Conformational Stabilization and Rapid Labeling of a 29-Residue Peptide by a Small Molecule Reaction Partner

Conformational stabilization and rapid labeling of a 29-residue peptide by a small molecule reaction partner

Authors

Ethan D. Evans[1], Zachary P. Gates[1], Zhen-Yu J. Sun[2,†], Alexander J. Mijalis[1], Bradley L. Pentelute[1,*]

Affiliations

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States.

[2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Ave., Boston, Massachusetts 02115, USA

Abstract

A 29-residue peptide (MP01) capable of performing a nucleophilic aromatic substitution reaction with a perfluoroaromatic was modified and characterized using experimental and computational techniques, with the goal of understanding and enhancing its reactivity. An improved MP01 variant (MP01-Gen4) exhibited properties reminiscent of larger proteins and enzymes, including structure-dependent reactivity and small molecule-mediated conformational change. MP01-Gen4 possessed a total of six point mutations and exhibited a reaction rate constant of 25.8 ± 1.8 M^{-1} s^{-1} at pH 7.4 and room temperature. This was approximately two orders of magnitude greater than its progenitor sequence and three orders of magnitude greater than background cysteine reactivity. We believe the majority of the reaction rate enhancement can be attributed to interaction of MP01-Gen4 with the perfluoroaromatic probe, which stabilizes a helical conformation. These results highlight the possibility of achieving reaction rate enhancement from midsized peptides.

Introduction

Proteins can access many different mechanisms in order to facilitate rapid chemical transformations. Studies on enzymes have suggested a host of general reaction rate-enhancing mechanisms including: residue activation[1], reaction pathway preorganization[2], transition state stabilization[3] and reactant destabilization.[4] It has been proposed that such mechanisms rely on combinations of electrostatics (proximal charges or dipole interactions for instance)[5], steric effects[6], in addition to energies of substrate binding[4] and desolvation[7] in order to achieve their rate enhancement.

The majority of reactive biopolymers are composed of long sequences of more than 100 amino acids. Some of the smallest, naturally occurring enzymes like lysozyme and barnase are still over this length.[8,9] Smaller still are 4-oxalocrotonate tautomerase (~62 residues, though it exists as a hexamer in solution)[10] and the 6 kDa mini-matrilysin enzyme fragment.[11]

Compared to these larger biopolymers, it is not obvious what mechanisms or features small to midsized, reactive peptides (≤30 amino acids) are capable of accessing. From a structural perspective, it may be challenging for short peptides to match the active sites often seen in larger proteins because they lack the size to adopt necessary folds. Although examples of short folded peptides exist[12–16], such peptides are often designed for structure and typically rely on disulfide bonds or cyclization to impart stability.[15] However, for reactivity, structural rigidity or even a single conformation is not necessarily required, as dynamics and coupled motions are often observed and suggested to play important roles in proteins.[17–19] In fact, enzymes displaying significant conformational dynamics have been reported[20] along with molten globule[21] and intrinsically disordered enzymes.[22] Along with this, significant advancements have been made developing amino acid and peptide-based catalysts for a host of different reactions suggesting that diverse reactivity can be achieved with minimal sequences.[23–26]

In vitro selections and screens are powerful approaches to identify short peptides capable of performing desired reactions. A specific vein of this research has focused on developing peptides that covalently react with small molecules at accelerated rates. These efforts have identified sequences that react, for example, with p-(chloromethyl)benzamide[27], palladium-activated iodofluorescein[28], 2-cyanobenzothiazole[29], diketones[30] or perfluoroaromatics.[31,32] Compared to rationally designed systems in which a biopolymer is biased toward specific structural features or mechanisms, less is known regarding the mechanistic solutions emergent from these library-based approaches.

Here we investigated the properties of a 29-residue peptide (MP01) originally isolated from an mRNA display selection to react with a perfluoroaromatic via nucleophilic aromatic substitution (S_NAr, Figure 1A). We studied MP01's mutational tolerance and conformational properties both experimentally and computationally. These techniques were used to alter and improve the MP01 sequence, culminating in a mutant containing six point mutations with a second order rate constant improved by a factor of ~90, delivering high conversion in minutes (Figure 1B and Table 1). Biochemical and biophysical analysis of the improved MP01-Gen4 demonstrated

increased helicity and a broad predicted structural landscape. Notably, both MP01-Gen4 and an unreactive Ser mutant were conformationally stabilized by the perfluoraromatic small molecule, suggestive of a binding interaction. We propose that this interaction between the peptide and the probe accounts for the majority of the reaction rate enhancement, as in protein molecules.[4]

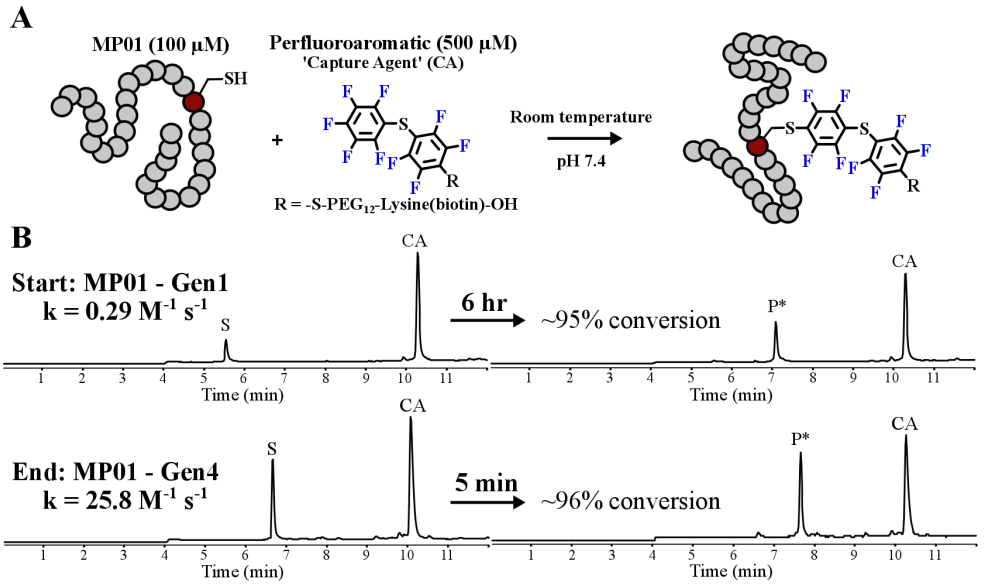


Figure 1. Rapid cysteine arylation by a 29-residue peptide. (A) Nucleophilic aromatic substitution reaction between MP01 and a pentafluorophenyl sulfide perfluoroaromatic (CA). (B) Liquid chromatography-mass spectrometry (LCMS) conversion analysis and rate constant comparison between the starting peptide (MP01-Gen1) and the final peptide (MP01-Gen4). S – Starting peptide, CA – capture agent, P* – peptide-CA conjugate following reaction.

Table 1. Key MP01 mutants show improved reaction rate constants relative to MP01-Gen1. All rates were measured using 100 µM peptide, 500 µM CA, 5 mM TCEP with 1x kinetics buffer (25 mM pH 7.45 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 100 mM NaCl, 5 mM $CaCl_2$ and 5 mM $MgCl_2$) at pH 7.4 and room temperature. Rate constant errors are standard deviations from either three individual replicates for MP01-Gen1 and MP01-Gen4 or the error of the linear regression fit of a single measurement for all others.

