	-1.9
3888	fEPAHGKWKKZ	69	326.4325	4	20.12	1301.703	-1.8
4551	LGARLFKWKKZ	69	415.9354	3	22.9	1244.787	-2.1
4580	LGARLFKWKKZ	69	312.2033	4	23.02	1244.787	-2.2
4658	dhAhgfKWKKZ	69	452.6194	3	23.35	1354.828	6.3
4665	dhAhgfKWKKZ	69	339.7161	4	23.38	1354.828	5.5
4914	dAQgfGKWKKZ	69	343.4569	4	24.4	1369.802	-2.6
5028	NFDSfdKWKKZ	69	484.2582	3	24.85	1449.756	-2
5068	dSGedaKWKKZ	69	346.203	4	25.02	1380.782	0.9
5097	GSdedaKWKKZ	69	461.2681	3	25.14	1380.782	0.5
5101	KTfhKfKWKKZ	69	369.9767	4	25.15	1475.881	-2.1

6015	DfdDfKKWKKZ	69	523.6177	3	28.85	1567.834	-1.7
6032	bDHFRDKWKKZ	69	346.9458	4	28.93	1383.752	1.1
6439	SNfDfbKWKKZ	69	492.9297	3	30.58	1475.772	-2.9
7303	PfgghPKWKKZ	69	336.9685	4	34.02	1343.848	-2.6
7431	GiGeDfKWKKZ	69	348.4285	4	34.54	1389.698	-9.6
8512	DhSLffKWKKZ	69	478.9348	3	39.05	1433.786	-2.3
8605	DhSLffKWKKZ	69	478.9348	3	39.44	1433.786	-2.4
9271	FFLEegKWKKZ	69	362.9661	4	42.31	1447.838	-1.7
9769	hfFgSfKWKKZ	69	494.2876	3	44.5	1479.843	-1.6
10630	NFGEPfKWKKZ	69	452.5809	3	48.26	1354.719	1.8
10854	TFDfgAKWKKZ	69	458.268	3	49.27	1371.77	8.8
11668	eGDQPfKWKKZ	69	469.2525	3	52.89	1404.734	1.1
11721	NeGEPfKWKKZ	69	469.2524	3	53.13	1404.734	0.9
6486	AaHFgSKWKKZ	68	419.9211	3	30.76	1256.751	-7.2
7260	eEQPbFKWKKZ	68	471.5992	3	33.84	1411.776	-0.4
2950	hALSQQKWKKZ	68	400.9113	3	16.04	1199.714	-1.4
2951	hALSQQKWKKZ	68	400.9113	3	16.05	1199.714	-1.4
3923	AAKDLFKWKKZ	68	309.191	4	20.27	1232.739	-3.4
3945	RKLgLKKWKKZ	68	339.2439	4	20.36	1352.95	-2.5
3966	RKLgKLKWKKZ	68	339.2439	4	20.45	1352.95	-2.3
4439	fhNRLTKWKKZ	68	345.9613	4	22.45	1379.819	-2.1
4446	fhNRLTKWKKZ	68	345.9611	4	22.48	1379.819	-2.8
4459	fhNRLTKWKKZ	68	345.9612	4	22.53	1379.819	-2.3
4472	hfNRLTKWKKZ	68	345.9612	4	22.58	1379.819	-2.6
4485	hfNRLTKWKKZ	68	345.9611	4	22.64	1379.819	-2.7
4498	hfNRLTKWKKZ	68	345.9611	4	22.69	1379.819	-2.8
4880	fThAfNKWKKZ	68	352.202	4	24.26	1404.771	5.7
4937	dAQgfGKWKKZ	68	343.4571	4	24.49	1369.802	-2.2
5012	NFDSfdKWKKZ	68	484.2582	3	24.79	1449.756	-2
5084	dSGedaKWKKZ	68	346.2031	4	25.08	1380.782	1.1
5710	ffKHgTKWKKZ	68	382.7293	4	27.61	1526.891	-2.2
5992	DfdDfKKWKKZ	68	523.6174	3	28.76	1567.834	-2.2
6036	iQPFRDKWKKZ	68	346.9455	4	28.94	1383.752	0.4
6099	bDHFRDKWKKZ	68	346.9455	4	29.21	1383.752	0.4
6626	TNFLGfKWKKZ	68	448.5913	3	31.33	1342.755	-2.1
6652	TNFLGfKWKKZ	68	448.5912	3	31.43	1342.755	-2.3
6691	TNFLGfKWKKZ	68	448.5912	3	31.58	1342.755	-2.3
6695	afNfgAKWKKZ	68	362.4648	4	31.6	1445.834	-2.5
7039	afNfgAKWKKZ	68	362.4647	4	32.97	1445.834	-2.7
7130	PfgghPKWKKZ	68	336.9684	4	33.34	1343.848	-2.8
7225	PfgghPKWKKZ	68	336.9685	4	33.71	1343.848	-2.6
7358	PfgghPKWKKZ	68	336.9685	4	34.24	1343.848	-2.6
7373	PfgghPKWKKZ	68	336.9685	4	34.3	1343.848	-2.6

8371	DhSLffKWKKZ	68	478.9346	3	38.44	1433.786	-2.7
8451	DhSLffKWKKZ	68	478.9348	3	38.78	1433.786	-2.2
8500	DhSLffKWKKZ	68	478.9349	3	39	1433.786	-2.2
8695	DhSLffKWKKZ	68	478.9349	3	39.83	1433.786	-2.2
8723	HGGFgSKWKKZ	68	400.9014	3	39.96	1199.693	-8.7
8725	DhSLffKWKKZ	68	478.9347	3	39.96	1433.786	-2.5
8741	GGHFgSKWKKZ	68	400.9017	3	40.04	1199.693	-8
8801	FFNeAeKWKKZ	68	366.2003	4	40.31	1460.776	-2.5
10568	NFGEPfKWKKZ	68	452.5806	3	47.98	1354.719	1.1
10642	NFGEPfKWKKZ	68	452.5809	3	48.32	1354.719	1.8
10651	NFGEPfKWKKZ	68	452.5805	3	48.36	1354.719	0.9
10856	TFDfgAKWKKZ	68	458.268	3	49.28	1371.77	8.8
11571	NeGEPfKWKKZ	68	469.2523	3	52.44	1404.734	0.6
11655	NeGEPfKWKKZ	68	469.2526	3	52.83	1404.734	1.2
11708	NeGEPfKWKKZ	68	469.2524	3	53.07	1404.734	0.8
6714	PLNDPgKWKKZ	67	417.9242	3	31.68	1250.75	0.6
6935	GHGhKfKWKKZ	67	425.9171	3	32.55	1274.74	-8.3
7114	eEiAgFKWKKZ	67	471.5994	3	33.27	1411.776	0.1
7396	DPDHDfKWKKZ	67	464.2354	3	34.4	1389.683	1
9719	bbEPTbKWKKZ	67	431.9278	3	44.28	1292.772	-7.8
11791	fENQPfKWKKZ	67	501.603	3	53.44	1501.787	0.1
2877	LhASNDKWKKZ	67	294.1735	4	15.73	1172.667	-1.5
2931	hLASNDKWKKZ	67	294.1733	4	15.96	1172.667	-1.9
4645	dhAhgfKWKKZ	67	452.619	3	23.29	1354.828	5.3
4924	dAQgfGKWKKZ	67	343.4571	4	24.44	1369.802	-2.3
4947	dAQgfGKWKKZ	67	343.4573	4	24.53	1369.802	-1.5
4970	dAQgfGKWKKZ	67	343.4571	4	24.62	1369.802	-2.3
4986	dAQgfGKWKKZ	67	343.4569	4	24.68	1369.802	-2.7
5065	GSdedaKWKKZ	67	461.2685	3	25.01	1380.782	1.4
5262	gFgNTQKWKKZ	67	333.9584	4	25.79	1331.808	-2.5
5272	gFgNTQKWKKZ	67	333.9584	4	25.83	1331.808	-2.4
5301	gFgNTQKWKKZ	67	333.9586	4	25.94	1331.808	-2.1
5463	FFNRhfKWKKZ	67	365.9625	4	26.6	1459.824	-2.2
5999	DfdDfKKWKKZ	67	523.6172	3	28.79	1567.834	-2.6
6028	DfdDfKKWKKZ	67	523.6177	3	28.91	1567.834	-1.7
6049	iQPFRDKWKKZ	67	346.9457	4	29	1383.752	0.8
6062	bHDFRDKWKKZ	67	346.9455	4	29.05	1383.752	0.3
6639	TNFLGfKWKKZ	67	448.5913	3	31.38	1342.755	-2.1
6665	TNFLGfKWKKZ	67	448.5914	3	31.48	1342.755	-1.9
6707	TNFLGfKWKKZ	67	448.5911	3	31.65	1342.755	-2.6
6723	TNFLGfKWKKZ	67	448.5912	3	31.71	1342.755	-2.3
7127	eEiAgFKWKKZ	67	471.5995	3	33.32	1411.776	0.1
7195	eEiAgFKWKKZ	67	471.5995	3	33.59	1411.776	0.3

7202	PfgghPKWKKZ	67	336.9686	4	33.62	1343.848	-2.4
7234	eEiAgFKWKKZ	67	471.5995	3	33.74	1411.776	0.3
7290	PfgghPKWKKZ	67	336.9686	4	33.96	1343.848	-2.2
7329	PfgghPKWKKZ	67	336.9686	4	34.12	1343.848	-2.2
8335	DhSLffKWKKZ	67	478.9348	3	38.28	1433.786	-2.2
8383	DhSLffKWKKZ	67	478.9346	3	38.49	1433.786	-2.7
8614	DhSLefKWKKZ	67	470.2629	3	39.48	1407.77	-2.5
8641	DhSLefKWKKZ	67	470.2625	3	39.59	1407.77	-3.2
8753	GGHFgSKWKKZ	67	400.9015	3	40.09	1199.693	-8.5
9896	GfLfGKWWKKZ	67	483.2721	3	45.06	1446.796	-1.4
10580	NFGEPfKWKKZ	67	452.5807	3	48.03	1354.719	1.3
10618	NFGEPfKWKKZ	67	452.581	3	48.21	1354.719	1.9
10632	NFGEPfKWKKZ	67	452.5809	3	48.27	1354.719	1.8
11526	NeDAPfKWKKZ	67	469.2528	3	52.23	1404.734	1.8
11573	NeGEPfKWKKZ	67	469.2523	3	52.45	1404.734	0.6
11657	NeGEPfKWKKZ	67	469.2526	3	52.84	1404.734	1.2
11689	NeGEPfKWKKZ	67	469.2521	3	52.99	1404.734	0.3
11710	NeGEPfKWKKZ	67	469.2524	3	53.08	1404.734	0.8
3249	hANgHNKWKKZ	66	412.9148	3	17.35	1235.725	-2
3462	ASTafNKWKKZ	66	428.9108	3	18.29	1283.714	-2.6
6744	PLEPTbKWKKZ	66	417.9243	3	31.79	1250.75	0.8
9967	NHAfFQKWKKZ	66	470.2568	3	45.38	1407.756	-5.6
2888	hLASNDKWKKZ	66	391.8953	3	15.77	1172.667	-2.2
2913	hLASNDKWKKZ	66	294.1733	4	15.88	1172.667	-2
2921	hLASNDKWKKZ	66	391.8954	3	15.92	1172.667	-1.7
2938	hLASNDKWKKZ	66	391.8952	3	15.99	1172.667	-2.2
2939	hLASNDKWKKZ	66	391.8952	3	15.99	1172.667	-2.2
4235	NhEKLeKWKKZ	66	339.4546	4	21.62	1353.792	-1.9
4345	AhAfgGKWKKZ	66	408.2564	3	22.08	1221.739	7
4358	AhAfgGKWKKZ	66	408.2563	3	22.13	1221.739	6.9
4530	fhNRLTKWKKZ	66	345.9611	4	22.82	1379.819	-2.8
4613	dhAhgfKWKKZ	66	339.7161	4	23.16	1354.828	5.5
4619	dhAhgfKWKKZ	66	452.6192	3	23.19	1354.828	5.8
4629	dhAhgfKWKKZ	66	452.6194	3	23.23	1354.828	6.2
4717	dhAhgfKWKKZ	66	339.7162	4	23.59	1354.828	5.6
4729	dhAhgfKWKKZ	66	339.7161	4	23.64	1354.828	5.5
4893	fTAhfNKWKKZ	66	352.2019	4	24.32	1404.771	5.5
4932	fTAhfNKWKKZ	66	352.2019	4	24.47	1404.771	5.5
4957	dAQgfGKWKKZ	66	343.4573	4	24.57	1369.802	-1.6
5285	gFgNTQKWKKZ	66	333.9586	4	25.88	1331.808	-1.9
5523	NgPeASKWKKZ	66	427.9197	3	26.85	1280.739	-1.5
5787	fhQhbfKWKKZ	66	365.465	4	27.92	1457.834	-2
5797	fhQhbfKWKKZ	66	365.4647	4	27.96	1457.834	-2.8

5823	fhQhbfKWKKZ	66	365.4651	4	28.06	1457.834	-1.8
6067	DfdDfKKWKKZ	66	523.6173	3	29.07	1567.834	-2.5
6080	DfdDfKKWKKZ	66	523.6174	3	29.12	1567.834	-2.2
6114	DbHFRDKWKKZ	66	346.9453	4	29.27	1383.752	-0.1
6505	hPGAfeKWKKZ	66	330.4411	4	30.84	1317.739	-2.4
6731	PLNDPgKWKKZ	66	417.9243	3	31.74	1250.75	0.8
6782	SEfTeGKWKKZ	66	461.579	3	31.94	1381.718	-2.1
6792	PLEPTbKWKKZ	66	313.6949	4	31.98	1250.75	0.5
6795	SEfTeGKWKKZ	66	461.579	3	31.99	1381.718	-2.2
6821	SEfTeGKWKKZ	66	461.5791	3	32.09	1381.718	-1.9
7101	eEiAgFKWKKZ	66	471.5996	3	33.22	1411.776	0.4
7176	PfgghPKWKKZ	66	336.9686	4	33.51	1343.848	-2.4
7182	eEiAgFKWKKZ	66	471.5995	3	33.54	1411.776	0.2
7208	eEiAgFKWKKZ	66	471.5995	3	33.64	1411.776	0.3
7221	eEiAgFKWKKZ	66	471.5995	3	33.69	1411.776	0.2
7247	eEiAgFKWKKZ	66	471.5996	3	33.79	1411.776	0.5
7264	PfgghPKWKKZ	66	336.9685	4	33.86	1343.848	-2.7
7277	PfgghPKWKKZ	66	336.9686	4	33.91	1343.848	-2.4
7408	GGieDfKWKKZ	66	464.2353	3	34.45	1389.698	-10.3
8161	egTQeEKWKKZ	66	489.942	3	37.52	1466.807	-2.1
8311	DhSLffKWKKZ	66	478.9347	3	38.17	1433.786	-2.5
8323	DTALffKWKKZ	66	478.9348	3	38.23	1433.786	-2.3
8347	DhSLffKWKKZ	66	478.9348	3	38.33	1433.786	-2.3
8359	DhSLffKWKKZ	66	478.9349	3	38.38	1433.786	-2.1
8407	DhSLffKWKKZ	66	478.9348	3	38.59	1433.786	-2.3
8419	DhSLffKWKKZ	66	478.9348	3	38.64	1433.786	-2.2
8429	DhSLffKWKKZ	66	478.9348	3	38.69	1433.786	-2.3
8441	DhSLffKWKKZ	66	478.9348	3	38.74	1433.786	-2.2
8458	DhSLffKWKKZ	66	478.9348	3	38.82	1433.786	-2.3
8470	DhSLffKWKKZ	66	478.9349	3	38.87	1433.786	-2.1
8488	DhSLffKWKKZ	66	478.9348	3	38.95	1433.786	-2.3
8524	DhSLffKWKKZ	66	478.9349	3	39.1	1433.786	-2.1
8532	EgTeQeKWKKZ	66	489.9418	3	39.13	1466.807	-2.5
8533	EgTeQeKWKKZ	66	489.9418	3	39.14	1466.807	-2.5
8536	DhSLffKWKKZ	66	478.9348	3	39.15	1433.786	-2.3
8590	DhSLffKWKKZ	66	478.9347	3	39.38	1433.786	-2.5
8686	DTALffKWKKZ	66	478.9349	3	39.79	1433.786	-2.2
8737	DhSLffKWKKZ	66	478.9347	3	40.02	1433.786	-2.6
8765	GGHFgSKWKKZ	66	400.9014	3	40.15	1199.693	-8.7
9188	SiESfGKKKKZ	66	422.905	3	41.96	1265.688	4.1
9200	ESiSfGKKKKZ	66	422.9051	3	42.01	1265.688	4.2
9671	bbEPTbKWKKZ	66	431.9281	3	44.06	1292.772	-7.2
9969	NHAfFQKWKKZ	66	470.2568	3	45.39	1407.756	-5.6

10570	NFGEPfKWKKZ	66	452.5806	3	47.99	1354.719	1.1
11528	NeDAPfKWKKZ	66	469.2528	3	52.24	1404.734	1.8
3283	QQLSPGKWKKZ	65	400.2392	3	17.51	1197.698	-2.1
6701	SdSGbaKWKKZ	65	407.9123	3	31.62	1220.714	0.8
6753	NPAGGfKWKKZ	65	403.2296	3	31.83	1206.666	0.6
11939	EgTeQeKWKKZ	65	489.9419	3	54.1	1466.807	-2.3
2897	hLASNDKWKKZ	65	391.8955	3	15.81	1172.667	-1.6
2906	hLASNDKWKKZ	65	391.8954	3	15.85	1172.667	-1.8
2914	AhLSQQKWKKZ	65	300.9351	4	15.89	1199.714	-2.2
2915	AhLSQQKWKKZ	65	300.9351	4	15.89	1199.714	-2.2
2946	hALSNDKWKKZ	65	391.8955	3	16.02	1172.667	-1.6
4103	AKQfFNKWKKZ	65	350.7045	4	21.05	1398.793	-2.6
4112	AKQfFNKWKKZ	65	350.7044	4	21.09	1398.793	-2.7
4121	AKQfFNKWKKZ	65	350.7045	4	21.13	1398.793	-2.5
4133	AKQfFNKWKKZ	65	350.7045	4	21.18	1398.793	-2.6
4145	AKQfFNKWKKZ	65	350.7047	4	21.23	1398.793	-2.1
5451	FFNRhfKWKKZ	65	365.9624	4	26.55	1459.824	-2.3
5478	FFNRhfKWKKZ	65	365.9625	4	26.66	1459.824	-2.2
5497	FFNRhfKWKKZ	65	365.9623	4	26.74	1459.824	-2.5
5507	NgPAeSKWKKZ	65	427.9193	3	26.78	1280.739	-2.6
5516	NLhLgQKWKKZ	65	423.6107	3	26.82	1267.813	-2.2
5810	fhQhbfKWKKZ	65	365.4648	4	28.01	1457.834	-2.6
5836	fhQhbfKWKKZ	65	365.4647	4	28.11	1457.834	-2.7
5861	fhQhbfKWKKZ	65	365.4646	4	28.22	1457.834	-3
6041	DfdDfKKWKKZ	65	523.6174	3	28.96	1567.834	-2.4
6054	DfdDfKKWKKZ	65	523.6176	3	29.01	1567.834	-2
6089	DfdDfKKWKKZ	65	523.6179	3	29.16	1567.834	-1.4
6678	TNFLGfKWKKZ	65	448.5912	3	31.53	1342.755	-2.2
7137	PfgghPKWKKZ	65	336.9685	4	33.36	1343.848	-2.6
7150	PfgghPKWKKZ	65	336.9685	4	33.41	1343.848	-2.6
7163	PfgghPKWKKZ	65	336.9686	4	33.46	1343.848	-2.2
7192	PfgghPKWKKZ	65	336.9686	4	33.58	1343.848	-2.4
7251	PfgghPKWKKZ	65	336.9684	4	33.81	1343.848	-2.9
7393	iGGeDfKWKKZ	65	348.4282	4	34.38	1389.698	-10.4
7405	GGieDfKWKKZ	65	348.4281	4	34.43	1389.698	-10.7
7418	GGieDfKWKKZ	65	348.4281	4	34.49	1389.698	-10.8
7421	GGieDfKWKKZ	65	464.2351	3	34.5	1389.698	-10.7
7437	GGieDfKWKKZ	65	464.2356	3	34.57	1389.698	-9.6
8395	DTALffKWKKZ	65	478.9345	3	38.54	1433.786	-2.9
8484	hfFgSgKWKKZ	65	346.9674	4	38.93	1383.843	-2.1
8515	EgTeQeKWKKZ	65	489.9418	3	39.06	1466.807	-2.5
8548	DTALffKWKKZ	65	478.935	3	39.2	1433.786	-1.9
8560	DhSLffKWKKZ	65	478.9347	3	39.25	1433.786	-2.5

