Determinants of protein-peptide interaction specificity in the Bcl-2 and TRAF families

Glenna Wink Foight

B.S. Biochemistry
North Carolina State University, 2009

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2015

Signature of Author: Signature redacted
Department of Biology
September 2015

Certified by: Signature redacted
Amy Keating
Professor of Biology
Thesis Supervisor

Accepted by: Signature redacted
Michael Hemann
Associate Professor of Biology
Co-Chair, Biology Graduate Committee

Determinants of protein-peptide interaction specificity in the Bcl-2 and TRAF families

by
Glenna Wink Foight

Submitted to the Department of Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology
September 2015

Abstract

Protein-peptide interactions have important roles in the majority of cellular processes. There are many families of peptide recognition domains in which homologous members display differential binding preferences for peptide sequence features. Peptide binding specificity is critical for the functional roles played by each family member, which can be overlapping or distinct. The two peptide recognition domain families discussed in this work, Bcl-2 and TRAF proteins, have roles in cellular processes including apoptosis, inflammation, and immunity. Aberrant function of these proteins has been linked to a variety of diseases. There is great interest in understanding the mechanistic basis of protein-peptide binding specificity in these families and others. An improved understanding will enable models of binding preferences for interactome prediction and design of specific peptide reagents for the inhibition and study of protein-peptide interactions.

The anti-apoptotic Bcl-2 family members bind α-helical Bcl-2-homology 3 (BH3) motifs in pro-apoptotic Bcl-2 family members to prevent apoptosis. Kaposi Sarcoma herpesvirus and Epstein Barr herpesvirus express viral homologs of the anti-apoptotic Bcl-2 proteins, KSBcl-2 and BHRF1, respectively, during viral replication to prevent host cell death. Because human Bcl-2 proteins are important in preventing apoptosis in cancers, there is interest in targeting the viral homologs, as they may also have a role in herpesvirus-associated malignancies. I designed and screened libraries of BH3 peptide variants for binding specificity to KSBcl-2 and BHRF1. From library screening and additional rational mutagenesis, I developed peptides that showed specific binding to KSBcl-2, BHRF1, or the human homolog Mcl-1, and displayed large margins of specificity over the other human Bcl-2 homologs.

TRAF proteins bind sequences in the unstructured regions of cell surface receptors and other adapter proteins in order to mediate downstream signaling events. TRAF-peptide binding preferences are relatively uncharacterized. I adapted a bacterial surface display system for screening peptides for TRAF binding. Using this system, I explored the binding preferences of TRAFs 2, 3, and 5 to single and double mutant libraries of two peptide interaction partners from CD40 and TANK. Comparison of the enriched peptide sequences reveals a surprising degree of difference between these three close TRAF homologs, yielding hypotheses relevant to TRAF function and inhibition.

Thesis Supervisor: Amy Keating
Title: Professor of Biology

Acknowledgements

I would like to thank my supervisor, Amy Keating for all of her support and valuable training over the years. I have learned a lot from observing her highly rigorous approach to science. I have also benefited from her attention to the development of her students on both scientific and professional levels.

I would also like to thank my committee members Bob Sauer and Thomas Schwartz. Yearly committee meetings with them have provided valuable re-focusing of my projects. I would especially like to thank them for providing me with many letters for applications over the years. Thanks to Joshua Kritzer for serving on my defense committee.

Thanks to Carla Mattos, my undergraduate mentor, who taught me to be a good structural biochemist and gave me many opportunities that got my research career off to a great start.

Thanks to the many members of the Keating lab past and present: Sanjib Dutta for valuable training in library display technology, and lots of great, unsolicited advice over the years. Joe DeBartolo, Raheleh Rezaei-Araghi and Justin Jenson for valuable discussions on Bcl-2s. Jenifer Kaplan for being a great lab mate and for reading many paper and thesis chapter drafts. Christos Kougentakis for making the lab a fun place to work and for being an exceptional technician who was always volunteering to do work on the weekends.