Name	Sequence	Second order rate (M^{-1} s^{-1})
MP01-Gen1	MHQKY KMTKD CFFSF LAHHK KRKLY PMSG	0.29 ± 0.01
MP01 3A	MHQKY KM**A**K**A** CFF**A**F LAHHK KRKLY PMSG	1.82 ± 0.03
MP01 H19L	MHQKY KMTKD CFFSF LAH**L**K KRKLY PMSG	2.17 ± 0.33
MP01-Gen2	MHQKY KM**A**K**A** CFF**A**F LAH**L**K KRKLY PMSG	9.35 ± 0.07
MP01-Gen3_1 A17E	MHQKY KM**A**K**A** CFF**A**F L**E**H**L**K KRKLY PMSG	25.2 ± 0.5
MP01-Gen3_2 H2N	M**N**QKY KM**A**K**A** CFF**A**F LAH**L**K KRKLY PMSG	12.8 ± 0.1
MP01-Gen4	M**N**QKY KM**A**K**A** CFF**A**F L**E**H**L**K KRKLY PMSG	25.8 ± 1.8

Results

Alanine scanning mutagenesis reveals beneficial and detrimental residues

To locate residues important for MP01's reactivity, we performed an alanine scan and measured the variants' reaction kinetics. This uncovered several residues that either helped or hindered reactivity (Figures 2A, S1–26). Synthetic peptide (0.1 mM) was treated with the perfluoroaromatic probe (0.5 mM and referred to as CA, Figure 2A inset), 5 mM tris(2-carboxyethyl)phosphine (TCEP) and 1x kinetics buffer (25 mM pH 7.45 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 100 mM NaCl, 5 mM $CaCl_2$ and 5 mM $MgCl_2$) at room temperature and analyzed by liquid chromatography-mass spectrometry (LCMS, details in SI, these were referred to as standard reaction conditions). The residues critical for full reactivity with the perfluoroaromatic probe were predominantly hydrophobic in nature and were concentrated within six residues of the active site cysteine. When the cysteine itself was mutated to serine, all reactivity was lost (Figure S11). Apart from these residues, the majority of the sequence appeared tolerant to alanine substitution. Several sites benefited from alanine placement, suggesting that the most reactive version of MP01 had yet to be found. These beneficial mutations enhanced the rate constant from 1.9- to 4.7-fold and appeared both close to and far from the active cysteine in primary sequence as evident by the H19A and P26A mutations. Additionally, the properties of the

residues varied from charged and polar to the rigid proline, providing little insight into their mechanism of action.

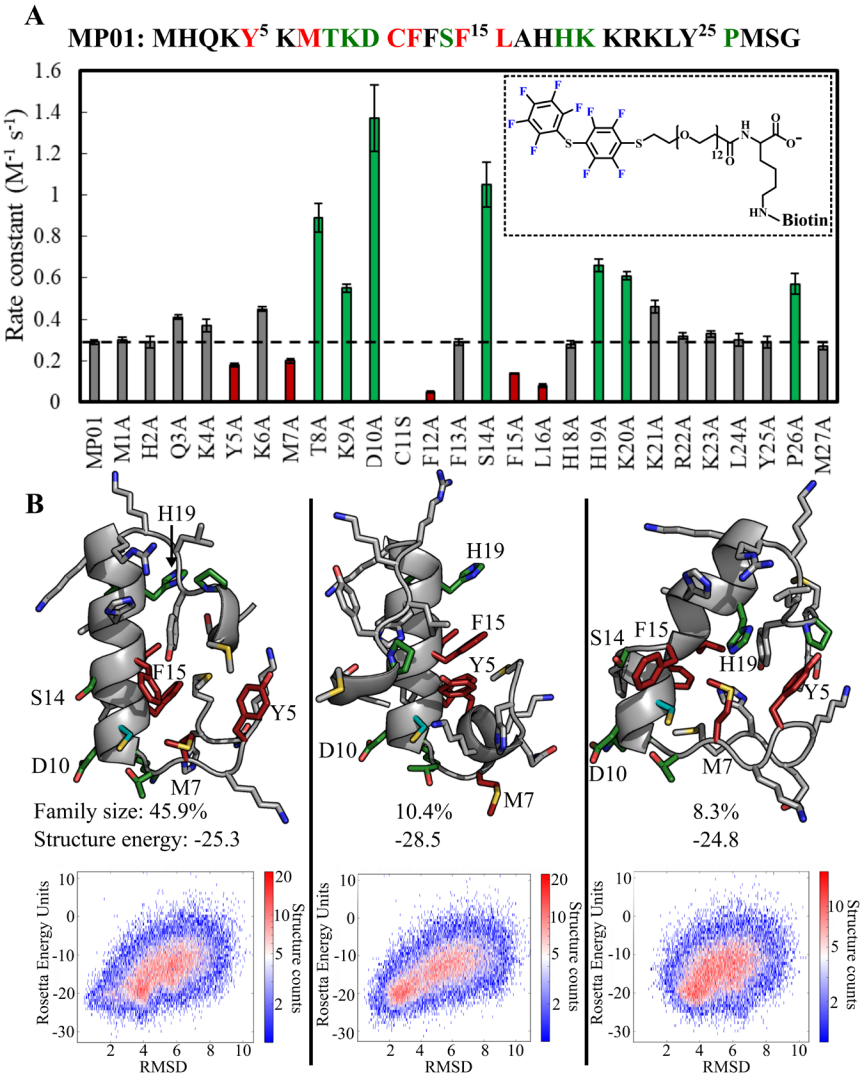


Figure 2. MP01's reactivity is modulated by alanine mutations and may possess helical elements. (A) Second order rate constants measured using standard reaction conditions for MP01 mutants. Beneficial mutations are colored in green, detrimental alanine mutations are in red. (B) Representative structures of the three largest clusters following structure prediction, clustering and family generation. Results of the alanine scan are mapped onto the structures with beneficial alanine sites in green, detrimental alanine positions in red and the active cysteine in teal. Family size percentages indicate how large that family was in the population with the number below corresponding to the Rosetta energy unit (REU) score for the representative structure. Energy vs root-mean-square deviation plots relative to the representative structure are shown at the bottom.

Structural modeling suggests the importance of a helical conformation

In an attempt to understand the alanine scan results and guide hypothesis driven experimental mutagenesis, we used Rosetta ab initio structure prediction[33,34] to model MP01. Structure clustering suggested a dominant structural family with several smaller, but similar neighbors (Figures S41 and S42). Representative structures were extracted from each family using a combination of low in-cluster energy and low in-cluster root-mean-square deviation (RMSD, Figures 2B top, S42). The predominant family accounted for 45.9% of structures; however, the energy difference between its representative structure and those of other clusters was minimal (-25.3 vs. -28.5 and -24.8 Rosetta energy units, Figure 2B). This similarity in energy, along with only minimally defined folding funnels (Figure 2B bottom) and its predominantly random coil-like circular dichroism (CD) signature[32] suggested that this sequence does not adopt a unique conformation and that a landscape approach to structure prediction would benefit interpretability.

A multistate modeling approach suggested that the majority of structural families adopted partial helical character with flanking random coil features. Analysis of these structures revealed that polar residues including K6, T8, D10 and S14 were often proximal to the active cysteine; however, alanine mutations at these positions did not hinder reactivity. These, in addition to other non-detrimental polar-to-alanine mutations (H2, K4, K9, H18, H19 and K20) suggested that polar / hydrogen bonding-based activation was not a predominant source of rate enhancement. In contrast to the polar residue mutations, several hydrophobic residues (Y5, M7, F12, F15 and L16) appeared critical for reactivity (Figure 2), perhaps due to stabilization of a reactive helical conformation or interaction with the CA. Further analysis revealed that in most families, the active cysteine was located in the N-terminal region of the primary alpha helix (shown in teal, Figure 2B), potentially taking advantage of helix dipole-based activation.[35] Based the alanine scan along with structural landscape modeling, we concluded that one or multiple helical conformations may be important for cysteine arylation, in conjunction with the critical hydrophobic residues.