8665	DhSLffKWKKZ	65	478.9348	3	39.69	1433.786	-2.2
8792	HGGFgSKWKKZ	65	400.9016	3	40.27	1199.693	-8.2
9263	EiSSfGKKKKZ	65	422.905	3	42.28	1265.688	4
9695	bbEPTbKWKKZ	65	431.9279	3	44.17	1292.772	-7.7
10620	NFGEPfKWKKZ	65	452.581	3	48.22	1354.719	1.9
10653	NFGEPfKWKKZ	65	452.5805	3	48.37	1354.719	0.9
11793	fENQPfKWKKZ	65	501.603	3	53.45	1501.787	0.1
2980	LhSANDKWKKZ	64	294.1733	4	16.17	1172.667	-2
3177	bGSgHNKWKKZ	64	412.9149	3	17.04	1235.725	-1.8
3672	fKAGGPKWKKZ	64	407.912	3	19.19	1220.718	-3.2
4269	ALDGbbKWKKZ	64	399.5806	3	21.76	1195.719	0.7
4349	KAdTTaKWKKZ	64	317.2017	4	22.09	1264.777	0.9
4590	PEdGSAKWKKZ	64	402.5649	3	23.07	1204.672	1.1
7473	eEiAgFKWKKZ	64	471.5994	3	34.72	1411.776	0.1
7596	fAGfFQKWKKZ	64	479.9317	3	35.23	1436.776	-1.6
7650	hEHfFQKWKKZ	64	479.9317	3	35.45	1436.772	1
8612	hASFFQKWKKZ	64	418.5767	3	39.47	1252.708	0.3
9455	GGaiegKWKKZ	64	427.2559	3	43.11	1278.735	8.6
10663	FDGQPfKWKKZ	64	452.581	3	48.42	1354.719	1.9
2945	hALSNDKWKKZ	64	391.8955	3	16.02	1172.667	-1.6
2958	hLASNDKWKKZ	64	294.1734	4	16.08	1172.667	-1.7
3201	GbSgHNKWKKZ	64	412.9148	3	17.14	1235.725	-2
3408	hFEEADKWKKZ	64	316.9222	4	18.06	1263.661	-1.1
4160	AKQfFNKWKKZ	64	350.7046	4	21.3	1398.793	-2.3
4289	AhAfgGKWKKZ	64	306.4441	4	21.85	1221.739	7
4313	AhAfgGKWKKZ	64	306.444	4	21.94	1221.739	6.8
4316	AhAfgGKWKKZ	64	408.256	3	21.96	1221.739	6.2
4329	AhAfgGKWKKZ	64	408.2564	3	22.01	1221.739	7
4352	AhAfgGKWKKZ	64	306.4441	4	22.11	1221.739	7
4365	AhAfgGKWKKZ	64	306.4441	4	22.16	1221.739	7
4378	AhAfgGKWKKZ	64	306.4438	4	22.21	1221.739	6.2
4394	AhAfgGKWKKZ	64	306.4439	4	22.28	1221.739	6.4
4410	AhAfgGKWKKZ	64	306.444	4	22.34	1221.739	6.7
4423	AhAfgGKWKKZ	64	306.4441	4	22.39	1221.739	6.9
4463	KhdSTaKWKKZ	64	317.2018	4	22.55	1264.777	0.8
4476	KAdTTaKWKKZ	64	317.2017	4	22.6	1264.777	0.7
4587	dhAhgfKWKKZ	64	339.7163	4	23.05	1354.828	6.1
4704	dhAhgfKWKKZ	64	339.7162	4	23.54	1354.828	5.7
4741	dhAhgfKWKKZ	64	339.7161	4	23.69	1354.828	5.4
4892	fThAfNKWKKZ	64	352.2019	4	24.31	1404.771	5.5
5565	NLhLgQKWKKZ	64	423.6106	3	27.02	1267.813	-2.3
5649	PSFfEAKWKKZ	64	448.2476	3	27.36	1341.723	-1.8
5656	hGGgfGKWKKZ	64	398.9089	3	27.39	1193.707	-2.1

5873	fhQhbfKWKKZ	64	365.4648	4	28.27	1457.834	-2.6
6166	FSfGgRKWKKZ	64	462.6103	3	29.48	1384.813	-3
6772	SEfTeGKWKKZ	64	461.579	3	31.9	1381.718	-2.1
6808	SEfTeGKWKKZ	64	461.579	3	32.04	1381.718	-2.1
6834	SEfTeGKWKKZ	64	461.5791	3	32.15	1381.718	-2
6847	SEfTeGKWKKZ	64	461.579	3	32.2	1381.718	-2.3
6860	SEfTeGKWKKZ	64	461.579	3	32.25	1381.718	-2.2
6895	SEfTeGKWKKZ	64	461.5793	3	32.39	1381.718	-1.6
7019	GaSLEfKWKKZ	64	433.2522	3	32.89	1296.734	0.5
7035	GaSLEfKWKKZ	64	433.2524	3	32.95	1296.734	1
7045	GaSLEfKWKKZ	64	433.2525	3	32.99	1296.734	1.2
7061	GaSLEfKWKKZ	64	433.2523	3	33.06	1296.734	0.6
7074	GaSLEfKWKKZ	64	433.2522	3	33.11	1296.734	0.5
7090	GaSLEfKWKKZ	64	433.2524	3	33.18	1296.734	1
7238	PfgghPKWKKZ	64	336.9686	4	33.76	1343.848	-2.2
7342	PfgghPKWKKZ	64	336.9686	4	34.17	1343.848	-2.3
7351	eEiAgFKWKKZ	64	471.5996	3	34.21	1411.776	0.4
7573	fAGfFQKWKKZ	64	479.9315	3	35.14	1436.776	-2.2
7637	fAGfFQKWKKZ	64	479.9315	3	35.4	1436.776	-2
8175	DhSLffKWKKZ	64	478.9349	3	37.58	1433.786	-2.1
8176	DhSLffKWKKZ	64	478.9349	3	37.58	1433.786	-2.1
8377	DTALffKWKKZ	64	359.4526	4	38.46	1433.786	-3.1
8410	ESALffKWKKZ	64	359.4525	4	38.6	1433.786	-3.4
8425	DhSLffKWKKZ	64	359.4531	4	38.67	1433.786	-1.8
8435	ESALffKWKKZ	64	359.4531	4	38.71	1433.786	-1.7
8602	DhSLefKWKKZ	64	470.2629	3	39.43	1407.77	-2.4
8772	DTALffKWKKZ	64	478.9347	3	40.18	1433.786	-2.5
8773	DTALffKWKKZ	64	478.9347	3	40.18	1433.786	-2.5
8777	GGHFgSKWKKZ	64	400.9012	3	40.2	1199.693	-9.1
9905	hPQEegKWKKZ	64	451.2631	3	45.1	1350.781	-10.2
10582	NFGEPfKWKKZ	64	452.5807	3	48.04	1354.719	1.3
10866	TFDfgAKWKKZ	64	458.2685	3	49.32	1371.77	9.8
11941	EgTeQeKWKKZ	64	489.9419	3	54.11	1466.807	-2.3
3114	SAGhGgKWKKZ	63	358.2252	3	16.76	1071.655	-1.3
4922	FPAgRDKWKKZ	63	326.2018	4	24.43	1300.777	0.9
5880	AaidGeKWKKZ	63	448.2526	3	28.3	1341.746	-7.1
6417	bLEQHNKWKKZ	63	445.9279	3	30.49	1334.757	3.5
8064	DhHfgAKWKKZ	63	449.5962	3	37.12	1345.766	0.7
9479	GGLAHfKWKKZ	63	416.242	3	43.21	1245.713	-7.4
2849	AhLSQQKWKKZ	63	300.9352	4	15.61	1199.714	-1.9
3660	fKAGGPKWKKZ	63	407.9124	3	19.14	1220.718	-2.3
4296	ALDbGbKWKKZ	63	399.5808	3	21.88	1195.719	1.3
4301	AhAfgGKWKKZ	63	306.444	4	21.9	1221.739	6.8

4304	AhAfgGKWKKZ	63	408.2562	3	21.91	1221.739	6.5
4326	AhAfgGKWKKZ	63	306.4441	4	22	1221.739	6.9
4339	AhAfgGKWKKZ	63	306.444	4	22.05	1221.739	6.8
4374	hAAfgGKWKKZ	63	408.2563	3	22.19	1221.739	6.8
4385	KAdTTaKWKKZ	63	317.2016	4	22.24	1264.777	0.5
4421	KAdTTaKWKKZ	63	317.2018	4	22.38	1264.777	1
4450	KhdSTaKWKKZ	63	317.2017	4	22.5	1264.777	0.7
4869	fThAfNKWKKZ	63	352.202	4	24.22	1404.771	5.9
4879	fThAfNKWKKZ	63	352.202	4	24.26	1404.771	5.7
4905	fThAfNKWKKZ	63	352.202	4	24.36	1404.771	5.7
4908	fTAhfNKWKKZ	63	469.2669	3	24.38	1404.771	5.8
4931	fThAfNKWKKZ	63	352.2019	4	24.47	1404.771	5.5
4940	fThAfNKWKKZ	63	469.2668	3	24.5	1404.771	5.6
5590	NLhLgQKWKKZ	63	423.6105	3	27.13	1267.813	-2.5
5616	AGAgfGKWKKZ	63	398.9091	3	27.23	1193.707	-1.5
6176	FSfGgRKWKKZ	63	462.6104	3	29.52	1384.813	-2.9
6520	AeFfdEKWKKZ	63	383.7109	4	30.89	1530.817	-1.9
6524	AeFfdEKWKKZ	63	383.7107	4	30.91	1530.817	-2.4
6550	AeFfdEKWKKZ	63	383.711	4	31.01	1530.817	-1.5
6563	AeFfdEKWKKZ	63	383.711	4	31.07	1530.817	-1.7
6879	SEfTeGKWKKZ	63	461.579	3	32.32	1381.718	-2.3
6948	GGHhKfKWKKZ	63	425.9173	3	32.6	1274.74	-7.9
6983	GaSLEfKWKKZ	63	433.2524	3	32.74	1296.734	0.9
7215	PfgghPKWKKZ	63	336.9685	4	33.67	1343.848	-2.5
7220	eEbPQFKWKKZ	63	471.5995	3	33.69	1411.776	0.2
7485	iGGeDfKWKKZ	63	348.4283	4	34.77	1389.698	-10.1
8099	DhHfgAKWKKZ	63	449.5964	3	37.26	1345.766	1.1
8108	DhHfgAKWKKZ	63	449.5966	3	37.3	1345.766	1.6
8120	DhHfgAKWKKZ	63	449.5966	3	37.35	1345.766	1.6
8144	DhHfgAKWKKZ	63	449.5963	3	37.45	1345.766	0.9
8180	DhHfgAKWKKZ	63	449.5963	3	37.6	1345.766	0.9
8204	DhHfgAKWKKZ	63	449.5965	3	37.71	1345.766	1.3
8494	EgTeQeKWKKZ	63	489.9419	3	38.97	1466.807	-2.5
8544	EgTeQeKWKKZ	63	489.9418	3	39.19	1466.807	-2.6
8545	EgTeQeKWKKZ	63	489.9418	3	39.19	1466.807	-2.6
8572	DTALffKWKKZ	63	478.9347	3	39.3	1433.786	-2.5
8626	DhSLefKWKKZ	63	470.2631	3	39.53	1407.77	-2.1
8990	QAEKLeKWKKZ	63	452.2705	3	41.12	1353.792	-1.6
9251	EiSSfGKKKKZ	63	422.9048	3	42.23	1265.688	3.7
9407	GGaiegKWKKZ	63	427.2559	3	42.9	1278.735	8.6
11552	NADePfKWKKZ	63	469.2522	3	52.35	1404.734	0.5
2643	LKhNEdKWKKZ	62	334.2076	4	14.69	1332.803	-1.2
3866	AiaETSKWKKZ	62	410.5601	3	20.03	1228.668	-7.6

6502	NbQSgSKWKKZ	62	419.9207	3	30.83	1256.735	3.9
8086	iPSfgAKWKKZ	62	449.5965	3	37.21	1345.766	1.3
8955	TGFPGfKWKKZ	62	424.2406	3	40.98	1269.702	-1.8
11626	NGEePfKWKKZ	62	469.2525	3	52.69	1404.734	1
11733	NDeAPfKWKKZ	62	469.2523	3	53.19	1404.734	0.6
2969	RNESHFKWKKZ	62	340.4408	4	16.12	1357.737	-2
3102	SAGhGgKWKKZ	62	358.2252	3	16.71	1071.655	-1.4
3294	QQLSPGKWKKZ	62	400.2391	3	17.56	1197.698	-2.4
3303	QQSPLGKWKKZ	62	400.2391	3	17.6	1197.698	-2.4
3878	AiaETSKWKKZ	62	410.5601	3	20.08	1228.668	-7.6
4031	GGaLgLKWKKZ	62	385.9281	3	20.74	1154.765	-2.3
4037	GGaLgLKWKKZ	62	385.928	3	20.76	1154.765	-2.6
4049	GGaLgLKWKKZ	62	385.928	3	20.82	1154.765	-2.5
4505	LAGRLFKWKKZ	62	415.9351	3	22.72	1244.787	-2.8
5064	dSGedaKWKKZ	62	461.2685	3	25	1380.782	1.4
5135	KTfhKfKWKKZ	62	492.9661	3	25.29	1475.881	-2.8
5539	NLhLgQKWKKZ	62	423.6108	3	26.91	1267.813	-1.9
5552	NLhLgQKWKKZ	62	423.6108	3	26.96	1267.813	-2
5599	AGAgfGKWKKZ	62	398.9088	3	27.16	1193.707	-2.4
5603	AGAgfGKWKKZ	62	398.909	3	27.18	1193.707	-1.8
5629	AGAgfGKWKKZ	62	398.9089	3	27.28	1193.707	-2
5651	hfFgSbKWKKZ	62	346.7124	4	27.37	1382.823	-1.7
5685	hGGgfGKWKKZ	62	398.9088	3	27.51	1193.707	-2.2
6128	FfQEhLKWKKZ	62	471.9383	3	29.33	1412.797	-2.6
6169	FfQEhLKWKKZ	62	471.938	3	29.49	1412.797	-3.4
6189	FSfGgRKWKKZ	62	462.6107	3	29.57	1384.813	-2.2
6537	AeFfdEKWKKZ	62	383.7108	4	30.96	1530.817	-2.3
6576	AeFfdEKWKKZ	62	383.7107	4	31.12	1530.817	-2.4
6591	SNFLGfKWKKZ	62	443.9192	3	31.18	1328.739	-2.6
7100	eEbPQFKWKKZ	62	471.5996	3	33.22	1411.776	0.4
7113	eEbPQFKWKKZ	62	471.5994	3	33.27	1411.776	0.1
7126	eEbPQFKWKKZ	62	471.5995	3	33.32	1411.776	0.1
7142	eEbPQFKWKKZ	62	471.5995	3	33.38	1411.776	0.3
7181	eEbPQFKWKKZ	62	471.5995	3	33.53	1411.776	0.2
7194	eEbPQFKWKKZ	62	471.5995	3	33.58	1411.776	0.3
7207	eEbPQFKWKKZ	62	471.5995	3	33.63	1411.776	0.3
7233	eEbPQFKWKKZ	62	471.5995	3	33.74	1411.776	0.3
7246	eEbPQFKWKKZ	62	471.5996	3	33.79	1411.776	0.5
7380	iGGeDfKWKKZ	62	348.4283	4	34.33	1389.698	-10
7386	GGieDfKWKKZ	62	464.2354	3	34.36	1389.698	-10
8077	DhHfgAKWKKZ	62	449.5963	3	37.17	1345.766	1
8132	DhHfgAKWKKZ	62	449.5964	3	37.4	1345.766	1.1
8156	DhHfgAKWKKZ	62	449.5964	3	37.5	1345.766	1.1

8192	DhHfgAKWKKZ	62	449.5963	3	37.65	1345.766	1
8318	TPaLgQKWKKZ	62	418.9348	3	38.21	1253.797	-11.6
8779	FFNeAeKWKKZ	62	366.2002	4	40.21	1460.776	-2.7
8902	fegQbFKWKKZ	62	384.9769	4	40.75	1535.881	-1.5
8928	QAEKLeKWKKZ	62	452.2707	3	40.87	1353.792	-1.3
8929	QAEKLeKWKKZ	62	452.2707	3	40.87	1353.792	-1.3
8968	TGFPGfKWKKZ	62	424.24	3	41.03	1269.702	-3.2
11561	NeEGPfKWKKZ	62	469.2525	3	52.39	1404.734	1
11584	NGeEPfKWKKZ	62	469.2523	3	52.5	1404.734	0.6
3900	hfPGhAKWKKZ	61	302.4363	4	20.17	1205.707	7.2
4092	KfGAdLKWKKZ	61	339.9623	4	21	1355.823	-2.1
4208	HGdKSDKWKKZ	61	430.2439	3	21.5	1287.72	-7.8
5383	QfNLHhKWKKZ	61	347.9532	4	26.27	1387.788	-2.9
6403	fQbPgTKWKKZ	61	464.2841	3	30.43	1389.829	1.4
9872	iGGGegKWKKZ	61	412.9087	3	44.95	1235.693	9.4
2870	AhLSQQKWKKZ	61	300.9353	4	15.7	1199.714	-1.4
2871	hALSQQKWKKZ	61	300.9353	4	15.7	1199.714	-1.4
2974	hALSNDKWKKZ	61	391.8954	3	16.15	1172.667	-1.9
2975	hALSNDKWKKZ	61	391.8954	3	16.15	1172.667	-1.9
2986	QcESHFKWKKZ	61	340.4406	4	16.2	1357.737	-2.5
2987	QcESHFKWKKZ	61	340.4406	4	16.2	1357.737	-2.5
3126	SAGhGgKWKKZ	61	358.225	3	16.81	1071.655	-2
3803	KGEgFdKWKKZ	61	338.9598	4	19.75	1351.813	-1.9
3824	KGEgFdKWKKZ	61	338.9598	4	19.84	1351.813	-1.9
3848	KGEgFdKWKKZ	61	338.9597	4	19.95	1351.813	-2.1
3918	hfPGhAKWKKZ	61	302.4362	4	20.25	1205.707	6.9
3957	hfPGhAKWKKZ	61	302.4361	4	20.42	1205.707	6.5
4712	GFSLLFKWKKZ	61	418.256	3	23.57	1251.749	-2.6
5041	dSGedaKWKKZ	61	461.2685	3	24.91	1380.782	1.3
5051	GSdedaKWKKZ	61	461.2683	3	24.95	1380.782	0.9
5074	KTfhKfKWKKZ	61	369.9769	4	25.04	1475.881	-1.5
5080	dSGedaKWKKZ	61	461.2686	3	25.07	1380.782	1.7
5096	dSGedaKWKKZ	61	461.2681	3	25.13	1380.782	0.5
5100	KTfhKfKWKKZ	61	369.9767	4	25.15	1475.881	-2.1
5113	KTfhKfKWKKZ	61	369.9766	4	25.2	1475.881	-2.3
5139	KTfhKfKWKKZ	61	369.9765	4	25.3	1475.881	-2.5
5152	KTfhKfKWKKZ	61	369.9766	4	25.35	1475.881	-2.2
5431	QeLbdgKWKKZ	61	364.7295	4	26.47	1454.891	-1.7
5578	NLhLgQKWKKZ	61	423.6109	3	27.08	1267.813	-1.6
5606	NLhLgQKWKKZ	61	423.6109	3	27.19	1267.813	-1.6
5642	AGAgfGKWKKZ	61	398.9089	3	27.34	1193.707	-2.1
5655	AGAgfGKWKKZ	61	398.9089	3	27.39	1193.707	-2.1
5671	AGAgfGKWKKZ	61	398.9087	3	27.45	1193.707	-2.5