I would also like to thank members of the MIT Koch flow cytometry core for assistance with sorting. Thanks to the other core facilities for much assistance: the MIT Biomicro Center, MIT Koch Biopolymers, and the Biophysical Instrumentation Facility. Special thanks to Bob Grant in the crystallography core for help and advice on screening crystals.

Thanks to the Biograd class of 2009. The coffee breaks, game nights, and Building-68-lunches have been invaluable in getting me through.

Finally, thanks to my family: My parents for supporting me in following my interests, and helping me keep things in perspective. My husband, Dillon for everything.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Protein-protein interactions are critical for essentially all cellular events. With ~30,000 proteins expressed by the human genome, a large number of pairings are possible. Many protein-protein interactions (PPI) occur between the folded domains of proteins, analogous to two puzzle pieces interlocking. Such domain-domain interactions have complex interfaces between tertiary structures, and we lack structural characterization of many of the known examples due to the relative difficulty of crystallizing protein complexes. An estimated 15-40% of PPIs occur between a folded domain of one protein and a linear polypeptide fragment of another protein (Petsalaki and Russell, 2008). Domain-peptide interfaces are regarded as simpler than domain-domain interfaces as the peptide generally lacks secondary structure when unbound, and in many cases, the domain surface does not change substantially upon peptide binding (London et al., 2010). Protein-peptide interactions are especially important for connecting proteins in signaling pathways. They also mediate the formation of complexes for a variety of processes such as the regulation of apoptosis or cell motility. Their essential biological roles and simple interfaces relative to domain-domain interactions make protein-peptide interactions an important frontier for the advancement of our understanding of the protein interactome.

Many examples of peptide recognition domains (PRDs) have been identified, and the peptide binding preferences of several families of homologous PRDs have been extensively characterized *in vitro* (Liu et al., 2012). The peptide binding preferences of a PRD family are often represented as a motif. A sampling of PRDs and their known motifs are presented in Table 1.1. Motifs contain the residues that are most invariant in the known binders of a PRD. They can be inferred from multiple sequence alignments of cellular interaction partners, or established by experimental mutagenesis. Motifs are a simplistic representation of PRD binding preferences. The functional specificity of individual PRD family members arises in part from recognition of

different motifs, or tolerance of significant variation on motifs. Context residues, here defined as the residues not restricted in the motif ("x") or the residues terminal to the motif, also influence binding specificity.

Table 1.1. Protein recognition domains discussed and a subset of their binding motifs

Protein recognition domain	Motifs recognized[a]
PDZ	(ST)x(ACVILF)>[b], (VILFY)x(ACVILF)>, (DE)x(ACVILF)>
SH3	(RKY)xxPxxP, PxxPx(RK), xxx(PV)xxP, KPxx(QK)xxx, PxxDY
SH2	pYxN, pY(QDEVAIL)(DENPYHI)(IPVGAHS), pYxxQ, pY(VLTFIC)xx
PTB	x(ILVMFY)xNxx(FYpY)x, x(ILVM)LGxxPx
WW	xx(p(ST))Px, PPxY, PPLP
GYF	(QHR)xP(PL)PP(GS)H(RH)
Bcl-2	ΦxxxΦxxΦ(GSA)DxΦ
TRAF	(PSAT)x(QE)E, PxQxD, xxPxExx(FYWHDE)
EVH1	(FYWL)PxPP, (FYWL)PP(ALIVTFY)P, PPxF, (FY)x(FW)5x(LMVIF)PxP(DE)

[a] Motifs are from the Eukatryotic Linear Motif database, except for the Bcl-2 motif.
[b] ">" denotes the position of the C-terminus.
[c] "Φ" denotes a hydrophobic residue.

This thesis addresses the topic of protein-peptide binding specificity, and the introduction covers the biological context of binding specificity, the binding specificity mechanisms of well-studied, and less-studied PRD families, as well as applicable methods and applications. It is important to first acknowledge that in the cell, the biophysical determinants of binding are only one of several factors that dictate whether two proteins will interact.