Targeted alteration of MP01

I. Beneficial Ala mutants synergize to improve reactivity

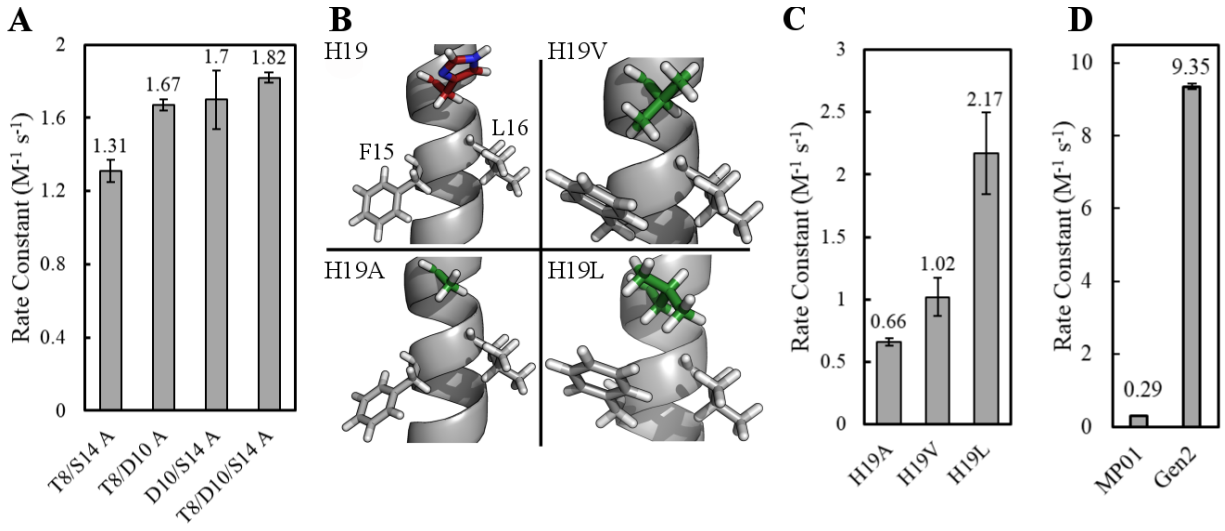


Figure 3. A combination of alanine scanning and experimental guided mutagenesis enhanced MP01's reactivity. (A) Second order rate constants of combined alanine mutations. (B) Model of residue position 19 relative to F15 and L16 for the representative structure of the largest cluster of MP01, MP01-H19A, MP01-H19V and MP01-H19L. H19 is shown in red with the hydrophobic mutations in green. (C) Second order rate constants of the three H19 variants. (D) Second order rate constant of MP01 vs. MP01-Gen2 composed of the T8/D10/S14 A mutations along with H19L.

Following up on the most beneficial Ala mutants (T8, D10, S14), we tested their synergy in an attempt to understand how removing sidechain functionality during the alanine scan improved reactivity. Each combination of double mutants showed improvement, leading to the triple alanine variant with a 1.82 ± 0.03 M^{-1} s^{-1} second order rate, representing a 6.3-fold increase in reactivity (MP01-T8/D10/S14 A, Figures 3A, S27–30). Analyzing the population changes between MP01 and its three alanine variant by modeling suggested a structural landscape expansion in terms of the family representation (Figures S43–46). The dominant family in MP01 decreased in size (-23.2%) and energy (-2.8), while several other families significantly increased in size and decreased in energy (Figures S45 and S46). In addition to potentially altering the structural landscape accessible, these mutations may have stabilized a helical conformation[36,37], improved accessibility of the CA to the reactive cysteine, or removed a negatively charged residue (D10) that could destabilize the Meisenheimer reaction intermediate.

II. A helix stabilizing mutation improves reactivity

The location of both H19 and its improved alanine mutant in reference to the critical F15 and L16 (*i, i+4* and *i, i+3*) suggested that substitution at this location with larger hydrophobic residues (V and L) may stabilize the predicted helical conformation[38] (Figure 3B). Under a model that placed importance on an alpha helix, stabilization of this feature was believed important in order to improve reactivity. In line with this notion, the chemically synthesized H19V and L mutations showed progressively increased rate constants of 3.5- and 7.5-fold respectively (Figures 3C, S31–32). Modeling of the H19L variant suggested that while the primary family from MP01 remained dominant and the energy of its representative structure decreased (-5.8 REU), its size slightly decreased (-5.19%, Figures S47–S49).

III. Combining mutations further improves reactivity

To determine whether the 3 Ala variant (MP01-T8/D10/S14 A) would synergize with the H19L mutation we synthesized a combined, four residue altered sequence (MP01-Gen2). This sequence afforded a 32.2-fold improved second order rate constant relative to MP01 and was superior to a 4 Ala variant where H19 was replaced with alanine (Figures 3D, S33 and S34). Computational modeling suggested that this peptide may adopt a number of different conformations (S50–S54). The major family from all previous variants was replaced by a new family (Gen2 cluster C2 in Figure S54); however, many of the previous dominant families still appeared at significant percentages. Additionally, none of the representative structures displayed a significant folding funnel (Figure S51). Attempts to combine other beneficial alanine point mutants with the Gen2 scaffold yielded no further benefits (Q3A, K9A and K20A for example, Figures S35–37).

IV. Additional helix stabilizing mutations further improve reactivity

Attempting to stabilize helical elements seen throughout Gen2's landscape led to a sequence possessing a second order rate constant ~90 times greater than the initial MP01 scaffold. Gen2's