5918	fbShbfKWKKZ	61	365.4652	4	28.45	1457.834	-1.4
6019	DHbFRDKWKKZ	61	346.9458	4	28.87	1383.752	1.2
6119	FfQEhLKWKKZ	61	354.2055	4	29.29	1412.797	-2.7
6138	FfQEhLKWKKZ	61	471.9384	3	29.37	1412.797	-2.5
6151	FfQEhLKWKKZ	61	471.9384	3	29.42	1412.797	-2.6
6760	PLEPTbKWKKZ	61	313.6949	4	31.86	1250.75	0.5
6779	PLEPTbKWKKZ	61	313.695	4	31.93	1250.75	0.6
7107	gTEgfPKWKKZ	61	464.9504	3	33.24	1391.833	-2.6
7259	eEbPQFKWKKZ	61	471.5992	3	33.84	1411.776	-0.4
7324	gTEgPfKWKKZ	61	464.9506	3	34.1	1391.833	-2.1
7521	fAfFGAKWKKZ	61	345.9449	4	34.93	1379.754	-2.6
8168	DhHfgAKWKKZ	61	449.5968	3	37.55	1345.766	2
8938	AQEKLeKWKKZ	61	452.2706	3	40.9	1353.792	-1.4
8964	AQEKLeKWKKZ	61	452.2701	3	41.01	1353.792	-2.6
9118	AQEKLeKWKKZ	61	452.2708	3	41.65	1353.792	-1.1
9488	GGLAHfKWKKZ	61	416.242	3	43.25	1245.713	-7.5
11604	NGEePfKWKKZ	61	469.2522	3	52.59	1404.734	0.5
11679	NGEePfKWKKZ	61	469.2524	3	52.94	1404.734	0.9
3974	bGGggHKWKKZ	60	305.6994	4	20.49	1218.772	-2.5
4678	hAdhgfKWKKZ	60	339.7161	4	23.43	1354.828	5.5
6478	NPPFgSKWKKZ	60	315.1924	4	30.73	1256.739	0.9
6491	HAaFgSKWKKZ	60	315.1925	4	30.78	1256.751	-7.5
7780	EHhfFQKWKKZ	60	479.9315	3	35.97	1436.772	0.6
8893	EgTQeeKWKKZ	60	489.9421	3	40.71	1466.807	-2
10859	gPEPTbKWKKZ	60	422.5961	3	49.29	1264.766	0.7
11517	eDGSfGKKWKZ	60	442.2335	3	52.19	1323.676	1.7
3057	SAGhGgKWKKZ	60	358.225	3	16.51	1071.655	-2
3833	KGEgFdKWKKZ	60	338.9601	4	19.88	1351.813	-1.2
3854	AiaETSKWKKZ	60	410.5599	3	19.97	1228.668	-8
4064	GGaLgLKWKKZ	60	289.6978	4	20.88	1154.765	-2.7
4165	KaTfQLKWKKZ	60	346.2164	4	21.32	1380.839	-2.2
4166	KaTfQLKWKKZ	60	346.2164	4	21.32	1380.839	-2.2
4199	HGdKSDKWKKZ	60	430.2436	3	21.46	1287.72	-8.6
4223	HGdKSDKWKKZ	60	430.2437	3	21.57	1287.72	-8.3
4849	TRfhNFKWKKZ	60	472.274	3	24.14	1413.804	-2.2
5002	NFDSfdKWKKZ	60	363.4454	4	24.75	1449.756	-2.2
5067	bHGedaKWKKZ	60	346.203	4	25.01	1380.793	-7.2
5083	dSGedaKWKKZ	60	346.2031	4	25.08	1380.782	1.1
5087	KTfhKfKWKKZ	60	369.9765	4	25.09	1475.881	-2.4
5106	KTfhKfKWKKZ	60	492.9666	3	25.17	1475.881	-1.8
5122	KTfhKfKWKKZ	60	492.9664	3	25.23	1475.881	-2.2
5126	KTfhKfKWKKZ	60	369.9763	4	25.25	1475.881	-2.9
5168	KTfhKfKWKKZ	60	369.9765	4	25.42	1475.881	-2.6

5526	NLhLgQKWKKZ	60	423.6104	3	26.86	1267.813	-2.8
5625	EgTeAGKWKKZ	60	424.2477	3	27.27	1269.723	-1.6
5638	EgTeAGKWKKZ	60	424.2476	3	27.32	1269.723	-1.9
6996	GaSLEfKWKKZ	60	325.1912	4	32.79	1296.734	1.3
7146	gTEgfPKWKKZ	60	464.9505	3	33.4	1391.833	-2.5
7159	gTEgfPKWKKZ	60	464.9506	3	33.45	1391.833	-2.1
7198	gTEgfPKWKKZ	60	464.9505	3	33.6	1391.833	-2.5
7311	gTEgPfKWKKZ	60	464.9504	3	34.05	1391.833	-2.5
7449	GGieDfKWKKZ	60	464.2353	3	34.62	1389.698	-10.3
7458	DPDHDfKWKKZ	60	348.4285	4	34.66	1389.683	1.4
7472	eEbPQFKWKKZ	60	471.5994	3	34.72	1411.776	0.1
8219	DhHfgAKWKKZ	60	449.5966	3	37.77	1345.766	1.7
8437	EgTeQeKWKKZ	60	489.9417	3	38.72	1466.807	-2.9
8438	EgTeQeKWKKZ	60	489.9417	3	38.73	1466.807	-2.9
8951	AQEKLeKWKKZ	60	452.2706	3	40.96	1353.792	-1.5
9016	AQEKLeKWKKZ	60	452.2701	3	41.22	1353.792	-2.6
9029	AQEKLeKWKKZ	60	452.2699	3	41.28	1353.792	-3
9055	AQEKLeKWKKZ	60	452.2701	3	41.39	1353.792	-2.6
9067	AQEKLeKWKKZ	60	452.2701	3	41.44	1353.792	-2.6
9080	AQEKLeKWKKZ	60	452.2704	3	41.49	1353.792	-2
9144	QAEKLeKWKKZ	60	452.2708	3	41.77	1353.792	-1.1
9145	QAEKLeKWKKZ	60	452.2708	3	41.77	1353.792	-1.1
3210	NhAgHNKWKKZ	59	412.9146	3	17.18	1235.725	-2.6
4081	LgHDHdKWKKZ	59	349.2108	4	20.96	1392.814	0
4332	fdSFGAKWKKZ	59	450.5878	3	22.02	1348.744	-2
4918	TfhAfNKWKKZ	59	352.2019	4	24.42	1404.771	5.5
5133	gGfFSAKWKKZ	59	325.9437	4	25.28	1299.749	-2.8
6492	eaASgSKWKKZ	59	315.1925	4	30.79	1256.739	1.4
2900	AhLSQQKWKKZ	59	300.9352	4	15.83	1199.714	-1.8
3397	PGNgGNKWKKZ	59	289.4247	4	18.01	1153.672	-2.3
3909	hfPGAhKWKKZ	59	302.4361	4	20.21	1205.707	6.7
4010	bGGggHKWKKZ	59	305.6994	4	20.65	1218.772	-2.3
4061	GGaLgLKWKKZ	59	385.9279	3	20.87	1154.765	-2.7
4085	GGaLgLKWKKZ	59	385.9281	3	20.97	1154.765	-2.3
4097	GGaLgLKWKKZ	59	385.928	3	21.02	1154.765	-2.5
4109	GGaLgLKWKKZ	59	385.928	3	21.08	1154.765	-2.6
4139	KaTfQLKWKKZ	59	346.2164	4	21.21	1380.839	-2
4153	KaTfQLKWKKZ	59	346.2165	4	21.27	1380.839	-1.7
4154	KaTfQLKWKKZ	59	346.2165	4	21.27	1380.839	-1.7
4885	TfhAfNKWKKZ	59	469.2668	3	24.28	1404.771	5.5
4898	TfhAfNKWKKZ	59	469.2668	3	24.34	1404.771	5.5
5054	bGHedaKWKKZ	59	346.203	4	24.96	1380.793	-7.2
5090	KTfhKfKWKKZ	59	492.9663	3	25.11	1475.881	-2.3

5275	FFTFThKWKKZ	59	439.5876	3	25.84	1315.744	-2.3
5848	fSbhbfKWKKZ	59	365.4651	4	28.16	1457.834	-1.7
6122	FfQEhLKWKKZ	59	354.2056	4	29.3	1412.797	-2.6
6135	FfQEhLKWKKZ	59	354.2056	4	29.36	1412.797	-2.5
6148	FfQEhLKWKKZ	59	354.2056	4	29.41	1412.797	-2.5
6163	FfQEhLKWKKZ	59	354.2052	4	29.47	1412.797	-3.7
6179	FfQEhLKWKKZ	59	354.2057	4	29.53	1412.797	-2.4
6674	afNfgAKWKKZ	59	482.9508	3	31.52	1445.834	-1.9
6684	afNfLhKWKKZ	59	482.9507	3	31.56	1445.834	-2.2
6697	afNfgAKWKKZ	59	482.9506	3	31.61	1445.834	-2.5
7350	eEbPQFKWKKZ	59	471.5996	3	34.21	1411.776	0.4
7767	fGAfFQKWKKZ	59	479.9316	3	35.92	1436.776	-1.8
8977	AQEKLeKWKKZ	59	452.2706	3	41.06	1353.792	-1.5
9003	AQEKLeKWKKZ	59	452.2705	3	41.17	1353.792	-1.7
9042	AQEKLeKWKKZ	59	452.2702	3	41.33	1353.792	-2.4
9092	AQEKLeKWKKZ	59	452.2703	3	41.54	1353.792	-2.1
9102	QAEKLeKWKKZ	59	452.2704	3	41.58	1353.792	-2
9132	QAEKLeKWKKZ	59	452.2707	3	41.72	1353.792	-1.3
9133	QAEKLeKWKKZ	59	452.2707	3	41.72	1353.792	-1.3
9497	GGLAHfKWKKZ	59	416.2421	3	43.29	1245.713	-7.2
2709	SELGGQKWKKZ	58	387.224	3	14.99	1158.651	-0.7
3305	KRDGPGKWKKZ	58	400.2391	3	17.61	1197.71	-11.8
3684	QPbANPKWKKZ	58	306.1861	4	19.24	1220.714	0.9
4190	fbGGgaKWKKZ	58	320.4507	4	21.43	1277.776	-1.9
4272	hPGPGHKWKKZ	58	373.5579	3	21.78	1117.651	0.9
5723	AhihLFKWKKZ	58	311.4423	4	27.66	1241.74	0.2
5888	fbATbfKWKKZ	58	365.4648	4	28.33	1457.834	-2.5
2861	hALSQQKWKKZ	58	300.9353	4	15.66	1199.714	-1.5
2926	hALSQQKWKKZ	58	300.9352	4	15.94	1199.714	-1.9
2927	hALSQQKWKKZ	58	300.9352	4	15.94	1199.714	-1.9
3812	GKEgFdKWKKZ	58	338.9597	4	19.79	1351.813	-2.2
3863	KGEgFdKWKKZ	58	338.9598	4	20.01	1351.813	-1.9
3872	KGEgFdKWKKZ	58	338.9599	4	20.05	1351.813	-1.7
3937	RKLgKLKWKKZ	58	339.2439	4	20.33	1352.95	-2.5
3938	RKLgKLKWKKZ	58	339.2439	4	20.33	1352.95	-2.5
3942	hfPGhAKWKKZ	58	302.4362	4	20.35	1205.707	7
3965	RKLgKLKWKKZ	58	339.2439	4	20.45	1352.95	-2.3
4019	bGGggHKWKKZ	58	305.6994	4	20.69	1218.772	-2.5
4021	RKLgKLKWKKZ	58	339.2439	4	20.7	1352.95	-2.4
4022	RKLgKLKWKKZ	58	339.2439	4	20.7	1352.95	-2.4
4073	GGaLgLKWKKZ	58	385.9279	3	20.92	1154.765	-2.7
4211	HGdKSDKWKKZ	58	322.9346	4	21.52	1287.72	-8.3
4218	PhGPGHKWKKZ	58	373.558	3	21.55	1117.651	1.2

4319	fdSFGAKWKKZ	58	450.5876	3	21.97	1348.744	-2.6
4950	fTAhfNKWKKZ	58	352.2022	4	24.54	1404.771	6.5
4963	GffGGaKWKKZ	58	435.9126	3	24.59	1304.718	-1.8
5120	gGfFSAKWKKZ	58	325.9438	4	25.23	1299.749	-2.5
5291	FFTFThKWKKZ	58	439.5879	3	25.9	1315.744	-1.7
5298	FFTFThKWKKZ	58	439.5878	3	25.93	1315.744	-1.9
5327	FFTFThKWKKZ	58	439.5877	3	26.05	1315.744	-2.2
5658	fgGAbFKWKKZ	58	447.272	3	27.4	1338.797	-1.9
5903	fQhhbfKWKKZ	58	365.4648	4	28.39	1457.834	-2.5
6235	FSfGgRKWKKZ	58	462.6106	3	29.76	1384.813	-2.4
6739	afNfLhKWKKZ	58	482.9507	3	31.77	1445.834	-2.3
6804	afNfgAKWKKZ	58	482.9505	3	32.03	1445.834	-2.6
7178	GFPLEfKWKKZ	58	452.2594	3	33.52	1353.76	-2.4
7227	gTEgfPKWKKZ	58	464.9505	3	33.71	1391.833	-2.5
7292	gTEgPfKWKKZ	58	464.9506	3	33.97	1391.833	-2.3
7392	iGGeDfKWKKZ	58	348.4282	4	34.38	1389.698	-10.4
7404	GiGeDfKWKKZ	58	348.4281	4	34.43	1389.698	-10.7
8234	DhHfgAKWKKZ	58	449.5969	3	37.84	1345.766	2.1
8788	FFNeAeKWKKZ	58	366.2003	4	40.25	1460.776	-2.4
8830	FFNeAeKWKKZ	58	366.2001	4	40.43	1460.776	-2.9
9509	GGLAHfKWKKZ	58	416.2421	3	43.34	1245.713	-7.1
9539	GGLAHfKWKKZ	58	416.2422	3	43.48	1245.713	-7
4342	RQDSFQKWKKZ	57	450.5878	3	22.06	1348.736	3.9
6449	fhiLgTKWKKZ	57	464.2836	3	30.62	1389.829	0.3
3399	hFEEADKWKKZ	57	316.9218	4	18.02	1263.661	-2.5
3420	hFEEADKWKKZ	57	316.9219	4	18.11	1263.661	-2
3818	PgLShAKWKKZ	57	390.2598	3	19.82	1167.749	7.1
3889	KGEgFdKWKKZ	57	338.9599	4	20.13	1351.813	-1.8
3890	KGEgFdKWKKZ	57	338.9599	4	20.13	1351.813	-1.8
3930	hfPGhAKWKKZ	57	302.4362	4	20.3	1205.707	6.9
3944	RKLgKLKWKKZ	57	339.2439	4	20.36	1352.95	-2.5
3949	TaDLKeKWKKZ	57	336.4546	4	20.38	1341.792	-2.1
3986	bGGggHKWKKZ	57	305.6993	4	20.54	1218.772	-2.7
4028	RKLgKLKWKKZ	57	339.2438	4	20.72	1352.95	-2.7
4175	KaTfQLKWKKZ	57	346.2163	4	21.36	1380.839	-2.3
4257	ALbDGbKWKKZ	57	299.9373	4	21.71	1195.719	0.9
4266	ALbDGbKWKKZ	57	299.9373	4	21.75	1195.719	0.9
4278	ALbDGbKWKKZ	57	299.9374	4	21.8	1195.719	1.3
4639	hAdhgfKWKKZ	57	339.7161	4	23.27	1354.828	5.5
4734	EFGGedKWKKZ	57	451.2475	3	23.66	1350.724	-2.2
4743	EFGGedKWKKZ	57	451.2473	3	23.7	1350.724	-2.7
4755	EFGGedKWKKZ	57	451.2474	3	23.75	1350.724	-2.4
4927	fTAhfNKWKKZ	57	469.2668	3	24.45	1404.771	5.5

4973	GffGGaKWKKZ	57	435.9124	3	24.63	1304.718	-2.3
4989	GffGGaKWKKZ	57	435.9122	3	24.7	1304.718	-2.7
5314	FFTFThKWKKZ	57	439.5876	3	26	1315.744	-2.3
5843	AhihLFKWKKZ	57	311.4423	4	28.14	1241.74	0.2
6459	AHaFgSKWKKZ	57	419.9207	3	30.66	1256.751	-8.2
6465	hPAGfeKWKKZ	57	440.2523	3	30.68	1317.739	-2.7
6472	AHaFgSKWKKZ	57	419.9207	3	30.71	1256.751	-8.2
6485	AHaFgSKWKKZ	57	419.9211	3	30.76	1256.751	-7.2
6498	hPGAfeKWKKZ	57	440.2524	3	30.81	1317.739	-2.4
6501	AHaFgSKWKKZ	57	419.9207	3	30.82	1256.751	-8.2
7155	GFPLEfKWKKZ	57	452.2594	3	33.43	1353.76	-2.6
7230	GFPLEfKWKKZ	57	452.2595	3	33.73	1353.76	-2.4
7253	gTEgfPKWKKZ	57	464.9501	3	33.82	1391.833	-3.2
7285	gTEgfPKWKKZ	57	464.9505	3	33.94	1391.833	-2.4
7305	GFPLEfKWKKZ	57	452.2596	3	34.02	1353.76	-2.2
7363	iGGeDfKWKKZ	57	348.4281	4	34.26	1389.698	-10.8
7457	GiGeDfKWKKZ	57	348.4285	4	34.65	1389.698	-9.7
7484	GiGeDfKWKKZ	57	348.4283	4	34.77	1389.698	-10.1
7725	PAPeeNKWKKZ	57	454.5879	3	35.75	1360.744	-1.8
9431	GaGiegKWKKZ	57	427.2557	3	43.01	1278.735	8.1
9432	FffbFfKWKKZ	57	420.2328	4	43.01	1676.906	-2.4
4274	PLDAGgKWKKZ	56	390.2441	3	21.78	1167.713	-2.1
4921	KSTaaeKWKKZ	56	326.2018	4	24.43	1300.777	1.1
5356	FhGSFeKWKKZ	56	436.9123	3	26.16	1307.718	-2.1
5594	hAPPhGKWKKZ	56	360.8946	3	27.14	1079.66	1.5
10561	DGFQPfKWKKZ	56	452.5807	3	47.95	1354.719	1.3
2886	hgGSNDKWKKZ	56	294.1734	4	15.77	1172.667	-2
3830	gPLShAKWKKZ	56	390.2599	3	19.87	1167.749	7.4
4284	PhGPGHKWKKZ	56	373.5581	3	21.83	1117.651	1.4
4554	GGgRLFKWKKZ	56	312.2033	4	22.92	1244.787	-2.3
4673	SFGLLFKWKKZ	56	313.9438	4	23.41	1251.749	-2.5
4773	EFGGedKWKKZ	56	451.2477	3	23.82	1350.724	-1.8
4999	GffGGaKWKKZ	56	435.9124	3	24.74	1304.718	-2.3
5146	gGfFTGKWKKZ	56	325.9438	4	25.33	1299.749	-2.6
5318	hPAPSEKWKKZ	56	385.5615	3	26.01	1153.661	1.4
5450	FFNRhfKWKKZ	56	365.9624	4	26.55	1459.824	-2.3
5477	FFNRhfKWKKZ	56	365.9625	4	26.66	1459.824	-2.2
5496	FFNRhfKWKKZ	56	365.9623	4	26.74	1459.824	-2.5
5677	fgAGbFKWKKZ	56	335.7057	4	27.48	1338.797	-2.2
5700	fgAGbFKWKKZ	56	335.7059	4	27.57	1338.797	-1.7
6070	bDHFRDKWKKZ	56	462.2582	3	29.08	1383.752	0.3
6083	bDHFRDKWKKZ	56	462.2583	3	29.14	1383.752	0.4
6475	hPGAfeKWKKZ	56	440.2523	3	30.72	1317.739	-2.5

6511	hPAGfeKWKKZ	56	440.2526	3	30.86	1317.739	-2
6632	afNfgAKWKKZ	56	362.4647	4	31.35	1445.834	-2.7
6658	afNfgAKWKKZ	56	362.4648	4	31.45	1445.834	-2.3
6681	afNfgAKWKKZ	56	362.4648	4	31.55	1445.834	-2.5
7165	GFPLEfKWKKZ	56	452.2596	3	33.47	1353.76	-2.2
7204	GFPLEfKWKKZ	56	452.2595	3	33.62	1353.76	-2.2
7217	GFPLEfKWKKZ	56	452.2594	3	33.67	1353.76	-2.4
7243	GFPLEfKWKKZ	56	452.2596	3	33.78	1353.76	-2.2
7282	GFPLEfKWKKZ	56	452.2596	3	33.93	1353.76	-2.2
7369	GiGeDfKWKKZ	56	348.4282	4	34.28	1389.698	-10.6
7385	GGieDfKWKKZ	56	464.2354	3	34.35	1389.698	-10
7430	iGGeDfKWKKZ	56	348.4285	4	34.54	1389.698	-9.6
7442	iGGeDfKWKKZ	56	348.4282	4	34.59	1389.698	-10.5
7754	PAPeeNKWKKZ	56	454.5879	3	35.87	1360.744	-1.9
8651	hASFFQKWKKZ	56	418.5767	3	39.63	1252.708	0.1
8800	FFNeAeKWKKZ	56	366.2003	4	40.3	1460.776	-2.5
8812	FFNeAeKWKKZ	56	366.2002	4	40.35	1460.776	-2.6
9216	EiSSfGKKKKZ	56	422.905	3	42.08	1265.688	4
9240	EiSSfGKKKKZ	56	422.905	3	42.18	1265.688	4.1
9419	GaGiegKWKKZ	56	427.2559	3	42.95	1278.735	8.7
9433	FffbFfKWKKZ	56	420.2328	4	43.01	1676.906	-2.4
9475	FffbFfKWKKZ	56	420.2331	4	43.19	1676.906	-1.6
9490	FffbFfKWKKZ	56	420.2332	4	43.26	1676.906	-1.5
10052	AGHLfGKWKKZ	56	416.2415	3	45.75	1245.713	-8.6
10594	DFGQPfKWKKZ	56	452.5807	3	48.1	1354.719	1.4
2858	GGGGLDKWKKZ	55	348.8692	3	15.64	1043.588	-1.8
4583	PEdTGGKWKKZ	55	402.5651	3	23.04	1204.672	1.5
5681	gfGAbFKWKKZ	55	447.272	3	27.49	1338.797	-1.9
9660	AeFhggKWKKZ	55	448.2839	3	44.01	1341.833	-2
10874	gPEbPTKWKKZ	55	422.5967	3	49.36	1264.766	2.1
2718	SELGGQKWKKZ	55	387.2238	3	15.03	1158.651	-1.2
2727	SELGGQKWKKZ	55	387.2241	3	15.07	1158.651	-0.4
2894	hALSNDKWKKZ	55	294.1734	4	15.8	1172.667	-1.8
3248	hANgHNKWKKZ	55	412.9148	3	17.34	1235.725	-2
3779	PgLSAhKWKKZ	55	390.2599	3	19.65	1167.749	7.3
3794	PgLSAhKWKKZ	55	390.2597	3	19.71	1167.749	6.8
4293	ALbDGbKWKKZ	55	299.9374	4	21.86	1195.719	1.4
4686	GFSLLFKWKKZ	55	313.9438	4	23.46	1251.749	-2.7
4823	TRfhNFKWKKZ	55	472.2743	3	24.03	1413.804	-1.7
4836	TRfhNFKWKKZ	55	472.2737	3	24.08	1413.804	-2.9
5158	LhLgAGKWKKZ	55	385.5965	3	25.38	1153.77	-2.1
5165	LhLgAGKWKKZ	55	385.5963	3	25.41	1153.77	-2.7
5181	LhLgAGKWKKZ	55	385.5963	3	25.47	1153.77	-2.6