Interaction determinants in the cell

Post-translational modifications

In vivo, protein-peptide interactions can be controlled by a variety of factors beyond their binding specificity such as post-translational modifications (PTMs), localization,

oligomerization, and coexpression. PTMs are required in the recognition motifs of several PRDs, including phosphorylation of serine, threonine, or tyrosine (bound by src-homolog domain 2 (SH2), phosphotyrosine-binding domain (PTB), forkhead-associated domain (FHA), 14-3-3 proteins, etc.), or acetylation or methylation of lysine or arginine (bromodomains and chromodomains) (Pawson and Nash, 2003). Other PRD families have subsets of members with preferences for binding sites with PTMs—such as class IV WW domains, which bind phosphorylated serine or threonine—and other subsets whose binding is inhibited by PTMs, such as class I WW domains, which bind a motif containing a tyrosine. Analysis of the co-localization of phosphosites and the motifs of SH3 (src-homology 3), SH2, WW, and PDZ (PSD-95/Discs-large/ZO-1) domains in the human proteome found a statistically significant association (Akiva et al., 2012). Thus, PTMs act as switches to turn interactions with peptide motifs on or off, depending on the specificity of the PRD. There are even cases of "double switches" in which motifs for two different PRDs overlap, such as that of an SH2 domain and an SH3 domain; in this case, the SH2 domain binds to the site when it is phosphorylated, and the SH3 domain binds when the site is not phosphorylated (Akiva et al., 2012).

An example of a protein-peptide interaction that is heavily regulated by both PTMs and subcellular localization is the interaction between the BH3 (Bcl-2 homology 3) motif of the pro-apoptotic protein Bad and its anti-apoptotic PRDs, Bcl-x_L and Bcl-2. Bad promotes apoptosis by binding and inhibiting Bcl-x_L and Bcl-2 when it is localized to the outer mitochondrial membrane (Yang et al., 1995). Bad contains three serines that are phosphorylated by survival kinases, such as Akt (Danial, 2008). Two of these phosphorylated serines form interaction sites for 14-3-3 proteins, and the combination of phosphorylation and binding of 14-3-3 proteins results in the relocation of Bad to the cytoplasm. The pro-apoptotic activity of Bad is further

inhibited by phosphorylation of a serine in the BH3 binding motif, which directly blocks Bcl-x_L and Bcl-2 binding (Danial, 2008). At another level of regulation, with an opposing effect, Bad is also a substrate for PRMT1, a protein arginine methyltransferase. Methylation of Bad arginines that form the recognition site for Akt (outside of the BH3 motif) block Bad phosphorylation, promoting its apoptotic activity (Sakamaki et al., 2011). Bad is just one example of how PTMs and localization can influence protein-peptide interactions. Further mapping of PTMs to motif sites in the proteome will continue to unveil complex levels of regulation for the formation of protein-peptide complexes.

Gene expression

Protein-protein interaction networks are also regulated at the levels of gene expression and protein abundance. Most models of protein interaction networks lack information about the dynamics of protein abundance or gene expression. Several studies have integrated temporal or conditional gene expression studies (e.g. different cell cycle stages or growth conditions) with protein interaction data from yeast and *E. coli* (de Lichtenberg et al., 2005; Hegde et al., 2008; Komurov and White, 2007; Tang et al., 2011). These analyses have created limited dynamic protein interaction networks, but the networks are enriched for the type of stable domain-domain interactions that are more easily identified in large yeast two-hybrid or affinity purification-mass spectrometry datasets. Transient interactions, such as many weak domain-peptide interactions, are often missed by yeast two-hybrid experiments (Vinayagam et al., 2009). Protein abundance and signaling events also play a role in transient domain-peptide interactions, and these are generally not directly reflected in gene expression studies. For example, weak interactions such as those between tumor necrosis factor receptor-associated factors (TRAFs) and peptides in TNF receptors rely on avidity created by the oligomerization of TNF receptors in response to

extracellular signals (Pullen et al., 1999b). Therefore, significant advances are needed before gene expression, protein abundance, subcellular localization, and protein interaction data can be combined to create a comprehensive view of transient protein interaction networks. A clear accounting of possible protein interactions as determined *in vitro* is a necessary first step. As it would be an impossible task to directly measure all possible protein interactions, it is important to define the rules governing interactions. Determining the peptide motifs that enable interaction with each PRD family will enable modeling of protein interaction networks in the future.