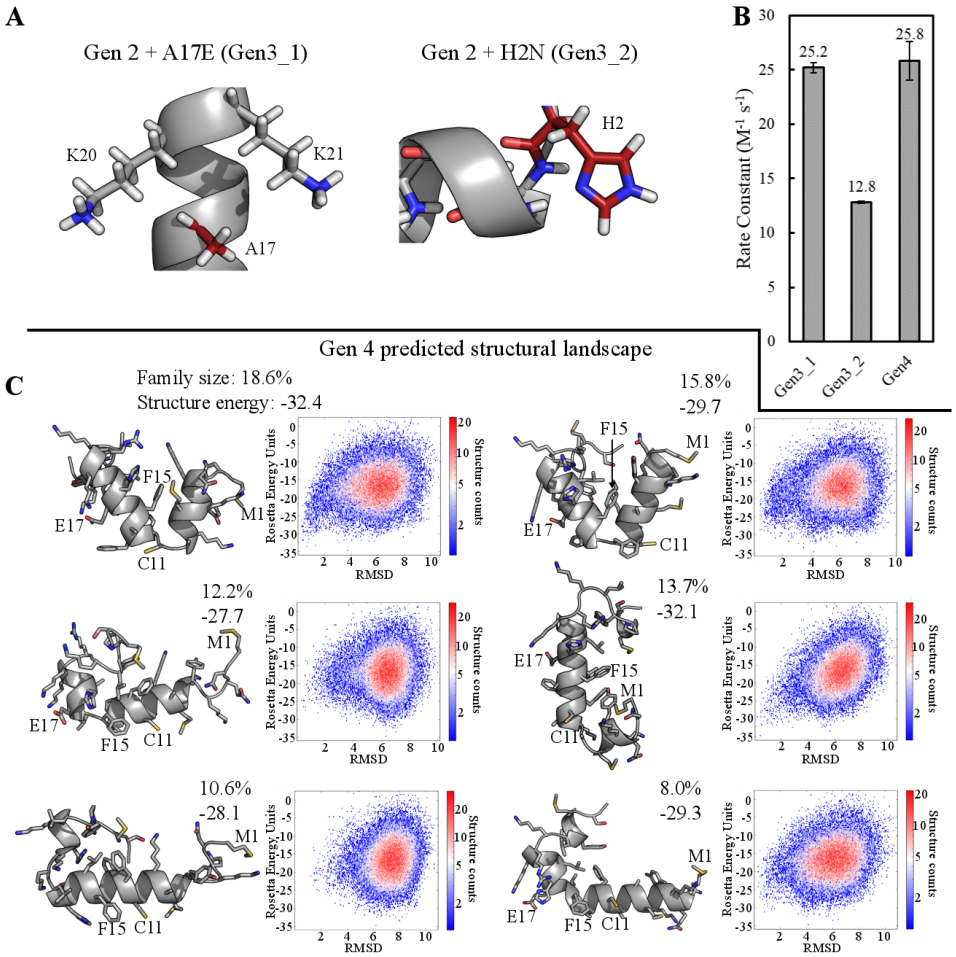


Figure 4. Structure guided mutations enhance MP01-Gen2's reactivity and broaden its predicted structural landscape with lower overall structure energies compared to Gen1. (A) Models showing the locations of mutation on the representative structure of the largest clusters of MP01-Gen2. (B) Second order rate constants of the Gen 3 and 4 sequences using standard reaction conditions. (C) Structural landscape analysis of Gen4 showing the size of the six largest families along with representative structures, their energies and energy vs. RMSD plots for the select structures.

A17 almost always appeared in an alpha helix three and four residues prior to a positively charged amino acid (K20/21) the structure of which could be stabilized by an *i*, *i+4* salt bridge (Figure 4A).[39] Furthermore we noted that H2 commonly adopted a position capping the N-terminal portion of a helix, a role that could also be stabilized by an asparagine residue.[40] Producing both of these variants (Gen3_1 for A17E and Gen3_2 for H2N) again each individually increased the rate of reaction with the CA, with a striking improvement provided by the A17E mutant (Figures 4B, S38 and S39). Combining these two alterations delivered the final MP01-Gen4 sequence possessing

six positions altered that was significantly more reactive than its progenitor scaffold and ~3.7×10^3 times more reactive than a random cysteine containing peptide (previously measured at 0.007 M^{-1} s^{-1}, Figures 4B and S40).[32]

Structural landscape modeling of Gen2 to Gen4 suggested a broad conformational space and no significant folding funnel for any of the representative structures (Figures 4C, S55–S60). Relative to the starting MP01 sequence, the energies of the representative structures had decreased, suggesting that structure itself was stabilized even though the structural landscape had broadened. Thus, while the mutations may have stabilized interactions, possibly helping the sequence adopt a reactive conformation, the overall conformational space was still predicted to be wide and similar in energy with transitions between states.

Biochemical and biophysical characterization of MP01-Gen4

Relative to MP01-Gen1, the fourth generation variant displayed increased alpha helicity as measured by CD at room temperature, consistent with lower Rosetta energies (Figure 5A). To test whether this secondary structure was important for reactivity we analyzed Gen4 under increasing concentrations of guanidinium chloride. Helicity decreased with denaturant concentration, even at the lowest concentration examined, demonstrating that the helical structure was minimally stable (Figure 5B and C). Thermal denaturation gave a similar outcome, reminiscent of helix-coil transitions for isolated helical peptides (Figure S96).[41,42] This loss of alpha helicity correlated with a commensurate drop in reactivity at low concentrations of guanidinium chloride with a near complete ablation of reaction from 2–3 M (Figures 5D, S62–S65). We note that denaturation has been shown previously not to alter the S_NAr mechanism and thus believe the decrease in reactivity to be attributed to the importance of structural elements.[32]

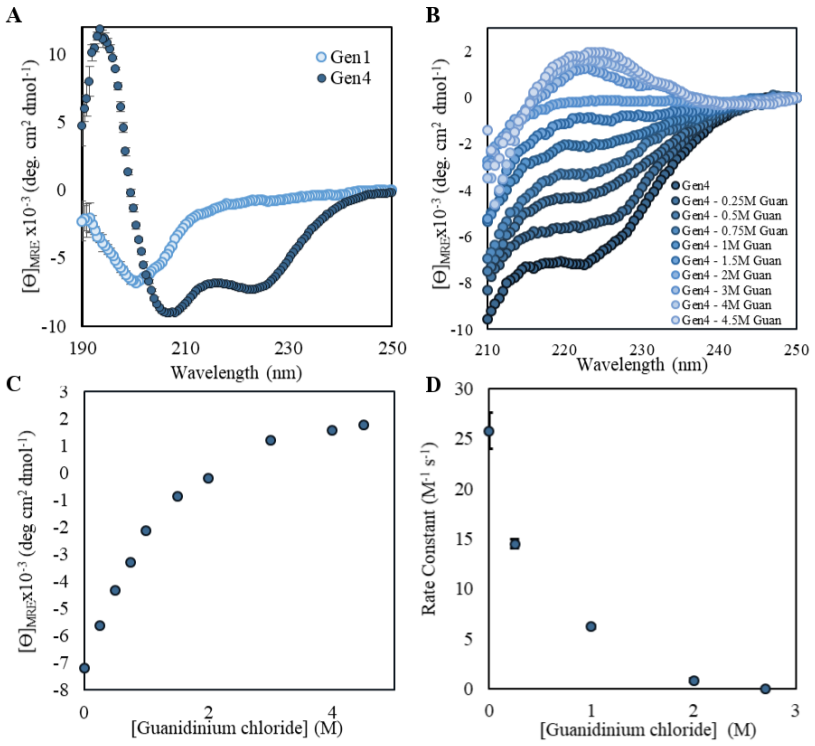


Figure 5. MP01-Gen4 displays increased helical features and reactivity that is sensitive to denaturant. (A) CD spectra of MP01-Gen1 and the final Gen4 variant using 50 µM MP in 10 mM phosphate buffer, 1 mM TCEP and pH 7.0 at room temperature. (B) CD spectra of MP01-Gen4 with select concentrations of guanidinum chloride (0–4.5 M). (C) Single wavelength mean residue ellipticity analysis of Gen4 as a function of denaturant concentration at 222 nm. (D) Guanidinium chloride-based Gen4 kinetics using standard reaction conditions with the addition of denaturant.

Gen4 displayed kinetics sensitive to truncations as well as select additives. Given a hypothesis that emphasized the importance of structural properties of this sequence, we synthesized a series of N and C-terminal truncations of the Gen4 sequence. This series of truncations displayed a progressive decrease in reactivity with longer truncations, perhaps suggesting a role for the majority of the sequence (Figures 6A, S69–S78). Having observed a secondary structure dependence along with a truncation sensitivity, we assayed Gen4 for reactivity under addition of additives that may impact its properties or its interactions with the CA. Using the structure stabilizing additive trimethylamine N-oxide (TMAO)[43] at 0.5 and 1 M, the rate constant decreased to 9.7 ± 0.6 M^{-1} s^{-1} and 7.5 ± 0.6 M^{-1} s^{-1} respectively, this may suggest the need for conformational flexibility prior to or during the reaction or that it stabilized a less reactive

conformation (Figure S66). Sodium chloride and ammonium sulfate both negatively impacted the reaction rate (3.9 ± 0.2 M^{-1} s^{-1} for 2 M ammonium sulfate and 6.1 ± 0.1 M^{-1} s^{-1} for 2 M sodium chloride, Figures S67–S68). This is in contrast to the four-residue π-clamp motif that also reacts with perfluoroaromatics, but displays a dramatic and beneficial salt-effect from ammonium sulfate.[44]

MP01-Gen4 exhibited pH-dependent kinetics. Profiling the reaction rate versus pH showed a roughly sigmoidal curve with a midpoint near pH 8 (Figures 6B and S79–S91). This observation is consistent with a thiolate acting as the nucleophile in the reaction and indirectly suggests that the pKa of the active cysteine is not significantly altered from that of a typical cysteine.[45] Therefore, we concluded that pKa modulation is not the primary determinant of MP01-Gen4's reactivity.