5184	gKfGTPKWKKZ	55	441.2751	3	25.48	1320.807	-2.9
5194	gKfGTPKWKKZ	55	441.2754	3	25.52	1320.807	-2.2
5210	gKfGTPKWKKZ	55	441.2751	3	25.58	1320.807	-2.8
5223	gKfGTPKWKKZ	55	441.2751	3	25.64	1320.807	-2.8
5421	ePhLLdKWKKZ	55	343.2176	4	26.42	1368.843	-1.6
5462	FFNRhfKWKKZ	55	365.9625	4	26.6	1459.824	-2.2
5536	hAPPhGKWKKZ	55	360.8944	3	26.9	1079.66	1
5546	hAPPhGKWKKZ	55	360.8947	3	26.94	1079.66	1.9
5559	hAPPhGKWKKZ	55	360.8947	3	26.99	1079.66	1.8
5569	hAPPhGKWKKZ	55	360.8946	3	27.04	1079.66	1.5
5582	hAPPhGKWKKZ	55	360.8948	3	27.09	1079.66	2
5612	PSFfEAKWKKZ	55	448.2477	3	27.21	1341.723	-1.6
6488	hPAGfeKWKKZ	55	440.2528	3	30.77	1317.739	-1.5
6645	afNfgAKWKKZ	55	362.4648	4	31.4	1445.834	-2.4
6704	afNfgAKWKKZ	55	362.4647	4	31.64	1445.834	-2.6
6713	PLEPTbKWKKZ	55	417.9242	3	31.67	1250.75	0.6
6717	PLEPTbKWKKZ	55	417.9243	3	31.69	1250.75	0.8
6720	afNfgAKWKKZ	55	362.4648	4	31.7	1445.834	-2.5
6730	PLEPTbKWKKZ	55	417.9243	3	31.74	1250.75	0.8
6736	afNfgAKWKKZ	55	362.4647	4	31.76	1445.834	-2.6
6749	afNfgAKWKKZ	55	362.4648	4	31.81	1445.834	-2.4
6769	PLEPTbKWKKZ	55	417.9243	3	31.89	1250.75	0.8
6785	PLEPTbKWKKZ	55	417.9242	3	31.95	1250.75	0.7
6801	PLEPTbKWKKZ	55	417.9242	3	32.01	1250.75	0.7
6954	afNfgAKWKKZ	55	362.4648	4	32.63	1445.834	-2.5
7084	gTEgfPKWKKZ	55	348.9647	4	33.15	1391.833	-2.4
7097	gTEgfPKWKKZ	55	348.9647	4	33.2	1391.833	-2.3
7188	GFPLEfKWKKZ	55	452.2595	3	33.56	1353.76	-2.2
7256	GFPLEfKWKKZ	55	452.2591	3	33.83	1353.76	-3.3
7272	GFPLEfKWKKZ	55	452.2593	3	33.89	1353.76	-2.6
7295	GFPLEfKWKKZ	55	452.2595	3	33.98	1353.76	-2.2
7321	GFPLEfKWKKZ	55	452.2596	3	34.09	1353.76	-2.1
7360	GFPLEfKWKKZ	55	452.2594	3	34.25	1353.76	-2.6
7375	GGieDfKWKKZ	55	464.2352	3	34.31	1389.698	-10.5
7379	iGGeDfKWKKZ	55	348.4283	4	34.33	1389.698	-10
8269	bATLgQKWKKZ	55	418.9351	3	37.99	1253.797	-11.1
8275	bATLgQKWKKZ	55	418.9352	3	38.02	1253.797	-10.7
8299	TbALgQKWKKZ	55	418.935	3	38.12	1253.797	-11.3
8632	LNTQeEKWKKZ	55	457.5858	3	39.55	1369.751	-10.8
9187	ESiSfGKKKKZ	55	422.905	3	41.95	1265.688	4.1
9499	FffFbfKWKKZ	55	420.2332	4	43.3	1676.906	-1.4
9516	FffFbfKWKKZ	55	420.2332	4	43.38	1676.906	-1.4
9517	FffFbfKWKKZ	55	420.2332	4	43.38	1676.906	-1.4

9552	FffFbfKWKKZ	55	420.233	4	43.53	1676.906	-1.8
9553	FffFbfKWKKZ	55	420.233	4	43.54	1676.906	-1.8
9661	AeFhggKWKKZ	55	448.2839	3	44.02	1341.833	-2
9983	AGHLfGKWKKZ	55	416.2413	3	45.45	1245.713	-9.1
4093	KTdSHdKWKKZ	54	349.211	4	21.01	1392.814	0.6
4414	bPSbGGKWKKZ	54	380.2334	3	22.36	1137.677	0.9
4691	KKAcTeKWKKZ	54	339.7162	4	23.49	1354.835	0.4
4856	RTfhNFKWKKZ	54	354.4575	4	24.17	1413.804	-1.9
5684	LhAGfGKWKKZ	54	398.9088	3	27.51	1193.707	-2.2
8197	PAFhffKWKKZ	54	478.9348	3	37.67	1433.801	-13.1
8272	DDKLffKWKKZ	54	502.6141	3	38	1504.823	-1.7
9046	LTSHLeKWKKZ	54	446.2669	3	41.35	1335.782	-2
2822	GGGGLDKWKKZ	54	348.8694	3	15.49	1043.588	-1.3
2830	AhLSQQKWKKZ	54	300.9353	4	15.53	1199.714	-1.4
2831	AhLSQQKWKKZ	54	300.9353	4	15.53	1199.714	-1.4
2882	AhLSQQKWKKZ	54	300.9352	4	15.75	1199.714	-1.7
2918	hALSNDKWKKZ	54	294.1733	4	15.9	1172.667	-2.1
3173	DFTDGGKWKKZ	54	394.2078	3	17.02	1179.604	-1.6
3293	QQSPLGKWKKZ	54	400.2391	3	17.55	1197.698	-2.4
4202	bfGGgaKWKKZ	54	320.4505	4	21.48	1277.776	-2.4
4286	PLDAGgKWKKZ	54	390.2436	3	21.83	1167.713	-3.4
4603	PEdTGGKWKKZ	54	402.5648	3	23.12	1204.672	0.7
4616	PEdTGGKWKKZ	54	402.5647	3	23.17	1204.672	0.6
4635	PEdTGGKWKKZ	54	402.5648	3	23.25	1204.672	0.9
4696	FGSLLFKWKKZ	54	313.9438	4	23.5	1251.749	-2.5
4722	GFSLLFKWKKZ	54	313.9438	4	23.61	1251.749	-2.6
4992	begDTbKWKKZ	54	345.9583	4	24.71	1379.808	-2.9
5236	gKfGPTKWKKZ	54	441.2754	3	25.69	1320.807	-2
5622	PSFfEAKWKKZ	54	448.2477	3	27.25	1341.723	-1.5
5668	fgGAbFKWKKZ	54	447.2719	3	27.44	1338.797	-2
5687	fgAGbFKWKKZ	54	335.7057	4	27.52	1338.797	-2
5694	fgGAbFKWKKZ	54	447.272	3	27.55	1338.797	-1.8
5703	LhAGfGKWKKZ	54	398.9089	3	27.58	1193.707	-2.1
5707	fgGAbFKWKKZ	54	447.2719	3	27.6	1338.797	-2
5719	fgGAbFKWKKZ	54	447.2714	3	27.65	1338.797	-3.1
6668	afNfgAKWKKZ	54	362.4649	4	31.5	1445.834	-2.1
6694	afNfgAKWKKZ	54	362.4648	4	31.6	1445.834	-2.5
6743	PLEPTbKWKKZ	54	417.9243	3	31.79	1250.75	0.8
6756	PLEPTbKWKKZ	54	417.9243	3	31.84	1250.75	0.8
6762	afNfgAKWKKZ	54	362.4647	4	31.86	1445.834	-2.6
6775	afNfgAKWKKZ	54	362.4648	4	31.91	1445.834	-2.4
6788	afNfgAKWKKZ	54	362.4648	4	31.96	1445.834	-2.4
6798	afNfgAKWKKZ	54	362.4647	4	32	1445.834	-2.7

6814	afNfgAKWKKZ	54	362.4648	4	32.07	1445.834	-2.5
6840	afNfgAKWKKZ	54	362.4648	4	32.17	1445.834	-2.2
6856	afNfgAKWKKZ	54	362.4647	4	32.23	1445.834	-2.6
6892	afNfgAKWKKZ	54	362.4649	4	32.38	1445.834	-2.1
6999	afNfgAKWKKZ	54	362.4648	4	32.81	1445.834	-2.4
7009	afNfgAKWKKZ	54	362.4645	4	32.85	1445.834	-3.2
7025	afNfgAKWKKZ	54	362.4648	4	32.91	1445.834	-2.5
7051	gTEgfPKWKKZ	54	348.9648	4	33.02	1391.833	-2.2
7077	gTEgfPKWKKZ	54	348.9647	4	33.12	1391.833	-2.5
7110	gTEgfPKWKKZ	54	348.9646	4	33.26	1391.833	-2.8
7123	gTEgfPKWKKZ	54	348.9646	4	33.31	1391.833	-2.8
7139	gTEgfPKWKKZ	54	348.9646	4	33.37	1391.833	-2.6
7152	gTEgfPKWKKZ	54	348.9646	4	33.42	1391.833	-2.8
7168	gTEgfPKWKKZ	54	348.9647	4	33.48	1391.833	-2.4
7331	GFPLEfKWKKZ	54	452.2595	3	34.13	1353.76	-2.2
7401	GFPLEfKWKKZ	54	452.2595	3	34.42	1353.76	-2.4
7417	GGieDfKWKKZ	54	348.4281	4	34.48	1389.698	-10.8
7475	GiGeDfKWKKZ	54	348.4283	4	34.73	1389.698	-10.1
8474	EgTeQeKWKKZ	54	367.7085	4	38.89	1466.807	-1.8
9217	EiSSfGKKKKZ	54	422.905	3	42.08	1265.688	4
9228	EiSSfGKKKKZ	54	422.9047	3	42.13	1265.688	3.4
9229	EiSSfGKKKKZ	54	422.9047	3	42.13	1265.688	3.4
9276	EiSSfGKKKKZ	54	422.9051	3	42.34	1265.688	4.3
9277	EiSSfGKKKKZ	54	422.9051	3	42.34	1265.688	4.3
10028	AGHLfGKWKKZ	54	416.242	3	45.64	1245.713	-7.3
2256	GGQAdDKWKKZ	53	398.2238	3	12.99	1191.651	-1.4
2819	AGhEGFKWKKZ	53	378.8847	3	15.48	1133.635	-1.9
4268	ALTThHKWKKZ	53	399.5806	3	21.76	1195.719	0.7
7347	GaKEEfKWKKZ	53	452.2596	3	34.19	1353.756	0.9
7548	GfPhPNKWKKZ	53	421.2455	3	35.04	1260.713	1.1
2694	SELGGQKWKKZ	53	387.2236	3	14.92	1158.651	-1.6
2853	LhASQQKWKKZ	53	400.9111	3	15.63	1199.714	-2
2891	GGGGLDKWKKZ	53	348.869	3	15.79	1043.588	-2.5
2930	hALSNDKWKKZ	53	294.1733	4	15.95	1172.667	-1.9
3282	QQSPLGKWKKZ	53	400.2392	3	17.5	1197.698	-2.1
3932	TaDLKeKWKKZ	53	336.4546	4	20.31	1341.792	-2.1
4094	KTdSHdKWKKZ	53	349.211	4	21.01	1392.814	0.6
4280	ALTThHKWKKZ	53	399.5807	3	21.81	1195.719	1
4355	fdSFGAKWKKZ	53	450.5878	3	22.12	1348.744	-2
4612	dhAhgfKWKKZ	53	339.7161	4	23.16	1354.828	5.5
4709	GFSLLFKWKKZ	53	313.9438	4	23.56	1251.749	-2.4
4862	TRfhNFKWKKZ	53	472.2744	3	24.19	1413.804	-1.5
5635	PSFfEAKWKKZ	53	448.2477	3	27.31	1341.723	-1.5

5728	aAidGeKWKKZ	53	336.441	4	27.68	1341.746	-8
5870	aAidGeKWKKZ	53	448.2537	3	28.26	1341.746	-4.7
6031	bDHFRDKWKKZ	53	346.9458	4	28.92	1383.752	1.1
6086	bDHFRDKWKKZ	53	346.9457	4	29.15	1383.752	1
6113	DbHFRDKWKKZ	53	346.9453	4	29.26	1383.752	-0.1
6869	afNfgAKWKKZ	53	362.4648	4	32.28	1445.834	-2.4
6882	afNfgAKWKKZ	53	362.4648	4	32.34	1445.834	-2.3
6902	afNfgAKWKKZ	53	362.4647	4	32.42	1445.834	-2.6
6928	afNfgAKWKKZ	53	362.4647	4	32.52	1445.834	-2.7
6941	afNfgAKWKKZ	53	362.4648	4	32.57	1445.834	-2.4
6973	afNfgAKWKKZ	53	362.465	4	32.7	1445.834	-1.8
7038	afNfgAKWKKZ	53	362.4647	4	32.97	1445.834	-2.7
8731	GGHFgSKWKKZ	53	400.9014	3	39.99	1199.693	-8.8
8740	HGGFgSKWKKZ	53	400.9017	3	40.03	1199.693	-8
8752	HGGFgSKWKKZ	53	400.9015	3	40.08	1199.693	-8.5
9199	EiSSfGKKKKZ	53	422.9051	3	42	1265.688	4.2
9241	EiSSfGKKKKZ	53	422.905	3	42.18	1265.688	4.1
9250	ESiSfGKKKKZ	53	422.9048	3	42.22	1265.688	3.7
11845	gETeQeKWKKZ	53	489.9422	3	53.68	1466.807	-1.8
2717	SELNAGKWKKZ	52	387.2238	3	15.02	1158.651	-1.2
4586	KFGhgfKWKKZ	52	339.7163	4	23.05	1354.828	6.1
4600	LGFagfKWKKZ	52	339.7161	4	23.11	1354.828	5.6
6420	iAgfgTKWKKZ	52	464.2836	3	30.5	1389.829	0.3
6468	PbQfgTKWKKZ	52	348.4644	4	30.69	1389.829	0.1
7660	GAfFGAKWKKZ	52	405.5613	3	35.49	1213.676	-11.5
2219	GGAQdDKWKKZ	52	398.2238	3	12.83	1191.651	-1.2
2810	GGGGLDKWKKZ	52	348.8692	3	15.44	1043.588	-1.9
2837	GGGGLDKWKKZ	52	348.8692	3	15.55	1043.588	-1.8
3950	TaDLKeKWKKZ	52	336.4546	4	20.39	1341.792	-2.1
4082	LgHDHdKWKKZ	52	349.2108	4	20.96	1392.814	0
4626	KKAcTeKWKKZ	52	339.7162	4	23.22	1354.835	0.7
4651	dAhhgfKWKKZ	52	339.7163	4	23.32	1354.828	6.2
4652	LGFagfKWKKZ	52	339.7163	4	23.32	1354.828	6.4
4872	RTfQAFKWKKZ	52	354.4573	4	24.23	1413.803	-2.1
5148	LhLgAGKWKKZ	52	385.5963	3	25.34	1153.77	-2.6
5178	KgfGTPKWKKZ	52	331.2082	4	25.46	1320.807	-2.6
5191	KgfGTPKWKKZ	52	331.2083	4	25.51	1320.807	-2.3
5204	KgfGTPKWKKZ	52	331.2082	4	25.56	1320.807	-2.7
5217	KgfGTPKWKKZ	52	331.2082	4	25.61	1320.807	-2.5
6035	DbHFRDKWKKZ	52	346.9455	4	28.94	1383.752	0.4
6048	DbHFRDKWKKZ	52	346.9457	4	28.99	1383.752	0.8
6061	DbHFRDKWKKZ	52	346.9455	4	29.04	1383.752	0.3
6074	DbHFRDKWKKZ	52	346.9457	4	29.1	1383.752	0.9

8491	SASSggKWKKZ	52	392.2487	3	38.96	1173.723	0.6
8503	SASSggKWKKZ	52	392.2488	3	39.01	1173.723	1
8518	SASSggKWKKZ	52	392.2486	3	39.07	1173.723	0.4
8527	SASSggKWKKZ	52	392.2489	3	39.11	1173.723	1.3
8584	NGfTeGKWKKZ	52	446.5716	3	39.35	1336.708	-11.3
8713	GGHFgSKWKKZ	52	400.9016	3	39.91	1199.693	-8.4
8722	GGHFgSKWKKZ	52	400.9014	3	39.95	1199.693	-8.7
8764	HGGFgSKWKKZ	52	400.9014	3	40.14	1199.693	-8.7
8776	HGGFgSKWKKZ	52	400.9012	3	40.19	1199.693	-9.1
9262	EiSSfGKKKKZ	52	422.905	3	42.27	1265.688	4
10004	GAHLfGKWKKZ	52	416.2419	3	45.54	1245.713	-7.7
3921	AKaiGPKWKKZ	51	299.4362	4	20.26	1193.715	1
4371	KhSdTaKWKKZ	51	317.2017	4	22.18	1264.777	0.7
9857	hASfGSKWKKZ	51	400.2299	3	44.89	1197.666	1.8
12665	EgTQeeKWKKZ	51	489.942	3	57.38	1466.807	-2.1
2885	LhASNDKWKKZ	51	294.1734	4	15.76	1172.667	-1.8
2972	hALSNDKWKKZ	51	294.1734	4	16.14	1172.667	-1.8
2981	hALSNDKWKKZ	51	294.1733	4	16.18	1172.667	-2
3304	QQSPLGKWKKZ	51	400.2391	3	17.6	1197.698	-2.4
4488	KhdSTaKWKKZ	51	317.2017	4	22.65	1264.777	0.5
4528	LGARLFKWKKZ	51	312.2032	4	22.81	1244.787	-2.4
4599	dhAhgfKWKKZ	51	339.7161	4	23.1	1354.828	5.5
4703	KFGhgfKWKKZ	51	339.7162	4	23.53	1354.828	5.7
4794	TRfhNFKWKKZ	51	354.4572	4	23.91	1413.804	-2.5
4827	RTfQAFKWKKZ	51	354.4572	4	24.05	1413.803	-2.5
4944	KSTaaeKWKKZ	51	326.202	4	24.52	1300.777	1.7
5161	KgfGTPKWKKZ	51	331.2084	4	25.39	1320.807	-2
5174	KgfGTPKWKKZ	51	331.2082	4	25.44	1320.807	-2.7
5233	gKfGTPKWKKZ	51	331.2084	4	25.68	1320.807	-2
5246	gKfGTPKWKKZ	51	331.2082	4	25.73	1320.807	-2.5
5468	NgPeASKWKKZ	51	427.9195	3	26.62	1280.739	-2.1
5596	PSFfEAKWKKZ	51	336.4374	4	27.15	1341.723	-2.2
5735	aAidGeKWKKZ	51	336.4413	4	27.71	1341.746	-7.2
6358	bGGFgfKWKKZ	51	332.2018	4	30.25	1324.781	-2.3
6700	SdGSbaKWKKZ	51	407.9123	3	31.62	1220.714	0.8
6746	PLEPTbKWKKZ	51	313.6949	4	31.8	1250.75	0.5
6759	PLEPTbKWKKZ	51	313.6949	4	31.85	1250.75	0.5
6778	PLEPTbKWKKZ	51	313.695	4	31.92	1250.75	0.6
6791	PLEPTbKWKKZ	51	313.6949	4	31.98	1250.75	0.5
8530	NEHTeGKWKKZ	51	441.8999	3	39.12	1322.688	-7.8
9992	AGHLfGKWKKZ	51	416.2419	3	45.49	1245.713	-7.7
10016	AGHLfGKWKKZ	51	416.2418	3	45.59	1245.713	-7.9
10037	AGHLfGKWKKZ	51	416.2417	3	45.68	1245.713	-8

5.10 Appendix IV. Peptides identified after cleaving 1,000-member PMO-library from beads

Scan	Peptide	ALC (%)	m/z	z	RT	Mass	ppm
4733	gLESTNKWKKX	99	420.5873	3	23.86	1258.74	0.2
5511	AADGfNKWKKX	99	310.6705	4	25.92	1238.656	-2.6
5517	AADGfNKWKKX	99	413.8917	3	25.93	1238.656	-2.3
6118	GSgFEQKWKKX	99	421.9114	3	27.64	1262.714	-1
6120	GSgFEQKWKKX	99	421.9114	3	27.65	1262.714	-1
6131	PTDGTGKWKKX	99	372.8765	3	27.68	1115.609	-1
6134	GSgFEQKWKKX	99	421.9113	3	27.69	1262.714	-1.1
6136	GSgFEQKWKKX	99	421.9113	3	27.69	1262.714	-1.1
6152	GSgFEQKWKKX	99	421.9116	3	27.74	1262.714	-0.5
6154	GSgFEQKWKKX	99	421.9116	3	27.74	1262.714	-0.5
6168	GSgFEQKWKKX	99	421.9118	3	27.78	1262.714	0
6170	GSgFEQKWKKX	99	421.9118	3	27.79	1262.714	0
6186	GSgFEQKWKKX	99	421.9116	3	27.83	1262.714	-0.5
6202	GSgFEQKWKKX	99	421.9117	3	27.88	1262.714	-0.2
12680	PTgSDeKWKKX	99	438.252	3	49.7	1311.734	0.1
4544	NSSbgLKWKKX	98	414.9274	3	23.37	1241.761	-0.5
4735	gLESTNKWKKX	98	420.5873	3	23.86	1258.74	0.2
5527	AADGfNKWKKX	98	413.8925	3	25.96	1238.656	-0.3
6161	GPEEGeKWKKX	98	418.8925	3	27.76	1253.656	0.2
6188	GSgFEQKWKKX	98	421.9116	3	27.84	1262.714	-0.5
6204	GSgFEQKWKKX	98	421.9117	3	27.89	1262.714	-0.2
6226	GSgFEQKWKKX	98	421.9117	3	27.95	1262.714	-0.2
6751	ANfSgaKWKKX	98	437.5958	3	29.52	1309.766	-0.3
6764	ANfSgaKWKKX	98	437.5955	3	29.57	1309.766	-0.9
6786	ANfSgaKWKKX	98	437.5951	3	29.64	1309.766	-1.9
7436	GDDfNSKWKKX	98	433.8874	3	31.91	1298.641	-0.2
7528	DTgPQgKWKKX	98	428.599	3	32.28	1282.776	-0.8
7532	DTgPQgKWKKX	98	428.5992	3	32.29	1282.776	-0.2
8405	PTDDNgKWKKX	98	419.903	3	35.17	1256.688	-0.6
12669	TPgSDeKWKKX	98	438.2518	3	49.67	1311.734	-0.2
3891	GRETgNKWKKX	97	424.9226	3	21.73	1271.746	-0.3
4278	AgEQEdKWKKX	97	337.9487	4	22.72	1347.766	-0.5
4815	GFPbLGKWKKX	97	395.9133	3	24.06	1184.718	-0.1
4855	ATQgEDKWKKX	97	420.5751	3	24.17	1258.703	0
5065	FgESATKWKKX	97	417.5798	3	24.72	1249.718	-0.5
5519	AADGfNKWKKX	97	413.8917	3	25.94	1238.656	-2.3
5547	AADGfNKWKKX	97	413.8925	3	26.01	1238.656	-0.2

6121	GPEEGeKWKKX	97	418.8921	3	27.65	1253.656	-0.9
6123	GPEEGeKWKKX	97	418.8921	3	27.65	1253.656	-0.9
6163	GPEEGeKWKKX	97	418.8925	3	27.77	1253.656	0.2
6228	GSgFEQKWKKX	97	421.9117	3	27.95	1262.714	-0.2
6477	NGgALDKWKKX	97	395.9081	3	28.69	1184.703	-0.5
6775	ANfSgaKWKKX	97	437.5954	3	29.61	1309.766	-1.2
7432	GDDfNSKWKKX	97	433.8874	3	31.9	1298.641	-0.1
7438	GDDfNSKWKKX	97	433.8874	3	31.92	1298.641	-0.2
7446	GDDfNSKWKKX	97	433.8875	3	31.94	1298.641	0.1
7543	DTgPQgKWKKX	97	428.5991	3	32.33	1282.776	-0.5
7557	DTgPQgKWKKX	97	428.5993	3	32.37	1282.776	0
7579	DTgPQgKWKKX	97	428.5989	3	32.43	1282.776	-0.9
7594	DTgPQgKWKKX	97	428.599	3	32.49	1282.776	-0.7
8389	PTDDNgKWKKX	97	419.9028	3	35.12	1256.688	-0.9
8394	PTDDNgKWKKX	97	419.9032	3	35.13	1256.688	0.1
8398	PTDDNgKWKKX	97	419.9033	3	35.15	1256.688	0.2
8407	PTDDNgKWKKX	97	419.903	3	35.17	1256.688	-0.6
12671	TPgSDeKWKKX	97	438.2518	3	49.68	1311.734	-0.2
3540	ARLREQKWKKX	96	336.211	4	20.87	1340.815	-0.1
3577	ARLREQKWKKX	96	336.2111	4	20.96	1340.815	0
3876	GRETgNKWKKX	96	318.9434	4	21.7	1271.746	-1.4
4049	GGAFNTKWKKX	96	379.2172	3	22.14	1134.63	0
4208	LRLQADKWKKX	96	321.9529	4	22.54	1283.783	0.1
4298	AgEQEdKWKKX	96	337.9487	4	22.77	1347.766	-0.6
4403	DNQhTFKWKKX	96	426.9028	3	23.03	1277.688	-1
4546	NSSbgLKWKKX	96	414.9274	3	23.38	1241.761	-0.5
4742	LgESTNKWKKX	96	315.6919	4	23.88	1258.74	-1.1
4772	GFPbLGKWKKX	96	395.9133	3	23.96	1184.718	-0.2
4774	GFPbLGKWKKX	96	395.9133	3	23.96	1184.718	-0.2
4817	GFPbLGKWKKX	96	395.9133	3	24.07	1184.718	-0.1
4839	GFPbLGKWKKX	96	395.9131	3	24.13	1184.718	-0.6
4846	ATQgEDKWKKX	96	315.683	4	24.15	1258.703	-0.2
5067	FgESATKWKKX	96	417.5798	3	24.72	1249.718	-0.5
5081	FgESATKWKKX	96	417.58	3	24.76	1249.718	-0.1
6133	PTDGTGKWKKX	96	372.8765	3	27.69	1115.609	-1
6149	PTDGTGKWKKX	96	372.8766	3	27.73	1115.609	-0.5
6493	NGgALDKWKKX	96	395.9081	3	28.74	1184.703	-0.5
6753	ANfSgaKWKKX	96	437.5958	3	29.53	1309.766	-0.3
6766	ANfSgaKWKKX	96	437.5955	3	29.58	1309.766	-0.9
6788	ANfSgaKWKKX	96	437.5951	3	29.65	1309.766	-1.9
7434	GDDfNSKWKKX	96	433.8874	3	31.91	1298.641	-0.1
7448	GDDfNSKWKKX	96	433.8875	3	31.95	1298.641	0.1
7530	DTgPQgKWKKX	96	428.599	3	32.29	1282.776	-0.8

8396	PTDDNgKWKKX	96	419.9032	3	35.14	1256.688	0.1
8400	PTDDNgKWKKX	96	419.9033	3	35.15	1256.688	0.2
12682	PTgSDeKWKKX	96	438.252	3	49.71	1311.734	0.1
12701	PTgSDeKWKKX	96	438.2519	3	49.76	1311.734	0
3667	NSdaFGKWKKX	95	423.912	3	21.18	1268.714	0
3799	QAFRNdKWKKX	95	345.9556	4	21.51	1379.794	-0.4
3825	NKTGghKWKKX	95	300.9449	4	21.57	1199.75	0.2
4841	GFPbLGKWKKX	95	395.9131	3	24.14	1184.718	-0.6
5083	FgESATKWKKX	95	417.58	3	24.76	1249.718	-0.1
5524	AADGfNKWKKX	95	310.6711	4	25.95	1238.656	-0.5
5572	AAAgHgKWKKX	95	398.2621	3	26.08	1191.761	3.3
6151	PTDGTGKWKKX	95	372.8766	3	27.74	1115.609	-0.5
6479	NGgALDKWKKX	95	395.9081	3	28.7	1184.703	-0.5
6768	ANfSgaKWKKX	95	437.5955	3	29.58	1309.766	-0.9
6803	ANfSgaKWKKX	95	437.5952	3	29.7	1309.766	-1.7
7534	DTgPQgKWKKX	95	428.5992	3	32.3	1282.776	-0.2
7536	DTgPQgKWKKX	95	428.5992	3	32.31	1282.776	-0.3
7545	DTgPQgKWKKX	95	428.5991	3	32.34	1282.776	-0.5
7555	DTgPQgKWKKX	95	428.5994	3	32.36	1282.776	0.3
7581	DTgPQgKWKKX	95	428.5989	3	32.44	1282.776	-0.9
7583	DTgPQgKWKKX	95	428.5992	3	32.45	1282.776	-0.3
7588	DTgPQgKWKKX	95	428.599	3	32.47	1282.776	-0.7
7603	DTgPQgKWKKX	95	428.599	3	32.53	1282.776	-0.8
7610	DTgPQgKWKKX	95	428.5992	3	32.56	1282.776	-0.4
7824	GAAELfKWKKX	95	418.2449	3	33.32	1251.713	0.3
8215	PTfhAKKWKKX	95	431.9324	3	34.57	1292.776	-0.2
8248	PTfhAKKWKKX	95	431.932	3	34.67	1292.776	-1.1
8391	PTDDNgKWKKX	95	419.9028	3	35.12	1256.688	-0.9
12719	PTgSDeKWKKX	95	438.252	3	49.81	1311.734	0.1
3652	ALaFGTKWKKX	94	393.2448	3	21.15	1176.713	-0.4
3846	ALaKAFKWKKX	94	305.4512	4	21.63	1217.776	-0.3
3864	ALaKAFKWKKX	94	305.4512	4	21.67	1217.776	-0.3
4025	QSKgFaKWKKX	94	327.2094	4	22.09	1304.808	0.2
4280	AgEQEdKWKKX	94	337.9487	4	22.72	1347.766	-0.5
4609	RFKDgGKWKKX	94	330.4581	4	23.53	1317.804	-0.2
5529	AADGfNKWKKX	94	413.8925	3	25.96	1238.656	-0.3
5544	hGDGfNKWKKX	94	310.6711	4	26	1238.656	-0.5
5549	AADGfNKWKKX	94	413.8925	3	26.01	1238.656	-0.2
5582	AAAgHgKWKKX	94	298.9472	4	26.1	1191.761	-0.7
6192	GPEEGeKWKKX	94	418.8923	3	27.85	1253.656	-0.4
6514	NGgALDKWKKX	94	395.908	3	28.8	1184.703	-0.6
7453	GDDfNSKWKKX	94	433.8874	3	31.97	1298.641	-0.3
7455	GDDfNSKWKKX	94	433.8874	3	31.97	1298.641	-0.3

7559	DTgPQgKWKKX	94	428.5993	3	32.38	1282.776	0
7568	DTgPQgKWKKX	94	428.5992	3	32.4	1282.776	-0.4
7614	DTgPQgKWKKX	94	428.5995	3	32.57	1282.776	0.4
8361	ATfNQeKWKKX	94	474.9272	3	35	1421.761	-0.6
3614	hAEbRLKWKKX	93	317.9541	4	21.06	1267.788	-0.4
3816	QAFRNdKWKKX	93	345.9556	4	21.55	1379.794	-0.2
3843	GDRSaFKWKKX	93	313.4333	4	21.62	1249.704	-0.2
3878	GRETgNKWKKX	93	318.9434	4	21.7	1271.746	-1.4
3893	GRETgNKWKKX	93	424.9226	3	21.74	1271.746	-0.3
4051	GGAFNTKWKKX	93	379.2172	3	22.15	1134.63	0
4300	AgEQEdKWKKX	93	337.9487	4	22.77	1347.766	-0.6
4550	ARgFGRKWKKX	93	326.4622	4	23.39	1301.82	0.1
5574	AAAgHgKWKKX	93	398.2621	3	26.08	1191.761	3.3
6194	GPEEGeKWKKX	93	418.8923	3	27.86	1253.656	-0.4
6495	NGgALDKWKKX	93	395.9081	3	28.74	1184.703	-0.5
6777	ANfSgaKWKKX	93	437.5954	3	29.61	1309.766	-1.2
7460	GDDfNSKWKKX	93	433.8875	3	31.99	1298.641	0
7520	DTgPQgKWKKX	93	428.5992	3	32.26	1282.776	-0.2
7522	DTgPQgKWKKX	93	428.5992	3	32.26	1282.776	-0.2
7538	DTgPQgKWKKX	93	428.5992	3	32.31	1282.776	-0.3
7566	DTgPQgKWKKX	93	428.5995	3	32.39	1282.776	0.4
7570	DTgPQgKWKKX	93	428.5992	3	32.41	1282.776	-0.4
7585	DTgPQgKWKKX	93	428.5992	3	32.45	1282.776	-0.3
7596	DTgPQgKWKKX	93	428.599	3	32.5	1282.776	-0.7
7826	GAAELfKWKKX	93	418.2449	3	33.33	1251.713	0.3
12703	PTgSDeKWKKX	93	438.2519	3	49.77	1311.734	0
3579	ARLREQKWKKX	92	336.2111	4	20.97	1340.815	0
3699	RDgSbQKWKKX	92	332.7039	4	21.26	1326.789	-1.5
4210	LRLQADKWKKX	92	321.9529	4	22.55	1283.783	0.1
4364	GGSbgLKWKKX	92	385.917	3	22.93	1154.729	0.4
4405	NDQhTFKWKKX	92	426.9028	3	23.03	1277.688	-1
4425	DNQhTFKWKKX	92	426.9034	3	23.08	1277.688	0.3
4744	LgESTNKWKKX	92	315.6919	4	23.89	1258.74	-1.1
4857	ATQgEDKWKKX	92	420.5751	3	24.18	1258.703	0
4868	ATQgEDKWKKX	92	315.683	4	24.21	1258.703	-0.5
5513	AADGfNKWKKX	92	310.6705	4	25.92	1238.656	-2.6
5566	AADGfNKWKKX	92	413.8925	3	26.06	1238.656	-0.2
5657	ALAATgKWKKX	92	286.4404	4	26.32	1141.734	-0.9
5806	GfGDGAKWKKX	92	390.2136	3	26.72	1167.619	0
6821	ANfSgaKWKKX	92	437.5955	3	29.76	1309.766	-0.9
7462	GDDfNSKWKKX	92	433.8875	3	32	1298.641	0
7590	DTgPQgKWKKX	92	428.599	3	32.47	1282.776	-0.7
7822	GAAELfKWKKX	92	418.2449	3	33.31	1251.713	0.1

7833	GAAELfKWKKX	92	418.2448	3	33.35	1251.713	-0.1
7840	GAAELfKWKKX	92	418.2449	3	33.37	1251.713	0.1
12721	TPgSDeKWKKX	92	438.252	3	49.82	1311.734	0.1
3542	ARLREQKWKKX	91	336.211	4	20.87	1340.815	-0.1
3840	QAFRNdKWKKX	91	345.9555	4	21.61	1379.794	-0.5
3921	GGNPgaKWKKX	91	380.9047	3	21.81	1139.693	-0.2
4230	FTKNTFKWKKX	91	442.9276	3	22.6	1325.761	0.3
4331	NDATGFKWKKX	91	398.5523	3	22.85	1192.635	-0.1
4848	ATQgEDKWKKX	91	315.683	4	24.15	1258.703	-0.2
5520	NGQSeDKWKKX	91	429.5584	3	25.94	1285.657	-2.6
5526	GhDGfNKWKKX	91	310.6711	4	25.96	1238.656	-0.5
5563	hGDGfNKWKKX	91	310.6712	4	26.05	1238.656	-0.1
5671	AhDgdLKWKKX	91	425.9356	3	26.35	1274.786	-1
5688	AhDgdLKWKKX	91	425.9355	3	26.39	1274.786	-1.3
5695	AhDgdLKWKKX	91	425.9357	3	26.41	1274.786	-0.9
5997	NGEeNDKWKKX	91	438.8909	3	27.28	1313.652	-0.6
6938	AGGQfFKWKKX	91	424.5728	3	30.13	1270.698	-0.8
7605	DTgPQgKWKKX	91	428.599	3	32.54	1282.776	-0.8
3616	AhEbRLKWKKX	90	317.9541	4	21.06	1267.788	-0.4
3669	NSdaFGKWKKX	90	423.912	3	21.19	1268.714	0
3845	GDRSaFKWKKX	90	313.4333	4	21.62	1249.704	-0.2
4239	AhLRPPKWKKX	90	302.7001	4	22.62	1206.771	0
5546	hGDGfNKWKKX	90	310.6711	4	26.01	1238.656	-0.5
5568	AADGfNKWKKX	90	413.8925	3	26.07	1238.656	-0.2
5584	AAAgHgKWKKX	90	298.9472	4	26.11	1191.761	-0.7
5594	AAAgHgKWKKX	90	398.2604	3	26.14	1191.761	-0.9
6304	GGLTDgKWKKX	90	386.9041	3	28.18	1157.692	-1.5
6805	ANfSgaKWKKX	90	437.5952	3	29.7	1309.766	-1.7
6899	ANGAfFKWKKX	90	424.5732	3	30.01	1270.698	0.1
6925	ANGAfFKWKKX	90	424.5732	3	30.09	1270.698	0.2
7464	GDDfNSKWKKX	90	433.8873	3	32	1298.641	-0.4
7850	GGhELfKWKKX	90	418.2448	3	33.4	1251.713	-0.4
8217	PTfhAKKWKKX	90	431.9324	3	34.58	1292.776	-0.2
3701	RDgSbQKWKKX	89	332.7039	4	21.27	1326.789	-1.5
3818	QAFRNdKWKKX	89	345.9556	4	21.56	1379.794	-0.2
3923	GGNPgaKWKKX	89	380.9047	3	21.81	1139.693	-0.2
4058	NQEQSFKWKKX	89	441.2383	3	22.17	1320.694	-0.6
4194	GfGAHAKWKKX	89	402.2295	3	22.51	1203.667	0.1
4333	NDATGFKWKKX	89	398.5523	3	22.85	1192.635	-0.1
4427	DNQhTFKWKKX	89	426.9034	3	23.09	1277.688	0.3
5216	GGThQeKWKKX	89	405.2329	3	25.11	1212.677	0.1
5379	aTDGTGKWKKX	89	373.8811	3	25.55	1118.62	1.7
5649	AhDgdLKWKKX	89	425.9356	3	26.29	1274.786	-1.2