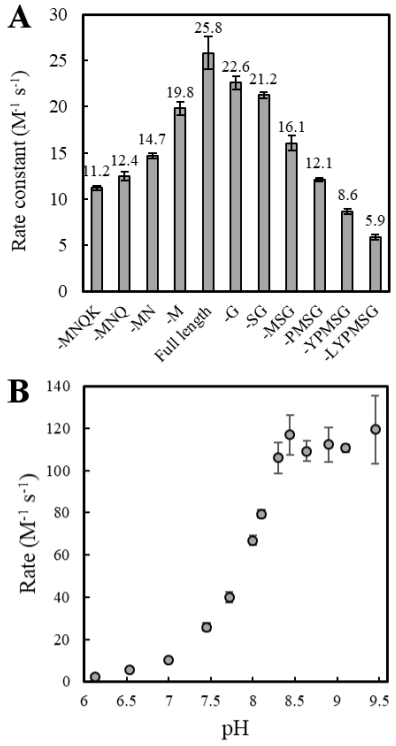


Figure 6. MP01-Gen4 requires a majority of its sequence for full reactivity and possesses pH sensitive kinetics. (A) Kinetic analysis of N- and C-terminal truncations of MP01-Gen4 using standard conditions. (B) Kinetic analysis of MP01 as a function of reaction pH. Both studies used 100 μM peptide and 500 μM CA.

To determine whether MP01-Gen4 adopted a tertiary structure, we performed natural abundance nuclear magnetic resonance (NMR) spectroscopy. ^{1}H NMR characterization of Gen4 revealed that at lower concentrations and temperatures (10 ºC), the amide proton region displayed increased signal intensities, an observation consistent with decreased conformational exchange and/or stabilization of structural hydrogen bonds (Figure S61). However, between 4 and 100 µM the amide proton resonances disappeared, due to apparent aggregation, which was previously observed for MP01-Gen1.[32] We acquired a ^{1}H-^{1}H NOESY spectrum at 5 mM MP01-Gen4. Few cross peaks were observed with the exception of several aromatic-to-aliphatic resonances, indicative of side chain interactions within (*i*, *i+3* or *i*, *i+4*) or between structural segments. Based on these results, we concluded that preparation of isotopically labeled MP01-Gen4 samples would be required for productive NMR analysis.

Small molecule induced conformational stabilization of MP01-Gen4

MP01-Gen4 possessed CA mediated structural properties. Following CA reaction and purification, the labeled version displayed a significantly increased alpha helical signature relative to its unlabeled state (50 µM Gen4-CA, Figure 7A). To determine whether this structural alteration was unique to Gen4 or had emerged from the selection[32], CD measurements were conducted on select MP01 variants before and after CA labeling (still in the reaction buffer with excess CA). This showed that all versions of MP01 altered their secondary structure upon labeling, suggesting that this structural alteration property was a feature emergent from the initial selection (Figure S92).

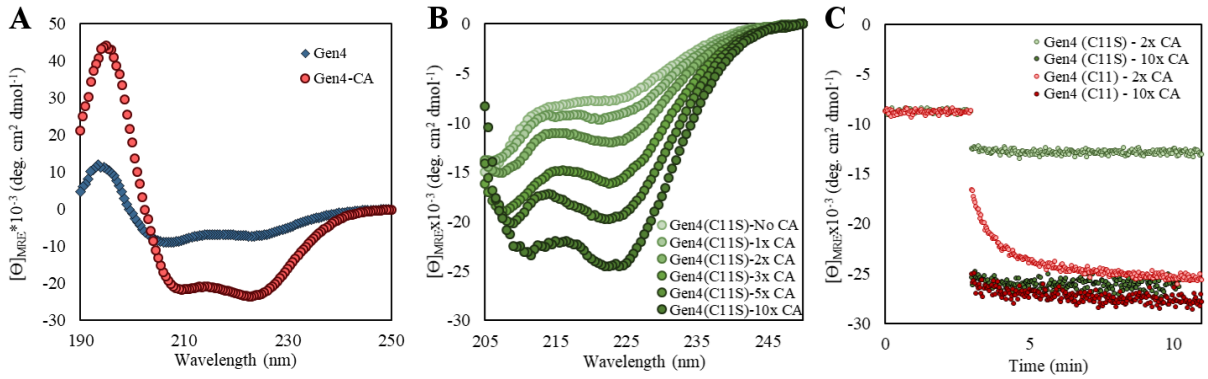


Figure 7. MP01-Gen4 displays CA-mediated structural alterations. (A) CD spectra of MP01-Gen4 and MP01-Gen4-CA at 24 ºC. (B) CD spectra of MP01-Gen4(C11S) at 50 µM with the CA at varying concentrations (0, 50, 100, 150, 250 and 500 µM). (C) Time-based CD measurement of MP01-Gen4 and MP01-Gen4(C11S) with the addition to 100 or 500 µM CA, followed at 222 nm.

To begin to test whether MP01 interacts with the CA we synthesized nonreactive variants for both MP01-Gen1 and Gen4: Gen1(C11S), Gen4(C11S) and Gen4(C11A). Increasing the concentration of CA present with all of these variants from 0 to a 10-fold excess (0–500 μM CA) showed progressively increasing helical signature, similar to the covalently labeled versions (Figures 7B, S93 and S94). The Gen1 version required higher CA concentrations to elicit conformational stabilization as up to a 2–3-fold excess showed minimal alteration while the Gen4 version already exhibited more pronounced change (Figures 7 and S93). Thermal melting of Gen4(C11S) at 50 μM with 10-fold CA (500 μM) showed a gradually increasing CD signal that remained well below the ellipticity values observed for either Gen4 or Gen4-CA, demonstrating that the CA conformationally stabilized MP01-Gen4(C11S) (Figure S96). We do not believe the CA to act as simply a structure promoting additive as the concentrations of CA used were much lower than those typically used with osmolytes in order to stabilize helicies (mM–M).[46,47] Instead, we believe that the CA directly interacts with the peptide, stabilizing one or multiple helical conformations.

Structural alteration was rapid and dependent on the CA concentration, as expected for a binding event. We characterized the rate at which this structural alteration occurred by measuring the CD signal at 222 nm for both the reactive form (C11) and the unreactive form (C11S) with select CA concentrations. For the unreactive form of Gen4 (C11S) we observed that the secondary structure alteration occurred rapidly and before measurements were started with ellipticity values similar to those obtained following overnight CA incubation (<10 s, Figure 7C). With a 2-fold CA excess, the reactive version not only displayed a rapid initial alteration but then exhibited a longer timescale alteration, arriving at a value similar to the purified, covalently labeled MP01-Gen4. This second phase of alteration occurred on a timescale commensurate with that observed for complete labeling via LCMS. A 10-fold excess of CA showed a large initial structural alteration for both the C11S and C11 Gen4 sequences. However, the reactive variant then only minimally increased in helicity suggesting that at this CA concentration, the fraction of Gen4 in the increased helical state immediately following the premeasurement phase was near unity. Thus while the MP would still undergo reaction, its secondary structure would not change. Extending this analysis to the C11 and C11S versions of Gen1 we observed a similar trend of rapid initial alteration dependent on the CA concentration used, in addition to a slow second phase of structural alteration for the 2-fold CA reaction mixture correlating with its much slower reaction kinetics (Figure S95).

The low concentration of CA used (μM), along with the rate of the structural transition, support an interpretation of the CD data as reporting on a binding interaction between MP01 and the CA.

Biophysical characterization of MP01-Gen4-CA

Gen4-CA exhibited properties reminiscent of globular proteins. Relative to its unlabeled version, higher concentrations of guanidinium chloride (greater than 1.5 M) were required for substantial alteration of Gen4-CA's secondary structure (Figure 8A). Nonlinear regression using a two state model of the mean residue molar ellipticity at 222 nm suggested a native-to-unfolded transition with a ΔG^o of 4.2 kcal mol^{-1} with an *m* value of 1.3 kcal mol^{-1} M^{-1} which is in the range of *m* values of known small proteins (Figure 8B).[48] Thermal melting of labeled and unlabeled Gen4 peptides (C11 present, followed from 4-95 ºC at 222 nm via CD) displayed a sharp unfolding transition for the CA-labeled version near 64 ºC compared with the unlabeled version that displayed a broad, gradually increasing trend similar to those seen from chemical denaturation (Figures 8C and S96).

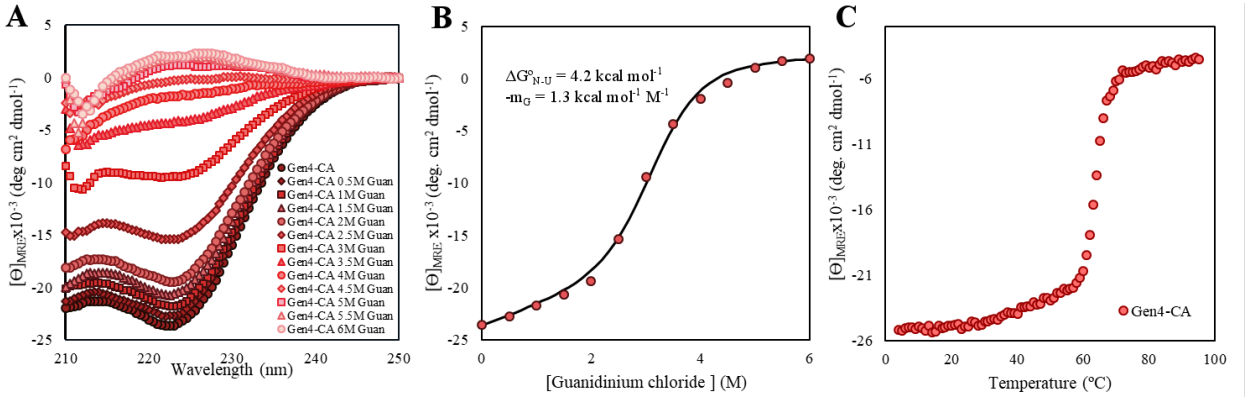


Figure 8. MP01-Gen4-CA displays increased stability relative to unlabeled MP01-Gen4. (A). CD analysis of MP01-Gen4-CA with increasing concentration of guanidinium chloride (0-6 M). (B). Two state modeling of chemical denaturant melt followed at 222 nm by CD (C). Thermal melting of 50 μM MP01-Gen4-CA followed by CD at 222 nm.

As an additional means of characterizing the structural stability of MP01-Gen4, we studied its protease stability. Gen4-CA displayed significantly increased protease resistance relative to the unlabeled Gen4 (Figures S97, S99 and S102). In line with the results from CD, Gen4(C11A and S) in the presence of 10-fold CA displayed an improved protease stability relative to Gen4 alone but was not as stable as the covalently bound form (Figures S100-S102). To probe the role of the CA in protease stability we synthesized a pentafluorophenyl sulfide (pfp) labeled Gen4, lacking the PEG-lysine-biotin portion of the CA. This variant displayed relatively high stability but was

less stable than the complete CA version (Figures S98 and S102). To gain a complete description of MP01-Gen4-CA's molecular structure, we attempted to crystallize this construct, with no success. This could reflect conformational heterogeneity or that the proper experimental conditions have yet to be determined.

Discussion

In this work, we demonstrated that MP01-Gen4 – a 29-amino-acid polypeptide with appreciable alpha helical structure – reacts with a small molecule probe ~3700-times greater than a random sequence cysteine-containing peptide. This rate enhancement is significantly larger than those observed in shorter peptide systems (<15 residues) where values of 2-[29], 5.2-[28] and 220-fold[49] relative to control nucleophiles have been observed and slightly greater than the 1000-fold value delivered by the π-clamp, a four-residue motif that accelerates cysteine arylation.[50] Of these systems, MP01 is the first to exhibit a link between extended secondary structure and enhanced chemical reactivity.

The data presented here indicate that midsized peptides can access mechanisms other than direct activation, like the modulation of cysteine pKa. We hypothesize that the MP01 variants interact with the perfluoroaromatic reaction partner in order to facilitate rapid reactivity due to an increased local concentation. A kinetic model in which rapid interaction precedes reaction yields an effective molarity (EM)[51] of 0.6 M (SI section 1.12 for derivation). This modest EM demonstrates that the observed rate enhancement can be accounted for by a relatively weak affinity interaction between MP01 and the CA (K_d ~150 μM, estimated from Figure 7B) without requiring a precise alignment of the cysteine and perfluoroarene or with perhaps multiple interaction conformations. While the structural basis of the interaction remains to be clarified, presumably even greater rate enhancements could be achieved with an optimized binding interaction geometry.

In the case of MP01, it appears that function is achieved without possessing a single conformation prior to reaction. In this regard, MP01-Gen4 is reminiscent of intrinsically disordered proteins that undergo disorder-to-order transitions upon performing their function.[52–54] This is not to say that MP01-Gen4 is completely unstructured since it possesses at least partial alpha helicity prior to reaction. Prediction of disorder suggests that the termini (residues 1-7 and 24-29) are disordered while the interior of the peptide may be more ordered.[55] These terminal regions may undergo structural alteration upon labeling (or interaction) and may be important for reactivity, as their removal is detrimental to function.

The use of structural modeling facilitated identification of helix stabilizing mutations (H2N, A17E and H19L) that improved the rate of cysteine arylation. However, we caution the interpretation that the presented models are necessarily the states populated by this sequence. Given such a short peptide and the experimentally observed CD spectra, the unlabeled MP01 peptide might actually display significantly more random-coil behavior than predicted due to conformational dynamics. Heavily structured landscapes may be due to the use of 9-mer fragments in the ab initio protocol. This could have shifted the results toward structure since fragments of this size represent 31% of the total length. Nonetheless, multistate modeling[56–58] may be applicable for identifying short, reactive polymers by de novo design, which has generally focused on longer polymers with defined tertiary structure.[59,60]

MP01-Gen4-CA displayed significant thermal and chemical stability. Often, isolated helical peptides exhibit low stability and broad thermal melting transitions[41], in contrast to the relatively sharp transition near 64 ºC for MP01-Gen4-CA. We propose that the stability of MP01-Gen4-CA can be explained by hydrophobic interactions between the CA and specific residues, helping lock the peptide into one or more helix-turn-helix conformations, accessible to the structural landscape. Additional studies are necessary to fully understand the conformation adopted MP01-Gen4-CA and the potential of perfluoroarene modification for stabilizing small polypeptide folds.

Conclusion

Here we studied the properties of a reactive midsized peptide. These features spanned from the effects of individual residues to structural considerations and substrate interactions. Mixing insights from structural landscape modeling with residue mutagenesis, we were able to alter and improve MP01's reactivity. We suggested that these mutations altered MP01's structural landscape, often through helix-stabilizing effects. Perhaps the most interesting observation was the conformational stabilization of MP01 in response to the small molecule probe, even under noncovalent conditions. While this study illustrates some of the features used by a single sequence, the extent of functional-sequence and -structure space conducive for a given reaction – S_NAr in this case – remains elusive. Further, it remains an open question as to the extent of possible reactions that can be performed by midsized peptides or whether they can fully rival their larger siblings in terms of breadth of reactivity. However, the rate acceleration and properties reported

here lends support to the notion that much can be achieved by a midsized biopolymer, hopefully kindling further studies on such sequences.