6516	NGgALDKWKKX	89	395.908	3	28.8	1184.703	-0.6
6837	ANfSgaKWKKX	89	437.596	3	29.81	1309.766	0.3
6909	AGGQfFKWKKX	89	424.5732	3	30.04	1270.698	0.1
7466	GDDfNSKWKKX	89	433.8873	3	32.01	1298.641	-0.4
8063	NGDDPeKWKKX	89	428.556	3	34.13	1282.646	0.4
8250	PTfhAKKWKKX	89	431.932	3	34.67	1292.776	-1.1
8306	PTFEGAKWKKX	89	397.5607	3	34.84	1189.661	-0.2
3827	KNTGghKWKKX	88	300.9449	4	21.58	1199.75	0.2
3866	ALaKAFKWKKX	88	305.4512	4	21.67	1217.776	-0.3
5240	GGThQeKWKKX	88	405.2328	3	25.17	1212.677	-0.3
5596	AAAgHgKWKKX	88	398.2604	3	26.14	1191.761	-0.9
7612	DTgPQgKWKKX	88	428.5992	3	32.56	1282.776	-0.4
7857	GAAELfKWKKX	88	418.2455	3	33.42	1251.713	1.5
8240	TPfhAKKWKKX	88	431.9321	3	34.64	1292.776	-1
3654	ALdAGTKWKKX	87	393.2448	3	21.15	1176.713	-0.4
3755	LLREGAKWKKX	87	409.9274	3	21.4	1226.761	-0.7
3848	ALaKAFKWKKX	87	305.4512	4	21.63	1217.776	-0.3
3953	PSPKbQKWKKX	87	313.6975	4	21.9	1250.761	-0.1
3987	NKAFDDKWKKX	87	426.9031	3	21.99	1277.688	-0.3
4027	QSKgFaKWKKX	87	327.2094	4	22.09	1304.808	0.2
4181	GGfhKSKWKKX	87	409.2448	3	22.47	1224.713	-0.4
4269	AhLPPRKWKKX	87	302.7	4	22.69	1206.771	-0.3
4524	GQcSSNKWKKX	87	401.8979	3	23.32	1202.663	7.1
4577	SPPEEaKWKKX	87	409.8994	3	23.46	1226.677	-0.4
5234	GGLNAgKWKKX	87	376.5731	3	25.15	1126.698	-0.1
5502	GGQSeDKWKKX	87	410.5518	3	25.89	1228.635	-1.3
5711	AGAgFSKWKKX	87	383.5691	3	26.46	1147.687	-0.9
6823	ANfSgaKWKKX	87	437.5955	3	29.76	1309.766	-0.9
6931	ANGAfFKWKKX	87	318.6814	4	30.11	1270.698	-0.9
7642	NLPNSeKWKKX	87	437.5837	3	32.65	1309.73	-0.2
8038	NGDDPeKWKKX	87	428.5559	3	34.06	1282.646	0.1
8040	NGDDPeKWKKX	87	428.5559	3	34.07	1282.646	0.1
8048	NGDDPeKWKKX	87	428.5559	3	34.09	1282.646	0
8050	NGDDPeKWKKX	87	428.5559	3	34.09	1282.646	0
3842	QAFRNdKWKKX	86	345.9555	4	21.62	1379.794	-0.5
4151	GSgDGGKWKKX	86	363.545	3	22.4	1087.614	-0.4
4232	FTKNTFKWKKX	86	442.9276	3	22.61	1325.761	0.3
4248	AhLPPRKWKKX	86	302.7	4	22.64	1206.771	-0.1
4271	AhLPPRKWKKX	86	302.7	4	22.7	1206.771	-0.3
4504	hPLPPGKWKKX	86	284.434	4	23.28	1133.707	-0.4
4777	hGGGRfKWKKX	86	306.6844	4	23.97	1222.709	-0.3
4870	ATQgEDKWKKX	86	315.683	4	24.21	1258.703	-0.5
4908	QaDbQGKWKKX	86	414.9068	3	24.3	1241.699	-0.4

5922	hAKATSKWKKX	86	377.9059	3	27.06	1130.692	3
6770	ANfSgaKWKKX	86	437.5955	3	29.59	1309.766	-0.9
6776	ANfSgaKWKKX	86	437.5954	3	29.61	1309.766	-1.2
7315	GLEHAaKWKKX	86	399.2378	3	31.46	1194.699	-5.9
8026	NGDDPeKWKKX	86	428.556	3	34.03	1282.646	0.4
8068	NGDDPeKWKKX	86	428.5559	3	34.14	1282.646	0.2
8070	NGDDPeKWKKX	86	428.5559	3	34.15	1282.646	0.2
3905	hSANdFKWKKX	85	317.9369	4	21.77	1267.719	-0.3
4497	DgAGGEKWKKX	85	382.2205	3	23.26	1143.64	-0.4
5194	NGfGAGKWKKX	85	389.8856	3	25.05	1166.635	0.2
5518	AADGfNKWKKX	85	413.8917	3	25.94	1238.656	-2.3
5522	NGQSeDKWKKX	85	429.5584	3	25.95	1285.657	-2.6
5528	AADGfNKWKKX	85	413.8925	3	25.96	1238.656	-0.3
5565	hGDGfNKWKKX	85	310.6712	4	26.06	1238.656	-0.1
6171	TPDGTGKWKKX	85	372.8768	3	27.79	1115.609	-0.1
6769	ANfSgaKWKKX	85	437.5955	3	29.59	1309.766	-0.9
8017	NGDDPeKWKKX	85	428.5561	3	34	1282.646	0.5
8019	NGDDPeKWKKX	85	428.5561	3	34.01	1282.646	0.5
8031	NDGDPeKWKKX	85	428.5561	3	34.04	1282.646	0.5
8033	NDGDPeKWKKX	85	428.5561	3	34.05	1282.646	0.5
8323	PTFEAGKWKKX	85	397.5608	3	34.89	1189.661	0
3981	KANFDDKWKKX	84	320.4291	4	21.97	1277.688	-0.6
3994	QEKDgGKWKKX	84	318.9409	4	22.01	1271.735	-0.4
4197	GGfhKSKWKKX	84	409.2451	3	22.52	1224.713	0.2
4216	GfAGHAKWKKX	84	402.2294	3	22.57	1203.667	0
4503	ALRgGAKWKKX	84	296.7001	4	23.27	1182.771	0
4760	hGGGRfKWKKX	84	306.6844	4	23.93	1222.709	-0.2
5606	AhGgHgKWKKX	84	298.9472	4	26.17	1191.761	-0.6
5729	AGAgFSKWKKX	84	383.5692	3	26.51	1147.687	-0.8
6787	ANfSgaKWKKX	84	437.5951	3	29.64	1309.766	-1.9
6989	RASPFEKWKKX	84	425.9154	3	30.28	1274.725	-0.2
6991	RASPFEKWKKX	84	425.9154	3	30.28	1274.725	-0.2
7727	iDDEhSKWKKX	84	424.8837	3	32.92	1271.626	2.5
3983	KANFDDKWKKX	83	320.4291	4	21.98	1277.688	-0.6
3989	NKAFDDKWKKX	83	426.9031	3	21.99	1277.688	-0.3
4865	hADbEHKWKKX	83	417.9033	3	24.2	1250.689	-0.4
4867	hADbEHKWKKX	83	417.9033	3	24.2	1250.689	-0.4
5381	aTDGTGKWKKX	83	373.8811	3	25.56	1118.62	1.7
5510	GGQSeDKWKKX	83	410.5515	3	25.92	1228.635	-2.2
5548	AADGfNKWKKX	83	413.8925	3	26.01	1238.656	-0.2
5978	DQgGGLKWKKX	83	395.9082	3	27.23	1184.703	-0.1
6609	PQGeEDKWKKX	83	437.8997	3	29.08	1310.677	0.1
6765	ANfSgaKWKKX	83	437.5955	3	29.57	1309.766	-0.9

6804	ANfSgaKWKKX	83	437.5952	3	29.7	1309.766	-1.7
7329	GLEHAaKWKKX	83	399.2377	3	31.5	1194.699	-6
12734	PTgSDeKWKKX	83	438.2514	3	49.85	1311.734	-1.1
4250	AhLPPRKWKKX	82	302.7	4	22.65	1206.771	-0.1
4526	GQcSSNKWKKX	82	401.8979	3	23.33	1202.663	7.1
4626	PSNDEbKWKKX	82	419.5629	3	23.58	1255.667	-0.4
4691	GNFGELKWKKX	82	402.5644	3	23.74	1204.672	-0.3
4856	ATQgEDKWKKX	82	420.5751	3	24.17	1258.703	0
5183	GNfGAGKWKKX	82	389.8856	3	25.02	1166.635	0.2
5512	AADGfNKWKKX	82	310.6705	4	25.92	1238.656	-2.6
5525	AADGfNKWKKX	82	310.6711	4	25.95	1238.656	-0.5
5590	AADGfNKWKKX	82	413.8925	3	26.13	1238.656	-0.2
5793	AhDgdLKWKKX	82	425.936	3	26.69	1274.786	-0.2
6119	GSgFEQKWKKX	82	421.9114	3	27.64	1262.714	-1
6951	AGGQfFKWKKX	82	424.5732	3	30.17	1270.698	0.2
7011	NaaPFEKWKKX	82	425.9156	3	30.34	1274.725	0.2
3892	GRETgNKWKKX	81	424.9226	3	21.74	1271.746	-0.3
3958	NAGNLRKWKKX	81	304.1873	4	21.91	1212.72	0
4241	AhLRPPKWKKX	81	302.7001	4	22.63	1206.771	0
4387	GAQdeQKWKKX	81	337.1936	4	22.99	1344.745	0
4734	gLESTNKWKKX	81	420.5873	3	23.86	1258.74	0.2
4910	QaDbQGKWKKX	81	414.9068	3	24.31	1241.699	-0.4
5714	AGAgFSKWKKX	81	287.9286	4	26.47	1147.687	-1
5760	AhDgdLKWKKX	81	425.9354	3	26.59	1274.786	-1.6
6377	FDSAgbKWKKX	81	421.252	3	28.4	1260.734	-0.2
6812	ANfSgaKWKKX	81	437.5958	3	29.72	1309.766	-0.2
3861	cEGSaFKWKKX	80	313.433	4	21.66	1249.704	-1
3877	GRETgNKWKKX	80	318.9434	4	21.7	1271.746	-1.4
3941	GGNPgaKWKKX	80	285.9302	4	21.87	1139.693	-0.9
4800	hGGGRfKWKKX	80	306.6844	4	24.02	1222.709	-0.2
4958	GGGLPDKWKKX	80	362.2136	3	24.43	1083.619	0.1
5213	GGGfQGKWKKX	80	389.8856	3	25.1	1166.635	0.2
5732	AGAgFSKWKKX	80	287.9287	4	26.52	1147.687	-0.7
6135	GSgFEQKWKKX	80	421.9113	3	27.69	1262.714	-1.1
7653	bTPNSeKWKKX	80	437.5838	3	32.69	1309.73	0
3615	AhEbRLKWKKX	79	317.9541	4	21.06	1267.788	-0.4
4279	LhEQEdKWKKX	79	337.9487	4	22.72	1347.766	-0.5
4987	GGGLPDKWKKX	79	362.2134	3	24.51	1083.619	-0.5
5102	FgEASTKWKKX	79	417.5759	3	24.81	1249.718	-9.9
5237	GGGfQGKWKKX	79	389.8855	3	25.16	1166.635	-0.1
5717	DfSAGAKWKKX	79	404.8885	3	26.48	1211.645	-1
5846	DSFNKFKWKKX	79	442.9151	3	26.84	1325.724	-0.7
6379	FDSAgbKWKKX	79	421.252	3	28.41	1260.734	-0.2

6393	FDSAgbKWKKX	79	421.252	3	28.45	1260.734	-0.1
3807	AGAgARKWKKX	78	381.2484	3	21.53	1140.724	-0.8
4040	hbNRSeKWKKX	78	339.203	4	22.12	1352.783	0
4066	GGNhFQKWKKX	78	392.8929	3	22.19	1175.657	0.3
4299	LhEQEdKWKKX	78	337.9487	4	22.77	1347.766	-0.6
4971	GGGLPDKWKKX	78	362.2136	3	24.47	1083.619	0
5545	AADGfNKWKKX	78	310.6711	4	26	1238.656	-0.5
6081	GNNGfNKWKKX	78	423.2276	3	27.53	1266.662	-0.9
7300	GbDLANKWKKX	78	395.569	3	31.4	1183.683	2.2
8081	NDGDPeKWKKX	78	428.5557	3	34.18	1282.646	-0.5
3975	KPPSbQKWKKX	77	313.6975	4	21.96	1250.761	-0.2
4012	GAFGGTKWKKX	77	360.21	3	22.05	1077.608	-0.2
4026	QSKgFaKWKKX	77	327.2094	4	22.09	1304.808	0.2
4399	DAGhTFKWKKX	77	388.8889	3	23.02	1163.645	-0.4
4572	GGGeDaKWKKX	77	391.217	3	23.44	1170.63	-0.4
4907	QaDbQGKWKKX	77	311.432	4	24.3	1241.699	-0.1
5255	GGLNAgKWKKX	77	376.5731	3	25.21	1126.698	-0.1
5567	AADGfNKWKKX	77	413.8925	3	26.06	1238.656	-0.2
5993	DQgGGLKWKKX	77	395.9079	3	27.27	1184.703	-0.8
5999	GNEeNDKWKKX	77	438.8909	3	27.28	1313.652	-0.6
6097	GNNGfNKWKKX	77	423.2277	3	27.58	1266.662	-0.6
6153	GSgFEQKWKKX	77	421.9116	3	27.74	1262.714	-0.5
6187	GSgFEQKWKKX	77	421.9116	3	27.84	1262.714	-0.5
6320	GGLTDgKWKKX	77	386.9044	3	28.23	1157.692	-0.7
8083	NDGDPeKWKKX	77	428.5557	3	34.18	1282.646	-0.5
3674	ARLPGGKWKKX	76	285.6844	4	21.2	1138.709	-0.3
3863	GcESaFKWKKX	76	313.433	4	21.67	1249.704	-1
4078	GAGFNTKWKKX	76	379.2173	3	22.22	1134.63	0
4084	ESSTGFKWKKX	76	399.5522	3	22.23	1195.635	-0.2
4506	hPLPPGKWKKX	76	284.434	4	23.28	1133.707	-0.4
4551	ARgFGRKWKKX	76	326.4622	4	23.39	1301.82	0.1
4563	AFLDhAKWKKX	76	298.4343	4	23.42	1189.697	9
5564	AADGfNKWKKX	76	310.6712	4	26.06	1238.656	-0.1
6203	GSgFEQKWKKX	76	421.9117	3	27.88	1262.714	-0.2
6371	FDSAgbKWKKX	76	421.2516	3	28.38	1260.734	-1.1
7281	LATeGGKWKKX	76	395.5694	3	31.33	1183.687	-0.1
9780	AaeiDeKWKKX	76	474.5873	3	40.03	1420.74	-0.1
3819	ANPLbQKWKKX	75	310.1935	4	21.56	1236.746	-0.7
3955	hARPbQKWKKX	75	313.6975	4	21.91	1250.773	-9.1
4421	DAGhTFKWKKX	75	388.889	3	23.07	1163.645	-0.1
5439	aTGDTGKWKKX	75	373.8771	3	25.72	1118.62	-9.1
5589	AADGfNKWKKX	75	413.8925	3	26.12	1238.656	-0.2
5751	AGAgFSKWKKX	75	383.5689	3	26.57	1147.687	-1.4

5815	LDHDLFKWKKX	75	443.5868	3	26.75	1327.74	-1.1
5848	DSFNKFKWKKX	75	442.9151	3	26.84	1325.724	-0.7
8107	GNDDPeKWKKX	75	428.5554	3	34.25	1282.646	-1.2
3977	KPPSbQKWKKX	74	313.6975	4	21.96	1250.761	-0.2
4803	GGLGFNKWKKX	74	378.5574	3	24.03	1132.651	-0.2
5670	AhDgdLKWKKX	74	425.9356	3	26.35	1274.786	-1
5687	AhDgdLKWKKX	74	425.9355	3	26.39	1274.786	-1.3
8109	GNDDPeKWKKX	74	428.5554	3	34.26	1282.646	-1.2
8307	KAmDAKKWKKX	74	431.9322	3	34.84	1292.78	-4.3
8309	KAmDAKKWKKX	74	431.9322	3	34.85	1292.78	-4.3
12736	PTgSDeKWKKX	74	438.2514	3	49.86	1311.734	-1.1
3653	LAaFGTKWKKX	73	393.2448	3	21.15	1176.713	-0.4
3904	hASNdFKWKKX	73	317.9369	4	21.77	1267.719	-0.3
4753	gLESTNKWKKX	73	420.5871	3	23.91	1258.74	-0.2
4773	aSNbLGKWKKX	73	395.9133	3	23.96	1184.714	3.3
5441	aTGDTGKWKKX	73	373.8771	3	25.72	1118.62	-9.1
5648	AhDgdLKWKKX	73	425.9356	3	26.29	1274.786	-1.2
5694	AhDgdLKWKKX	73	425.9357	3	26.41	1274.786	-0.9
6169	GSgFEQKWKKX	73	421.9118	3	27.79	1262.714	0
6565	FaDhhfKWKKX	73	448.5919	3	28.95	1342.755	-1
4332	NDATGFKWKKX	72	398.5523	3	22.85	1192.635	-0.1
4743	LgESTNKWKKX	72	315.6919	4	23.88	1258.74	-1.1
4816	aSNbLGKWKKX	72	395.9133	3	24.07	1184.714	3.4
6822	ANfSgaKWKKX	72	437.5955	3	29.76	1309.766	-0.9
6924	AGGQfFKWKKX	72	424.5732	3	30.09	1270.698	0.2
8702	kaGhFGKWKKX	72	416.5742	3	36.25	1246.707	-5.1
3551	ARLREQKWKKX	71	336.211	4	20.9	1340.815	-0.2
3714	GbcSTQKWKKX	71	308.1833	4	21.3	1228.715	-9.2
3920	KDGGdFKWKKX	71	317.9369	4	21.8	1267.719	-0.3
4050	GGAFNTKWKKX	71	379.2172	3	22.15	1134.63	0
4517	DgAGGEKWKKX	71	382.2206	3	23.31	1143.64	0
4714	AhNRNGKWKKX	71	395.9001	3	23.81	1184.689	-9
5066	FgESATKWKKX	71	417.5798	3	24.72	1249.718	-0.5
5274	NLGGAgKWKKX	71	376.5729	3	25.26	1126.698	-0.5
6122	PGEEGeKWKKX	71	418.8921	3	27.65	1253.656	-0.9
6162	PGEEGeKWKKX	71	418.8925	3	27.76	1253.656	0.2
6908	AGGQfFKWKKX	71	424.5732	3	30.04	1270.698	0.1
6937	AGGQfFKWKKX	71	424.5728	3	30.13	1270.698	-0.8
6950	AGGQfFKWKKX	71	424.5732	3	30.17	1270.698	0.2
6963	AGGQfFKWKKX	71	424.573	3	30.2	1270.698	-0.3
8633	hdHGFGKWKKX	71	416.5744	3	36.03	1246.709	-5.8

3541	ARLREQKWKKX	70	336.211	4	20.87	1340.815	-0.1
3578	ARLREQKWKKX	70	336.2111	4	20.96	1340.815	0
3735	hagTGhKWKKX	70	286.6893	4	21.35	1142.729	-0.8
4209	LRLhNDKWKKX	70	321.9529	4	22.55	1283.783	-0.1
4642	PGAGhgKWKKX	70	361.5659	3	23.62	1081.676	-0.1
4840	aSNbLGKWKKX	70	395.9131	3	24.13	1184.714	2.9
4847	ATQgEDKWKKX	70	315.683	4	24.15	1258.703	-0.2
5082	FgESATKWKKX	70	417.58	3	24.76	1249.718	-0.1
5509	GGQSeDKWKKX	70	410.5515	3	25.91	1228.635	-2.2
5583	AAAgHgKWKKX	70	298.9472	4	26.11	1191.761	-0.7
5659	ALAATgKWKKX	70	286.4404	4	26.32	1141.734	-0.9
5735	DfSGGhKWKKX	70	404.8886	3	26.53	1211.645	-0.9
6984	hcSPFEKWKKX	70	425.9154	3	30.26	1274.725	-0.4
7024	hcSPFEKWKKX	70	425.9155	3	30.38	1274.725	-0.1
7821	AGAELfKWKKX	70	418.2449	3	33.31	1251.713	0.1
7832	AGAELfKWKKX	70	418.2448	3	33.34	1251.713	-0.1
7839	AGAELfKWKKX	70	418.2449	3	33.37	1251.713	0.1
8372	ATfNQeKWKKX	70	474.9277	3	35.03	1421.761	0.4
9831	KiiSDeKWKKX	70	474.588	3	40.19	1420.736	4.1
3668	NSdaFGKWKKX	69	423.912	3	21.19	1268.714	0
3961	LNhPbQKWKKX	69	417.9276	3	21.92	1250.761	-0.2
4359	dTPPADKWKKX	69	312.1815	4	22.92	1244.703	-4.9
5317	GSmAGdKWKKX	69	410.2377	3	25.38	1227.696	-4
5335	GSmAGdKWKKX	69	410.2378	3	25.43	1227.696	-3.6
6752	ANfSgaKWKKX	69	437.5958	3	29.52	1309.766	-0.3
6898	GGAQfFKWKKX	69	424.5732	3	30.01	1270.698	0.1
8362	ATfNQeKWKKX	69	474.9272	3	35.01	1421.761	-0.6
3800	QAFRNdKWKKX	68	345.9556	4	21.51	1379.794	-0.4
3815	AGAgARKWKKX	68	286.1882	4	21.55	1140.724	-0.5
3844	RDGSaFKWKKX	68	313.4333	4	21.62	1249.704	-0.2
3982	hQaFDDKWKKX	68	320.4291	4	21.97	1277.688	-0.6
3988	KNAFDDKWKKX	68	426.9031	3	21.99	1277.688	-0.3
4308	GGFNKFKWKKX	68	413.5766	3	22.79	1237.708	-0.2
4543	LEAGGEKWKKX	68	382.2208	3	23.37	1143.64	0.6
5501	NAGSeDKWKKX	68	410.5518	3	25.89	1228.635	-1.3
5759	DAhgdLKWKKX	68	425.9354	3	26.59	1274.786	-1.6
5780	DAhgdLKWKKX	68	425.9359	3	26.65	1274.786	-0.3
6608	ANPeEDKWKKX	68	437.8997	3	29.08	1310.677	0.1
7351	GLEAHaKWKKX	68	399.2375	3	31.57	1194.699	-6.6
7667	bTPNSeKWKKX	68	437.5835	3	32.73	1309.73	-0.6
7849	AGAELfKWKKX	68	418.2448	3	33.4	1251.713	-0.2
7856	AGAELfKWKKX	68	418.2455	3	33.42	1251.713	1.5
9719	RPDiDeKWKKX	68	474.5871	3	39.85	1420.736	2.3