Associated Content

Supporting Information

This material is available free of charge at http://pubs.acs.org. Complete experimental protocols and reagents can be found in the supporting information. This includes all synthetic and kinetic methods in addition to CD, NMR and computational protocols, instrumentation and analysis.

Author Information

Corresponding Authors

*blp@mit.edu

Present Addresses

[†]Department of Cancer Biology, Dana Farber Cancer Institute, 360 Longwood Avenue, Boston, MA 02215.

Acknowledgments

We would like to acknowledge G. Wagner for establishing our NMR collaboration as well as G. Bhardwaj, V. Mulligan and D. Baker for initial guidance using Rosetta. We thank D. Dunkelmann for critical reading of the manuscript. We thank the Massachusetts Green High Performance Computing Center, in particular the Commonwealth Computational Cloud for Data Driven Biology without which this work would not have been feasible. The Biophysical Instrumentation Facility for the Study of Complex Macromolecular Systems (NSF-0070319) is gratefully acknowledged. This work was supported by an NSF graduate research fellowship (#112237) and the Martin Family Society of Fellows for Sustainability for E.D.E. and DARPA (Award #023504-001) for B.L.P. ZJS was supported by grants from the NIH to G. Wagner, GM047467 and EB002026 (P41 MIT).

References

(1) Blow, D. M.; Birktoft, J. J.; Hartley, B. S. Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin. *Nature* **1969**, *221* (5178), 337.

(2) Villà, J.; Warshel, A. Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B* **2001**, *105* (33), 7887–7907.

(3) Pauling, L. Nature of Forces between Large Molecules of Biological Interest*. *Nature* **1948**, *161*, 707.

(4) Jencks, W. P. Binding Energy, Specificity, and Enzymic Catalysis: The Circe Effect. In *Advances in Enzymology and Related Areas of Molecular Biology*; Meister, A., Ed.; John Wiley & Sons, Inc., 1975; pp 219–410.

(5) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106* (8), 3210–3235.

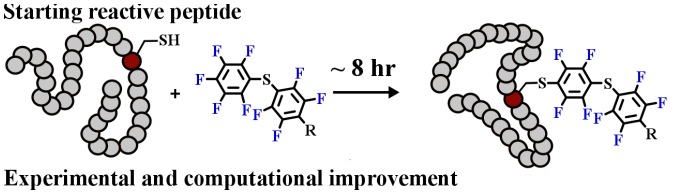
(6) Estell, D. A.; Graycar, T. P.; Miller, J. V.; Powers, D. B.; Wells, J. A.; Burnier, J. P.; Ng, P. G. Probing Steric and Hydrophobic Effects on Enzyme-Substrate Interactions by Protein Engineering. *Science* **1986**, *233* (4764), 659–663.
(7) Rucker, V. C.; Byers, L. D. An Assessment of Desolvation on Rates of Acetyl Transfer: Insights into Enzyme Catalysis. *J. Am. Chem. Soc.* **2000**, *122* (35), 8365–8369.
(8) Canfield, R. E. The Amino Acid Sequence of Egg White Lysozyme. *J. Biol. Chem.* **1963**, *238* (8), 2698–2707.
(9) Hartley, R. W.; Barker, E. A. Amino-Acid Sequence of Extracellular Ribonuclease (Barnase) of Bacillus Amyloliquefaciens. *Nature. New Biol.* **1972**, *235*, 15.
(10) Chen, L. H.; Kenyon, G. L.; Curtin, F.; Harayama, S.; Bembenek, M. E.; Hajipour, G.; Whitman, C. P. 4-Oxalocrotonate Tautomerase, an Enzyme Composed of 62 Amino Acid Residues per Monomer. *J. Biol. Chem.* **1992**, *267* (25), 17716–17721.
(11) Yu, W.-H.; Huang, P.-T.; Lou, K.-L.; Yu, S.-S. C.; Lin, C. A Smallest 6 Kda Metalloprotease, Mini-Matrilysin, in Living World: A Revolutionary Conserved Zinc-Dependent Proteolytic Domain- Helix-Loop-Helix Catalytic Zinc Binding Domain (ZBD). *J. Biomed. Sci.* **2012**, *19* (1), 54.
(12) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997**, *278* (5335), 82–87.
(13) Struthers, M. D.; Cheng, R. P.; Imperiali, B. Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure. *Science* **1996**, *271* (5247), 342–345.
(14) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan Zippers: Stable, Monomeric β-Hairpins. *Proc. Natl. Acad. Sci.* **2001**, *98* (10), 5578–5583.
(15) Bhardwaj, G.; Mulligan, V. K.; Bahl, C. D.; Gilmore, J. M.; Harvey, P. J.; Cheneval, O.; Buchko, G. W.; Pulavarti, S. V. S. R. K.; Kaas, Q.; Eletsky, A.; et al. Accurate de Novo Design of Hyperstable Constrained Peptides. *Nature* **2016**, *538* (7625), 329–335.
(16) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Mol. Biol.* **2002**, *9* (6), 425–430.
(17) Hammes, G. G. Multiple Conformational Changes in Enzyme Catalysis. *Biochemistry (Mosc.)* **2002**, *41* (26), 8221–8228.
(18) Tsou, C. L. Conformational Flexibility of Enzyme Active Sites. *Science* **1993**, *262* (5132), 380–381.
(19) Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. T. R.; Benkovic, S. J.; Hammes-Schiffer, S. Network of Coupled Promoting Motions in Enzyme Catalysis. *Proc. Natl. Acad. Sci.* **2002**, *99* (5), 2794–2799.
(20) Chao, F.-A.; Morelli, A.; Iii, J. C. H.; Churchfield, L.; Hagmann, L. N.; Shi, L.; Masterson, L. R.; Sarangi, R.; Veglia, G.; Seelig, B. Structure and Dynamics of a Primordial Catalytic Fold Generated by *in Vitro* Evolution. *Nat. Chem. Biol.* **2013**, *9* (2), 81.
(21) Vamvaca, K.; Vögeli, B.; Kast, P.; Pervushin, K.; Hilvert, D. An Enzymatic Molten Globule: Efficient Coupling of Folding and Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (35), 12860–12864.
(22) Palombo, M.; Bonucci, A.; Etienne, E.; Ciurli, S.; Uversky, V. N.; Guigliarelli, B.; Belle, V.; Mileo, E.; Zambelli, B. The Relationship between Folding and Activity in UreG, an Intrinsically Disordered Enzyme. *Sci. Rep.* **2017**, *7* (1), 5977.
(23) Gutte, B.; Däumigen, M.; Wittschieber, E. Design, Synthesis and Characterisation of a 34-Residue Polypeptide That Interacts with Nucleic Acids. *Nature* **1979**, *281*, 650.
(24) Davie, E. A. C.; Mennen, S. M.; Xu, Y.; Miller, S. J. Asymmetric Catalysis Mediated by Synthetic Peptides. *Chem. Rev.* **2007**, *107* (12), 5759–5812.
(25) Johnsson, K.; Allemann, R. K.; Widmer, H.; Benner, S. A. Synthesis, Structure and Activity of Artificial, Rationally Designed Catalytic Polypeptides. *Nature* **1993**, *365*, 530.
(26) Hahn, K.; Klis, W.; Stewart, J. Design and Synthesis of a Peptide Having Chymotrypsin-like Esterase Activity. *Science* **1990**, *248* (4962), 1544.
(27) Kawakami, T.; Ogawa, K.; Goshima, N.; Natsume, T. DIVERSE System: De Novo Creation of Peptide Tags for Non-Enzymatic Covalent Labeling by In Vitro Evolution for Protein Imaging Inside Living Cells. *Chem. Biol.* **2015**, *22* (12), 1671–1679.
(28) Lim, R. K. V.; Li, N.; Ramil, C. P.; Lin, Q. Fast and Sequence-Specific Palladium-Mediated Cross-Coupling Reaction Identified from Phage Display. *ACS Chem. Biol.* **2014**, *9* (9), 2139–2148.
(29) Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-Specific 2-Cyanobenzothiazole Ligation. *J. Am. Chem. Soc.* **2016**, *138* (17), 5499–5502.
(30) Tanaka, F.; Fuller, R.; Asawapornmongkol, L.; Warsinke, A.; Gobuty, S.; Barbas, III, C. F. Development of a Small Peptide Tag for Covalent Labeling of Proteins. *Bioconjug. Chem.* **2007**, *18* (4), 1318–1324.