3659	hATdFHKWKKX	67	327.1949	4	21.17	1304.751	-0.1
3778	hAANFbKWKKX	67	401.5765	3	21.46	1201.709	-0.7
3919	hSANdFKWKKX	67	317.9369	4	21.8	1267.719	-0.5
4404	DNQhTFKWKKX	67	426.9028	3	23.03	1277.688	-1
4412	GGPAELKWKKX	67	371.5574	3	23.05	1111.65	0.3
4645	GGSLheKWKKX	67	395.5694	3	23.63	1183.687	-0.1
4869	ATQgEDKWKKX	67	315.683	4	24.21	1258.703	-0.5
4893	oSgAGGKWKKX	67	414.9071	3	24.27	1241.701	-1.2
5998	ANDeNDKWKKX	67	438.8909	3	27.28	1313.652	-0.6
6227	GSgFEQKWKKX	67	421.9117	3	27.95	1262.714	-0.2
6340	ADdGSTKWKKX	67	399.2256	3	28.29	1194.651	3.3
6360	ADdGSTKWKKX	67	399.2253	3	28.35	1194.651	2.5
6368	ADdGSTKWKKX	67	399.2254	3	28.38	1194.651	3
6376	ADdGSTKWKKX	67	399.2256	3	28.4	1194.651	3.3
6478	NGgALDKWKKX	67	395.9081	3	28.69	1184.703	-0.5
7322	GLEAHaKWKKX	67	399.2377	3	31.48	1194.699	-6
7333	GLEAHaKWKKX	67	399.2378	3	31.51	1194.699	-5.7
7343	GLEAHaKWKKX	67	399.2377	3	31.55	1194.699	-6.1
7347	GLEAHaKWKKX	67	399.2378	3	31.56	1194.699	-5.9
7355	GLEAHaKWKKX	67	399.2378	3	31.58	1194.699	-5.9
7359	GLEAHaKWKKX	67	399.2375	3	31.59	1194.699	-6.5
7364	GLEAHaKWKKX	67	399.238	3	31.61	1194.699	-5.3
7368	GLEAHaKWKKX	67	399.2379	3	31.63	1194.699	-5.6
7373	GLEAHaKWKKX	67	399.2377	3	31.65	1194.699	-6
7825	AGAELfKWKKX	67	418.2449	3	33.32	1251.713	0.3
8375	ATfNQeKWKKX	67	474.9276	3	35.04	1421.761	0.2
8376	ATfNQeKWKKX	67	474.9276	3	35.05	1421.761	0.2
8598	ALaAiLKWKKX	67	403.9215	3	35.88	1208.751	-6.4
3741	KAAATiKWKKX	66	296.6818	4	21.37	1182.699	-0.2
3817	QAFRNdKWKKX	66	345.9556	4	21.56	1379.794	-0.2
3902	KAGiTQKWKKX	66	307.4332	4	21.76	1225.704	-0.5
4802	GGLGFNKWKKX	66	378.5574	3	24.03	1132.651	-0.2
5280	GSdGSdKWKKX	66	410.2377	3	25.27	1227.688	3
5640	hAcSDFKWKKX	66	412.5729	3	26.27	1234.693	2.9
5710	AGAgFSKWKKX	66	383.5691	3	26.46	1147.687	-0.9
6303	GGLTDgKWKKX	66	386.9041	3	28.18	1157.692	-1.5
6494	NGgALDKWKKX	66	395.9081	3	28.74	1184.703	-0.5
4231	TFKNTFKWKKX	65	442.9276	3	22.6	1325.761	0.3
4361	dTPPADKWKKX	65	312.1815	4	22.92	1244.703	-4.9
4363	GGSbgLKWKKX	65	385.917	3	22.93	1154.729	0.4
4449	hAGSaTKWKKX	65	363.8904	3	23.14	1088.646	3.7
4545	NSSbgLKWKKX	65	414.9274	3	23.38	1241.761	-0.5
5595	AGhgHgKWKKX	65	398.2604	3	26.14	1191.761	-0.9

5792	hGEgdLKWKKX	65	425.936	3	26.68	1274.786	-0.2
5812	DAaPiEKWKKX	65	418.5645	3	26.74	1252.668	3.2
6193	PEGEGeKWKKX	65	418.8923	3	27.85	1253.656	-0.4
6515	NGgALDKWKKX	65	395.908	3	28.8	1184.703	-0.6
3754	LLhAEQKWKKX	64	409.9274	3	21.4	1226.75	8.5
3836	NhgAhAKWKKX	64	286.1882	4	21.6	1140.713	9.1
4389	GGSbgLKWKKX	64	385.9168	3	22.99	1154.729	-0.2
4571	GGGDeaKWKKX	64	391.217	3	23.44	1170.63	-0.4
4895	ATQgEDKWKKX	64	315.683	4	24.27	1258.703	-0.2
5193	GGGfQGKWKKX	64	389.8856	3	25.05	1166.635	0.2
5521	GNQSeDKWKKX	64	429.5584	3	25.94	1285.657	-2.6
5573	AGhgHgKWKKX	64	398.2621	3	26.08	1191.761	3.3
5728	AGAgFSKWKKX	64	383.5692	3	26.51	1147.687	-0.8
6319	GGLTDgKWKKX	64	386.9044	3	28.23	1157.692	-0.7
6930	AGGQfFKWKKX	64	318.6814	4	30.11	1270.698	-0.9
8054	lPDGgPKWKKX	64	440.9242	3	34.1	1319.743	5.9
3824	AGAAgRKWKKX	63	381.2486	3	21.57	1140.724	-0.3
3949	LTLGGTKWKKX	63	377.5728	3	21.89	1129.697	-0.6
4426	NDbSTFKWKKX	63	426.9034	3	23.08	1277.688	0.3
4458	HHGAGGKWKKX	63	368.8767	3	23.16	1103.61	-1.7
4509	hLPPPGKWKKX	63	378.9096	3	23.29	1133.707	-0.3
4562	AFLDhAKWKKX	63	298.4343	4	23.42	1189.697	9
4612	agghAGKWKKX	63	385.9291	3	23.54	1154.765	0.1
4667	GELbNGKWKKX	63	395.5692	3	23.68	1183.683	2.5
5182	GGGfQGKWKKX	63	389.8856	3	25.02	1166.635	0.2
5212	GGGfQGKWKKX	63	389.8856	3	25.1	1166.635	0.2
5750	AGAgFSKWKKX	63	383.5689	3	26.57	1147.687	-1.4
3695	ARLPGGKWKKX	62	285.6843	4	21.25	1138.709	-0.7
3841	QAFRNdKWKKX	62	345.9555	4	21.61	1379.794	-0.5
3954	SPPKbQKWKKX	62	313.6975	4	21.9	1250.761	-0.1
4175	GSLEGGKWKKX	62	363.5454	3	22.46	1087.614	0.7
4502	ALAhgQKWKKX	62	296.7001	4	23.27	1182.76	9.2
5236	GGGfQGKWKKX	62	389.8855	3	25.16	1166.635	-0.1
5713	AGAgFSKWKKX	62	287.9286	4	26.47	1147.687	-1
5977	GGEgNLKWKKX	62	395.9082	3	27.22	1184.703	-0.1
6838	ANfLTaKWKKX	62	437.596	3	29.82	1309.766	0.3
7537	DTgPQgKWKKX	62	428.5992	3	32.31	1282.776	-0.3
4531	ALRgAGKWKKX	61	296.7001	4	23.34	1182.771	0.1
4532	ALhAgQKWKKX	61	296.7001	4	23.34	1182.76	9.3
5239	GGhTQeKWKKX	61	405.2328	3	25.17	1212.677	-0.3
5605	AGhgHgKWKKX	61	298.9472	4	26.17	1191.761	-0.6
5805	GfGDGAKWKKX	61	390.2136	3	26.72	1167.619	0

7533	DTPgQgKWKKX	61	428.5992	3	32.3	1282.776	-0.2
7641	egGGSeKWKKX	61	437.5837	3	32.65	1309.733	-3.2
8204	ALaPAiKWKKX	61	398.5775	3	34.54	1192.719	-7.1
8395	PTDDNgKWKKX	61	419.9032	3	35.14	1256.688	0.1
8632	AHaFFGKWKKX	61	416.5744	3	36.03	1246.709	-5.8
8676	AHaFFGKWKKX	61	416.5737	3	36.17	1246.709	-7.7
8694	AHaFFGKWKKX	61	416.5747	3	36.23	1246.709	-5.1
8720	AHaFFGKWKKX	61	416.5748	3	36.32	1246.709	-4.9
9818	KiiSDeKWKKX	61	474.5873	3	40.15	1420.736	2.7
3587	hGgAdEKWKKX	60	411.9202	3	20.99	1232.739	-0.4
3826	KNTGghKWKKX	60	300.9449	4	21.58	1199.75	0.2
4077	GGAFNTKWKKX	60	379.2173	3	22.21	1134.63	0
4180	GGfhKSKWKKX	60	409.2448	3	22.47	1224.713	-0.4
4196	GGfhKSKWKKX	60	409.2451	3	22.51	1224.713	0.2
4525	GQcSNSKWKKX	60	401.8979	3	23.33	1202.663	7.1
4675	SSfAGGKWKKX	60	390.8857	3	23.7	1169.635	0.5
4892	LDDbQGKWKKX	60	414.9071	3	24.26	1241.688	9.3
5101	FgESATKWKKX	60	417.5759	3	24.81	1249.718	-9.9
5614	PADTgDKWKKX	60	405.5672	3	26.19	1213.682	-1.7
5731	AGAgFSKWKKX	60	287.9287	4	26.52	1147.687	-0.7
5768	AGAgFSKWKKX	60	383.569	3	26.62	1147.687	-1.1
6405	DFSAgbKWKKX	60	421.2521	3	28.48	1260.734	0.2
7652	PbTNSeKWKKX	60	437.5838	3	32.69	1309.73	0
8683	AHaFFGKWKKX	60	416.5746	3	36.19	1246.709	-5.5
8701	AHaFFGKWKKX	60	416.5742	3	36.25	1246.709	-6.5
8705	AHaFFGKWKKX	60	416.5743	3	36.26	1246.709	-6.2
9722	QiNbDeKWKKX	60	474.5874	3	39.86	1420.736	2.9
9772	QbiNDeKWKKX	60	474.5875	3	40.01	1420.736	3
3783	LLhAEQKWKKX	59	409.9274	3	21.47	1226.75	8.5
3847	ALaKAFKWKKX	59	305.4512	4	21.63	1217.776	-0.3
4672	SSfAGGKWKKX	59	390.8854	3	23.69	1169.635	-0.2
4759	hGNhAfKWKKX	59	306.6844	4	23.92	1222.698	8.9
4776	AAGGRfKWKKX	59	306.6844	4	23.97	1222.709	-0.3
5215	GGhTQeKWKKX	59	405.2329	3	25.11	1212.677	0.1
5344	GSGSddKWKKX	59	410.2379	3	25.46	1227.688	3.7
5992	GGEgNLKWKKX	59	395.9079	3	27.26	1184.703	-0.8
6370	DSFAgbKWKKX	59	421.2516	3	28.38	1260.734	-1.1
7280	GgTeGGKWKKX	59	395.5694	3	31.33	1183.687	-0.1
7558	DTPgQgKWKKX	59	428.5993	3	32.37	1282.776	0
7666	PbTNSeKWKKX	59	437.5835	3	32.73	1309.73	-0.6
8677	AeaSFGKWKKX	59	416.5737	3	36.18	1246.697	1.5
9807	aSHeDeKWKKX	59	474.5871	3	40.12	1420.74	-0.5
3680	LFNASSKWKKX	58	302.6818	4	21.22	1206.687	9

3946	NAPPLHKWKKX	58	305.1866	4	21.88	1216.719	-1.7
4059	QNEQSFKWKKX	58	441.2383	3	22.17	1320.694	-0.6
5716	DfSAGAKWKKX	58	404.8885	3	26.47	1211.645	-1
5909	hAKhSSKWKKX	58	377.906	3	27.02	1130.692	3.3
6132	TPDGTGKWKKX	58	372.8765	3	27.68	1115.609	-1
7544	DTgPQgKWKKX	58	428.5991	3	32.33	1282.776	-0.5
3887	AKGiTQKWKKX	57	307.4331	4	21.72	1225.704	-0.7
4740	ADNbLGKWKKX	57	395.5693	3	23.88	1183.683	2.9
5257	GShAQeKWKKX	57	405.2331	3	25.21	1212.677	0.6
5609	GShRDFKWKKX	57	412.5729	3	26.18	1234.694	2.7
7580	DTPgQgKWKKX	57	428.5989	3	32.44	1282.776	-0.9
7589	DTgPQgKWKKX	57	428.599	3	32.47	1282.776	-0.7
7949	LKGEiKKWKKX	57	432.9326	3	33.76	1295.783	-4.9
3795	ASSTcbKWKKX	56	401.5767	3	21.5	1201.704	3.2
3865	ALaKAFKWKKX	56	305.4512	4	21.67	1217.776	-0.3
4011	AGFGGTKWKKX	56	360.21	3	22.05	1077.608	-0.2
4097	hGGGGFKWKKX	56	350.2066	3	22.26	1047.598	0.1
4398	DAGhTFKWKKX	56	388.8889	3	23.02	1163.645	-0.4
4576	PPSEEaKWKKX	56	409.8994	3	23.45	1226.677	-0.4
4799	GhGGRfKWKKX	56	306.6844	4	24.02	1222.709	-0.2
5930	DhAgdLKWKKX	56	425.9355	3	27.08	1274.786	-1.3
7010	NPSLFEKWKKX	56	425.9156	3	30.34	1274.714	8.8
7529	DTgPQgKWKKX	56	428.599	3	32.29	1282.776	-0.8
7595	DTPgQgKWKKX	56	428.599	3	32.49	1282.776	-0.7
8296	KAADmKKWKKX	56	431.9322	3	34.81	1292.78	-4.3
3862	cEGSaFKWKKX	55	313.433	4	21.66	1249.704	-1
3960	QAPLbQKWKKX	55	417.9276	3	21.92	1250.761	-0.2
4150	GSgDGGKWKKX	55	363.545	3	22.4	1087.614	-0.4
4318	QFDHGGKWKKX	55	308.1689	4	22.82	1228.647	0.1
4579	AFLDhAKWKKX	55	298.4342	4	23.46	1189.697	8.7
5397	aTDGTGKWKKX	55	373.879	3	25.6	1118.62	-4
6343	ADdGSTKWKKX	55	399.2255	3	28.3	1194.651	3.2
7007	PNSLFEKWKKX	55	319.6884	4	30.33	1274.714	8.6
7020	PNSLFEKWKKX	55	319.6884	4	30.37	1274.714	8.6
7299	GgTeGGKWKKX	55	395.569	3	31.39	1183.687	-1.1
7615	DTQgPgKWKKX	55	428.5995	3	32.57	1282.776	0.4
8328	AEGHagKWKKX	55	403.9091	3	34.9	1208.714	-7.1
12670	PTgSDeKWKKX	55	438.2518	3	49.67	1311.734	-0.2
3558	LHHAEdKWKKX	54	451.2632	3	20.91	1350.767	0.3
3691	NSdaFGKWKKX	54	423.9124	3	21.24	1268.714	0.9
4329	GGFNKFKWKKX	54	413.5768	3	22.84	1237.708	0.1
5380	TaDGTGKWKKX	54	373.8811	3	25.56	1118.62	1.7

6316	ADdGSTKWKKX	54	399.2253	3	28.22	1194.651	2.5
6347	ADdGSTKWKKX	54	399.2253	3	28.31	1194.651	2.6
6355	ADdGSTKWKKX	54	399.2256	3	28.34	1194.651	3.3
8406	PTDDNgKWKKX	54	419.903	3	35.17	1256.688	-0.6
3567	hGgAdEKWKKX	53	411.9207	3	20.93	1232.739	0.7
3679	AhFcSSKWKKX	53	302.6818	4	21.21	1206.699	-0.3
4909	QaDbQGKWKKX	53	414.9068	3	24.31	1241.699	-0.4
5734	DfSAGAKWKKX	53	404.8886	3	26.52	1211.645	-0.9
6300	ADdGSTKWKKX	53	399.2249	3	28.17	1194.651	1.6
6351	ADdGSTKWKKX	53	399.2253	3	28.32	1194.651	2.7
6367	ADdGSTKWKKX	53	399.2254	3	28.37	1194.651	3
6388	ADdGSTKWKKX	53	399.2255	3	28.43	1194.651	3.2
6564	FaEAhfKWKKX	53	448.5919	3	28.94	1342.755	-0.8
3868	GGEgHPKWKKX	52	398.2362	3	21.68	1191.688	-0.8
4193	GfAGAHKWKKX	52	402.2295	3	22.51	1203.667	0.1
4240	AhLKHhKWKKX	52	302.7001	4	22.63	1206.772	-0.2
5334	GSAGmdKWKKX	52	410.2378	3	25.43	1227.696	-3.6
5343	GSAGmdKWKKX	52	410.2379	3	25.45	1227.696	-3.3
5516	GhiHSNKWKKX	52	306.1679	4	25.93	1220.653	-8.4
5615	APDgTDKWKKX	52	405.5672	3	26.19	1213.682	-1.7
6339	ADdGSTKWKKX	52	399.2256	3	28.29	1194.651	3.3
6359	ADdGSTKWKKX	52	399.2253	3	28.35	1194.651	2.5
6363	ADdGSTKWKKX	52	399.2253	3	28.36	1194.651	2.7
6375	ADdGSTKWKKX	52	399.2256	3	28.4	1194.651	3.3
6401	ADdGSTKWKKX	52	399.2256	3	28.47	1194.651	3.3
6425	ADdGSTKWKKX	52	399.2251	3	28.54	1194.651	2.2
4957	GGGPLDKWKKX	51	362.2136	3	24.43	1083.619	0.1
5931	DhAgdLKWKKX	51	425.9355	3	27.08	1274.786	-1.3
7358	GLEAHaKWKKX	51	399.2375	3	31.59	1194.699	-6.5
8322	PTFEGAKWKKX	51	397.5608	3	34.89	1189.661	0

5.11 Appendix V. Peptides identified from experimental samples

Whole cell extract, PMO-Library

Scan	Peptide	ALC (%)	m/z	z	RT	Mass	ppm
3025	hPPKWKKX	95	289.8567	3	17.64	866.5491	-1.1
3055	hPPKWKKX	95	289.8566	3	17.79	866.5491	-1.2
3038	PPhKWKKX	94	289.8566	3	17.71	866.5491	-1.2
3027	hPPKWKKX	93	289.8567	3	17.65	866.5491	-1.1
3040	PPhKWKKX	93	289.8566	3	17.72	866.5491	-1.2
3057	hPPKWKKX	92	289.8566	3	17.8	866.5491	-1.2
2634	LTERKWKKX	92	363.2253	3	15.78	1086.666	-11.3
3045	hPPKWKKX	91	289.8565	3	17.74	866.5491	-1.6
3047	hPPKWKKX	90	289.8565	3	17.75	866.5491	-1.6
3077	hPPKWKKX	89	289.8564	3	17.89	866.5491	-2.1
2630	LTERKWKKX	89	363.2252	3	15.76	1086.666	-11.5
2641	LTERKWKKX	88	363.2254	3	15.81	1086.666	-11
2781	LPDETKWKKX	87	381.8884	3	16.49	1142.645	-1.3
2805	LPDRjWKKX	82	381.8882	3	16.59	1142.635	7.1
2812	LPDETKWKKX	81	381.8882	3	16.62	1142.645	-1.8
2782	LPDETKWKKX	78	381.8884	3	16.49	1142.645	-1.3
3085	AhShgKWKKX	78	348.5618	3	17.93	1042.665	-1.7
2783	KeETKWKKX	76	381.8884	3	16.5	1142.66	-14.6
2648	LTERKKWKX	75	363.2253	3	15.85	1086.666	-11.1
2793	LPDRWjKKX	75	381.8884	3	16.54	1142.635	7.5
2759	LPDETKWKKX	75	381.8884	3	16.39	1142.645	-1.2
2792	LPDETKWKKX	75	381.8884	3	16.54	1142.645	-1.2
2760	LPDETKWKKX	74	381.8884	3	16.39	1142.645	-1.2
2804	LPDETKWKKX	74	381.8882	3	16.59	1142.645	-1.6
2631	LTERKWKKX	73	363.2252	3	15.77	1086.666	-11.5
2621	LTERKWKKX	72	363.2255	3	15.72	1086.666	-10.8
2635	LTERKWKKX	72	363.2253	3	15.78	1086.666	-11.3
3078	LDagKWKKX	72	348.5616	3	17.9	1042.665	-1.9
2392	KDLQQKWKKX	71	300.9354	4	14.6	1199.714	-1.3
2394	KDLQQKWKKX	70	300.9354	4	14.61	1199.714	-1.3
2813	LPDETKWKKX	69	381.8882	3	16.63	1142.645	-1.8
2621	LTEAhKWKKX	68	363.2255	3	15.72	1086.655	-0.4
2753	LPDRjWKKX	67	381.8884	3	16.36	1142.635	7.6
2615	SbSGgKWKKX	67	358.2254	3	15.69	1071.655	-0.8
2635	LTPTSKWKKX	67	363.2253	3	15.78	1086.655	-0.9
2631	LTPTSKWKKX	66	363.2252	3	15.77	1086.655	-1.2
3087	AhATgKWKKX	65	348.5618	3	17.93	1042.665	-1.7
3080	LDagKWKKX	64	348.5616	3	17.9	1042.665	-1.9
2805	LPDETKWKKX	61	381.8882	3	16.59	1142.645	-1.6