(31) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π-Clamp-Mediated Cysteine Conjugation. *Nat. Chem.* **2016**, *8*, 120–128.
(32) Evans, E. D.; Pentelute, B. L. Discovery of a 29-Amino-Acid Reactive Abiotic Peptide for Selective Cysteine Arylation. *ACS Chem. Biol.* **2018**, *13* (3), 527–532.
(33) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences Using Simulated Annealing and Bayesian Scoring Functions. *J. Mol. Biol.* **1997**, *268* (1), 209–225.
(34) Bradley, P.; Misura, K. M. S.; Baker, D. Toward High-Resolution de Novo Structure Prediction for Small Proteins. *Science* **2005**, *309* (5742), 1868–1871.
(35) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. The α-Helix Dipole and the Properties of Proteins. *Nature* **1978**, *273* (5662), 443–446.
(36) Nick Pace, C.; Martin Scholtz, J. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998**, *75* (1), 422–427.
(37) Myers, J. K.; Pace, C. N.; Scholtz, J. M. Helix Propensities Are Identical in Proteins and Peptides. *Biochemistry (Mosc.)* **1997**, *36* (36), 10923–10929.
(38) Padmanabhan, S.; Baldwin, R. L. Tests for Helix-Stabilizing Interactions between Various Nonpolar Side Chains in Alanine-Based Peptides. *Protein Sci.* **1994**, *3* (11), 1992–1997.
(39) Marqusee, S.; Baldwin, R. L. Helix Stabilization by Glu-...Lys+ Salt Bridges in Short Peptides of de Novo Design. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (24), 8898–8902.
(40) Chakrabartty, A.; Doig, A. J.; Baldwin, R. L. Helix Capping Propensities in Peptides Parallel Those in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (23), 11332–11336.
(41) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. Unusually Stable Helix Formation in Short Alanine-Based Peptides. *Proc. Natl. Acad. Sci.* **1989**, *86* (14), 5286.
(42) Richardson, J. M.; Makhatadze, G. I. Temperature Dependence of the Thermodynamics of Helix–Coil Transition. *J. Mol. Biol.* **2004**, *335* (4), 1029–1037.
(43) Zou, Q.; Bennion, B. J.; Daggett, V.; Murphy, K. P. The Molecular Mechanism of Stabilization of Proteins by TMAO and Its Ability to Counteract the Effects of Urea. *J. Am. Chem. Soc.* **2002**, *124* (7), 1192–1202.
(44) Dai, P.; Zhang, C.; Welborn, M.; Shepherd, J. J.; Zhu, T.; Van Voorhis, T.; Pentelute, B. L. Salt Effect Accelerates Site-Selective Cysteine Bioconjugation. *ACS Cent. Sci.* **2016**, *2* (9), 637–646.
(45) Grimsley Gerald R.; Scholtz J. Martin; Pace C. Nick. A Summary of the Measured pK Values of the Ionizable Groups in Folded Proteins. *Protein Sci.* **2008**, *18* (1), 247–251.
(46) Dempsey, C. E.; Mason, P. E.; Brady, J. W.; Neilson, G. W. The Reversal by Sulfate of the Denaturant Activity of Guanidinium. *J. Am. Chem. Soc.* **2007**, *129* (51), 15895–15902.
(47) Celinski, S. A.; Scholtz, J. M. Osmolyte Effects on Helix Formation in Peptides and the Stability of Coiled-Coils. *Protein Sci. Publ. Protein Soc.* **2002**, *11* (8), 2048–2051.
(48) Myers, J. K.; Pace, C. N.; Scholtz, J. M. Denaturant M Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Sci. Publ. Protein Soc.* **1995**, *4* (10), 2138–2148.
(49) Chi, Z.; Peng, D.; Vinogradov, A. A.; Gates, Z. P.; Pentelute, B. L. Site-Selective Cysteine–Cyclooctyne Conjugation. *Angew. Chem. Int. Ed.* **2018**.
(50) Dai, P.; Williams, J. K.; Zhang, C.; Welborn, M.; Shepherd, J. J.; Zhu, T.; Van Voorhis, T.; Hong, M.; Pentelute, B. L. A Structural and Mechanistic Study of π-Clamp-Mediated Cysteine Perfluoroarylation. *Sci. Rep.* **2017**, *7* (1), 7954.
(51) Kirby, A. J. Effective Molarities for Intramolecular Reactions. In *Advances in Physical Organic Chemistry*; Gold, V., Bethell, D., Eds.; Academic Press, 1980; Vol. 17, pp 183–278.
(52) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. Intrinsic Disorder and Protein Function. *Biochemistry (Mosc.)* **2002**, *41* (21), 6573–6582.
(53) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (3), 197–208.
(54) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Mutual Synergistic Folding in Recruitment of CBP/p300 by p160 Nuclear Receptor Coactivators. *Nature* **2002**, *415* (6871), 549–553.
(55) Ishida, T.; Kinoshita, K. PrDOS: Prediction of Disordered Protein Regions from Amino Acid Sequence. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W460–W464.
(56) Ramezanghorbani, F.; Dalgicdir, C.; Sayar, M. A Multi-State Coarse Grained Modeling Approach for an Intrinsically Disordered Peptide. *J. Chem. Phys.* **2017**, *147* (9), 094103.

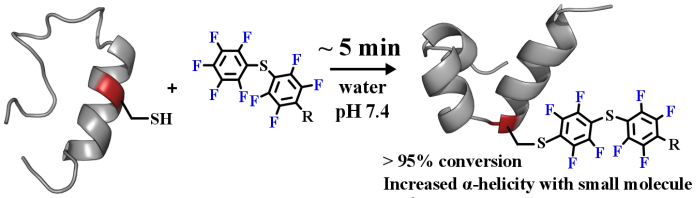
(57) Monticelli, L.; Sorin, E. J.; Tieleman, D. P.; Pande, V. S.; Colombo, G. Molecular Simulation of Multistate Peptide Dynamics: A Comparison between Microsecond Timescale Sampling and Multiple Shorter Trajectories. *J. Comput. Chem.* **2008**, *29* (11), 1740–1752.
(58) Löffler, P.; Schmitz, S.; Hupfeld, E.; Sterner, R.; Merkl, R. Rosetta:MSF: A Modular Framework for Multi-State Computational Protein Design. *PLOS Comput. Biol.* **2017**, *13* (6), e1005600.
(59) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; et al. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453* (7192), 190–195.
(60) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; et al. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319* (5868), 1387–1391.

TOC Graphic

Starting reactive peptide



Experimental and computational improvement



> 95% conversion
Increased α-helicity with small molecule
~10^3 rate enchancement