2616	KDNgKWKKX	59	358.2254	3	15.7	1071.655	-0.8
3026	hPPKWKKX	58	289.8567	3	17.64	866.5491	-1.1
3056	hPPKWKKX	58	289.8566	3	17.79	866.5491	-1.2
2753	LPDETKWKKX	58	381.8884	3	16.36	1142.645	-1.2
3039	hPPKWKKX	57	289.8566	3	17.71	866.5491	-1.2
3060	hPPKWKKX	57	289.8566	3	17.81	866.5491	-1.4
3112	hPPKWKKX	57	289.8567	3	18.05	866.5491	-1.1
2625	AEaGgKWKKX	56	358.225	3	15.74	1071.655	-1.9
2663	LTERKWKKX	53	363.2252	3	15.92	1086.666	-11.6

Cytosolic extract, PMO-Library

Scan	Peptide	ALC (%)	m/z	z	RT	Mass	ppm
3566	ARRhaHKWKKX	90	324.2084	4	21.38	1292.805	-0.8
3584	ARRhaHKWKKX	89	324.2085	4	21.46	1292.805	-0.3
3504	ARhRaHKWKKX	88	324.2084	4	21.11	1292.805	-0.7
3568	ARRhaHKWKKX	88	324.2084	4	21.39	1292.805	-0.8
3494	ARhRaHKWKKX	87	324.2085	4	21.07	1292.805	-0.5
3586	ARRhaHKWKKX	87	324.2085	4	21.47	1292.805	-0.3
3514	LNgLTHKWKKX	86	324.2084	4	21.15	1292.808	-2.9
3496	ARhRaHKWKKX	85	324.2085	4	21.08	1292.805	-0.5
3516	LNgLTHKWKKX	85	324.2084	4	21.16	1292.808	-2.9
3506	ARhRaHKWKKX	84	324.2084	4	21.12	1292.805	-0.7
3698	ARhRaHKWKKX	80	324.2086	4	21.97	1292.805	-0.1
3579	cNLNHKWKKX	79	302.9305	4	21.44	1207.705	-9.9
3611	iHgKWKKX	78	335.8711	3	21.58	1004.603	-11.8
3612	iHgKWKKX	78	335.8711	3	21.58	1004.603	-11.8
3700	ARhRaHKWKKX	76	324.2086	4	21.98	1292.805	-0.1
3579	RkNHKWKKX	71	302.9305	4	21.44	1207.682	9.1
3574	KAhRHKWKKX	68	292.1911	4	21.42	1164.736	-0.3
3579	hGNLNHKWKKX	67	302.9305	4	21.44	1207.694	-0.6
3574	KRRHKWKKX	66	292.1911	4	21.42	1164.747	-9.9
3567	ARLcfKWKKX	59	324.2084	4	21.39	1292.798	4.9
3658	LNKAcHKWKKX	59	324.2082	4	21.79	1292.794	7.3
3664	LgmPFKWKKX	56	316.9525	4	21.81	1263.794	-10.7
3657	ARhRaHKWKKX	56	324.2082	4	21.78	1292.805	-1.4
3495	ARcfLKWKKX	51	324.2085	4	21.07	1292.798	5.2

5.12 References

(1) Tripathi, P. P.; Arami, H.; Banga, I.; Gupta, J.; Gandhi, S. Cell Penetrating Peptides in Preclinical and Clinical Cancer Diagnosis and Therapy. *Oncotarget* **2018**, *9* (98), 37252–37267. https://doi.org/10.18632/oncotarget.26442.
(2) Xie, J.; Bi, Y.; Zhang, H.; Dong, S.; Teng, L.; Lee, R. J.; Yang, Z. Cell-Penetrating Peptides in Diagnosis and Treatment of Human Diseases: From Preclinical Research to Clinical Application. *Frontiers in Pharmacology* **2020**, *11*, 697. https://doi.org/10.3389/fphar.2020.00697.
(3) Bechara, C.; Sagan, S. Cell-Penetrating Peptides: 20years Later, Where Do We Stand? *FEBS Letters* **2013**, *587* (12), 1693–1702. https://doi.org/10.1016/j.febslet.2013.04.031.
(4) Ingram, D. Clinical Update: MOMENTUM Multiple- Ascending Dose Study of SRP-5051 for Duchenne Muscular Dystrophy. 33.
(5) Kim, G. C.; Cheon, D. H.; Lee, Y. Challenge to Overcome Current Limitations of Cell-Penetrating Peptides. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2021**, *1869* (4), 140604. https://doi.org/10.1016/j.bbapap.2021.140604.
(6) Frankel, A. D.; Pabo, C. O. Cellular Uptake of the Tat Protein from Human Immunodeficiency Virus. *Cell* **1988**, *55* (6), 1189–1193. https://doi.org/10.1016/0092-8674(88)90263-2.
(7) Green, M.; Loewenstein, P. M. Autonomous Functional Domains of Chemically Synthesized Human Immunodeficiency Virus Tat Trans-Activator Protein. *Cell* **1988**, *55* (6), 1179–1188. https://doi.org/10.1016/0092-8674(88)90262-0.
(8) Jearawiriyapaisarn, N.; Moulton, H. M.; Buckley, B.; Roberts, J.; Sazani, P.; Fucharoen, S.; Iversen, P. L.; Kole, R. Sustained Dystrophin Expression Induced by Peptide-Conjugated Morpholino Oligomers in the Muscles of Mdx Mice. *Molecular Therapy* **2008**, *16* (9), 1624–1629. https://doi.org/10.1038/mt.2008.120.
(9) Wimley, W. C. Synthetic Molecular Evolution of Cell Penetrating Peptides. In *Cell Penetrating Peptides: Methods and Protocols*; Langel, Ü., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2022; pp 73–89. https://doi.org/10.1007/978-1-0716-1752-6_5.
(10) Kauffman, W. B.; Guha, S.; Wimley, W. C. Synthetic Molecular Evolution of Hybrid Cell Penetrating Peptides. *Nature Communications* **2018**, *9* (1), 2568. https://doi.org/10.1038/s41467-018-04874-6.
(11) Porosk, L.; Gaidutšik, I.; Langel, Ü. Approaches for the Discovery of New Cell-Penetrating Peptides. *Expert Opin Drug Discov* **2021**, *16* (5), 553–565. https://doi.org/10.1080/17460441.2021.1851187.
(12) Lee, E. Y.; Wong, G. C. L.; Ferguson, A. Machine Learning-Enabled Discovery and Design of Membrane-Active Peptides. *Bioorganic & medicinal chemistry* **2018**, *26* (10), 2708–2718. https://doi.org/10.1016/j.bmc.2017.07.012.
(13) Manavalan, B.; Subramaniyam, S.; Shin, T. H.; Kim, M. O.; Lee, G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *J. Proteome Res.* **2018**, *17* (8), 2715–2726. https://doi.org/10.1021/acs.jproteome.8b00148.
(14) Pandey, P.; Patel, V.; George, N. V.; Mallajosyula, S. S. KELM-CPPpred: Kernel Extreme Learning Machine Based Prediction Model for Cell-Penetrating Peptides. *J. Proteome Res.* **2018**, *17* (9), 3214–3222. https://doi.org/10.1021/acs.jproteome.8b00322.
(15) Schissel, C. K.; Mohapatra, S.; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep

Learning to Design Nuclear-Targeting Abiotic Miniproteins. *Nat. Chem.* **2021**, *13* (10), 992–1000. https://doi.org/10.1038/s41557-021-00766-3.

(16) López-Vidal, E. M.; Schissel, C. K.; Mohapatra, S.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L. Deep Learning Enables Discovery of a Short Nuclear Targeting Peptide for Efficient Delivery of Antisense Oligomers. *JACS Au* **2021**. https://doi.org/10.1021/jacsau.1c00327.

(17) Wolfe, J. M.; Fadzen, C. M.; Choo, Z.-N.; Holden, R. L.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Machine Learning To Predict Cell-Penetrating Peptides for Antisense Delivery. *ACS Cent Sci* **2018**, *4* (4), 512–520. https://doi.org/10.1021/acscentsci.8b00098.

(18) Gao, S.; Simon, M. J.; Hue, C. D.; Morrison, B.; Banta, S. An Unusual Cell Penetrating Peptide Identified Using a Plasmid Display-Based Functional Selection Platform. *ACS Chem. Biol.* **2011**, *6* (5), 484–491. https://doi.org/10.1021/cb100423u.

(19) Kamide, K.; Nakakubo, H.; Uno, S.; Fukamizu, A. Isolation of Novel Cell-Penetrating Peptides from a Random Peptide Library Using in Vitro Virus and Their Modifications. *International Journal of Molecular Medicine* **2010**, *25* (1), 41–51. https://doi.org/10.3892/ijmm_00000311.

(20) Hoffmann, K.; Milech, N.; Juraja, S. M.; Cunningham, P. T.; Stone, S. R.; Francis, R. W.; Anastasas, M.; Hall, C. M.; Heinrich, T.; Bogdawa, H. M.; Winslow, S.; Scobie, M. N.; Dewhurst, R. E.; Florez, L.; Ong, F.; Kerfoot, M.; Champain, D.; Adams, A. M.; Fletcher, S.; Viola, H. M.; Hool, L. C.; Connor, T.; Longville, B. A. C.; Tan, Y.-F.; Kroeger, K.; Morath, V.; Weiss, G. A.; Skerra, A.; Hopkins, R. M.; Watt, P. M. A Platform for Discovery of Functional Cell-Penetrating Peptides for Efficient Multi-Cargo Intracellular Delivery. *Scientific Reports* **2018**, *8* (1), 12538. https://doi.org/10.1038/s41598-018-30790-2.

(21) Carney, R. P.; Thillier, Y.; Kiss, Z.; Sahabi, A.; Heleno Campos, J. C.; Knudson, A.; Liu, R.; Olivos, D.; Saunders, M.; Tian, L.; Lam, K. S. Combinatorial Library Screening with Liposomes for Discovery of Membrane Active Peptides. *ACS Comb. Sci.* **2017**, *19* (5), 299–307. https://doi.org/10.1021/acscombsci.6b00182.

(22) Quartararo, A. J.; Gates, Z. P.; Somsen, B. A.; Hartrampf, N.; Ye, X.; Shimada, A.; Kajihara, Y.; Ottmann, C.; Pentelute, B. L. Ultra-Large Chemical Libraries for the Discovery of High-Affinity Peptide Binders. *Nat Commun* **2020**, *11* (1), 3183. https://doi.org/10.1038/s41467-020-16920-3.

(23) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C.; Santi, D. V. Identification of Highest-Affinity Ligands by Affinity Selection from Equimolar Peptide Mixtures Generated by Robotic Synthesis. *Proc Natl Acad Sci U S A* **1992**, *89* (10), 4505–4509. https://doi.org/10.1073/pnas.89.10.4505.

(24) Beck, S.; Jin, X.; Yin, J.; Kim, S.-H.; Lee, N.-K.; Oh, S.-Y.; Jin, X.; Kim, M.-K.; Kim, E.-B.; Son, J.-S.; Kim, S.-C.; Nam, D.-H.; Kim, S.-H.; Kang, S.-K.; Kim, H.; Choi, Y.-J. Identification of a Peptide That Interacts with Nestin Protein Expressed in Brain Cancer Stem Cells. *Biomaterials* **2011**, *32* (33), 8518–8528. https://doi.org/10.1016/j.biomaterials.2011.07.048.

(25) Wu, C.-H.; Kuo, Y.-H.; Hong, R.-L.; Wu, H.-C. α-Enolase–Binding Peptide Enhances Drug Delivery Efficiency and Therapeutic Efficacy against Colorectal Cancer. *Science Translational Medicine* **2015**, *7* (290), 290ra91-290ra91. https://doi.org/10.1126/scitranslmed.aaa9391.

(26) Wu, C.-H.; Liu, I.-J.; Lu, R.-M.; Wu, H.-C. Advancement and Applications of Peptide Phage Display Technology in Biomedical Science. *Journal of Biomedical Science* **2016**, *23* (1), 8. https://doi.org/10.1186/s12929-016-0223-x.
(27) Ren, Y.; Zhan, C.; Gao, J.; Zhang, M.; Wei, X.; Ying, M.; Liu, Z.; Lu, W. A D-Peptide Ligand of Integrins for Simultaneously Targeting Angiogenic Blood Vasculature and Glioma Cells. *Mol Pharm* **2018**, *15* (2), 592–601. https://doi.org/10.1021/acs.molpharmaceut.7b00944.
(28) Wei, X.; Zhan, C.; Shen, Q.; Fu, W.; Xie, C.; Gao, J.; Peng, C.; Zheng, P.; Lu, W. A D-Peptide Ligand of Nicotine Acetylcholine Receptors for Brain-Targeted Drug Delivery. *Angew Chem Int Ed Engl* **2015**, *54* (10), 3023–3027. https://doi.org/10.1002/anie.201411226.
(29) Huang, L.; Xie, J.; Bi, Q.; Li, Z.; Liu, S.; Shen, Q.; Li, C. Highly Selective Targeting of Hepatic Stellate Cells for Liver Fibrosis Treatment Using a D-Enantiomeric Peptide Ligand of Fn14 Identified by Mirror-Image MRNA Display. *Mol Pharm* **2017**, *14* (5), 1742–1753. https://doi.org/10.1021/acs.molpharmaceut.6b01174.
(30) Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S. Inhibiting HIV-1 Entry: Discovery of D-Peptide Inhibitors That Target the Gp41 Coiled-Coil Pocket. *Cell* **1999**, *99* (1), 103–115. https://doi.org/10.1016/s0092-8674(00)80066-5.
(31) Loftis, A. R.; Zhang, G.; Backlund, C.; Quartararo, A. J.; Pishesha, N.; Hanna, C. C.; Schissel, C. K.; Garafola, D.; Loas, A.; Collier, R. J.; Ploegh, H.; Irvine, D. J.; Pentelute, B. L. An in Vivo Selection-Derived d-Peptide for Engineering Erythrocyte-Binding Antigens That Promote Immune Tolerance. *PNAS* **2021**, *118* (34). https://doi.org/10.1073/pnas.2101596118.
(32) Petersen, L. K.; Christensen, A. B.; Andersen, J.; Folkesson, C. G.; Kristensen, O.; Andersen, C.; Alzu, A.; Sløk, F. A.; Blakskjær, P.; Madsen, D.; Azevedo, C.; Micco, I.; Hansen, N. J. V. Screening of DNA-Encoded Small Molecule Libraries inside a Living Cell. *J. Am. Chem. Soc.* **2021**, *143* (7), 2751–2756. https://doi.org/10.1021/jacs.0c09213.
(33) Fischer, P. m.; Zhelev, N. z.; Wang, S.; Melville, J. e.; Fåhraeus, R.; Lane, D. p. Structure–Activity Relationship of Truncated and Substituted Analogues of the Intracellular Delivery Vector Penetratin. *The Journal of Peptide Research* **2000**, *55* (2), 163–172. https://doi.org/10.1034/j.1399-3011.2000.00163.x.
(34) Vinogradov, A. A.; Gates, Z. P.; Zhang, C.; Quartararo, A. J.; Halloran, K. H.; Pentelute, B. L. Library Design-Facilitated High-Throughput Sequencing of Synthetic Peptide Libraries. *ACS Comb. Sci.* **2017**, *19* (11), 694–701. https://doi.org/10.1021/acscombsci.7b00109.
(35) Schmidt, S.; Adjobo-Hermans, M. J. W.; Kohze, R.; Enderle, T.; Brock, R.; Milletti, F. Identification of Short Hydrophobic Cell-Penetrating Peptides for Cytosolic Peptide Delivery by Rational Design. *Bioconjugate Chem.* **2017**, *28* (2), 382–389. https://doi.org/10.1021/acs.bioconjchem.6b00535.
(36) Som, A.; Reuter, A.; Tew, G. N. Protein Transduction Domain Mimics: The Role of Aromatic Functionality. *Angewandte Chemie International Edition* **2012**, *51* (4), 980–983. https://doi.org/10.1002/anie.201104624.
(37) Lönn, P.; Dowdy, S. F. Cationic PTD/CPP-Mediated Macromolecular Delivery: Charging into the Cell. *Expert Opin Drug Deliv* **2015**, *12* (10), 1627–1636. https://doi.org/10.1517/17425247.2015.1046431.

(38) Najjar, K.; Erazo-Oliveras, A.; Mosior, J. W.; Whitlock, M. J.; Rostane, I.; Cinclair, J. M.; Pellois, J.-P. Unlocking Endosomal Entrapment with Supercharged Arginine-Rich Peptides. *Bioconjugate Chem.* **2017**, *28* (12), 2932–2941. https://doi.org/10.1021/acs.bioconjchem.7b00560.

(39) Pomplun, S.; Shugrue, C. R.; Schmitt, A. M.; Schissel, C. K.; Farquhar, C. E.; Pentelute, B. L. Secondary Amino Alcohols: Traceless Cleavable Linkers for Use in Affinity Capture and Release. *Angewandte Chemie International Edition* **2020**, *59* (28), 11566–11572. https://doi.org/10.1002/anie.202003478.

(40) Fadzen, C. M.; Holden, R. L.; Wolfe, J. M.; Choo, Z.-N.; Schissel, C. K.; Yao, M.; Hanson, G. J.; Pentelute, B. L. Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. *Biochemistry* **2019**. https://doi.org/10.1021/acs.biochem.9b00413.

(41) Verdurmen, W. P. R.; Bovee-Geurts, P. H.; Wadhwani, P.; Ulrich, A. S.; Hällbrink, M.; van Kuppevelt, T. H.; Brock, R. Preferential Uptake of L- versus D-Amino Acid Cell-Penetrating Peptides in a Cell Type-Dependent Manner. *Chemistry & Biology* **2011**, *18* (8), 1000–1010. https://doi.org/10.1016/j.chembiol.2011.06.006.

(42) Schissel, C. K.; Farquhar, C. E.; Malmberg, A. B.; Loas, A.; Pentelute, B. L. *Cell-Penetrating D-Peptides Retain Antisense Morpholino Oligomer Delivery Activity*; 2021; p 2021.09.30.462617. https://doi.org/10.1101/2021.09.30.462617.

(43) Burlina, F.; Sagan, S.; Bolbach, G.; Chassaing, G. A Direct Approach to Quantification of the Cellular Uptake of Cell-Penetrating Peptides Using MALDI-TOF Mass Spectrometry. *Nature protocols* **2006**, *1* (1), 200.

(44) Rabideau, A. E.; Pentelute, B. L. Delivery of Non-Native Cargo into Mammalian Cells Using Anthrax Lethal Toxin. *ACS Chem. Biol.* **2016**, *11* (6), 1490–1501. https://doi.org/10.1021/acschembio.6b00169.

(45) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014**, *15* (5), 713–720. https://doi.org/10.1002/cbic.201300796.

(46) Verdurmen, W. P. R.; Thanos, M.; Ruttekolk, I. R.; Gulbins, E.; Brock, R. Cationic Cell-Penetrating Peptides Induce Ceramide Formation via Acid Sphingomyelinase: Implications for Uptake. *Journal of Controlled Release* **2010**, *147* (2), 171–179. https://doi.org/10.1016/j.jconrel.2010.06.030.

(47) Brock, D. J.; Kustigian, L.; Jiang, M.; Graham, K.; Wang, T.-Y.; Erazo-Oliveras, A.; Najjar, K.; Zhang, J.; Rye, H.; Pellois, J.-P. Efficient Cell Delivery Mediated by Lipid-Specific Endosomal Escape of Supercharged Branched Peptides. *Traffic* **2018**, *19* (6), 421–435. https://doi.org/10.1111/tra.12566.

(48) Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Liu, S.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by Automated Flow Chemistry. *Science* **2020**, *368* (6494), 980–987. https://doi.org/10.1126/science.abb2491.

(49) Kjeken, R.; Mousavi, S. A.; Brech, A.; Griffiths, G.; Berg, T. Wortmannin-Sensitive Trafficking Steps in the Endocytic Pathway in Rat Liver Endothelial Cells. *Biochem J* **2001**, *357* (Pt 2), 497–503.

(50) Dutta, D.; Donaldson, J. G. Search for Inhibitors of Endocytosis: Intended Specificity and Unintended Consequences. *Cellular Logistics* **2012**, *2* (4), 203–208. https://doi.org/10.4161/cl.23967.