Methods for determining peptide interaction motifs

Peptide interaction motifs can be identified through a variety of experimental and computational methods. In this section, a brief survey of classical experimental methods and a more in-depth description of a new technique, deep mutational scanning, will be followed by a discussion of computational methods for motif and specificity determination.

Experimental methods

Experimentally, peptide interaction sites are often initially found by mutagenesis, with alanine scanning or hydrophile scanning being used to identify the residues most important for binding (Boersma et al., 2008). More in-depth examination of PRD binding preferences requires analysis of binding to many more peptides than are accessible by standard solution binding assays. Techniques such as SPOT arrays or protein microarrays can be used to semi-quantitatively analyze binding to a modest number of peptides (hundreds to thousands) (Liu et al., 2012). Larger libraries of peptides are accessible by high throughput screening strategies such as phage display, cell surface display, and ribosome or mRNA display (Levin and Weiss, 2006). Phage, ribosome and mRNA display offer the capacity to screen very large libraries (routinely 10^{10} for phage and $>10^{12}$ for ribosome and mRNA). Cell surface display techniques

have library sizes limited by the efficiency of getting the library DNA into cells, but offer the advantage of real-time, affinity-based resolution of binding partners when combined with fluorescence activated cell sorting (FACS). Yeast surface display libraries of 10^7 can be easily achieved, with larger libraries made possible by combining hundreds of transformations. With bacterial surface display in *E. coli*, libraries of 10^{10} are easily achievable due to the higher transformation efficiency of gram negative bacteria relative to yeast (Löfblom, 2011). These library-screening techniques have been successful in screening large random peptide libraries or natural sequences and variants thereof. When combined with deep sequencing of the enriched pools, these techniques can greatly expand the binding sequence space beyond that of the known natural partners. These screens excel at providing peptide reagents of high affinity or specificity for a PRD of interest. Screens can also be performed in the other direction – to find a variety of domain sequences that bind a given peptide (Chen et al., 2013; Ernst et al., 2010; Gold et al., 2013).

One disadvantage of these large library screens is that they are generally used to supply sequences of only the best leads and lack information on relative affinity or specificity of sequences of a broader spectrum. A recent method by Reich et al. utilizes the affinity discrimination offered by FACS to sort a yeast surface display library of peptides into pools of different affinity (Reich et al., 2014). The relative affinity for each clone is computationally extracted from its distribution across the pools as determined by deep sequencing. The combination of cell surface display methods, FACS, and deep sequencing has great potential to provide datasets containing sequence and affinity information from which to elucidate peptide-binding determinants.

The power to systematically explore binding determinants has been greatly improved by the advent of a technique known as deep mutational scanning (Fowler et al., 2010). In its most general form, deep mutational scanning involves screening a comprehensive library of mutants of a protein of interest by an assay designed to measure some aspect of protein function. Subsequent deep sequencing of the naïve and enriched library pools provides relative frequencies of sequences that can be used as a measure of the protein function assayed (Figure 1.1). This general framework has been applied to study sequence determinants of protein stability, enzyme function, protein-ligand interactions, and more general conglomerates of overall protein function (Araya et al., 2012; Fujino et al., 2012; Melamed et al., 2013; 2015; Starita et al., 2015; Tinberg et al., 2013; Traxlmayr et al., 2012; Whitehead et al., 2012). Detailed information about the relative contributions to function, or fitness, of all 20 amino acids at each protein position has the potential to provide rich datasets for applications such as protein structure prediction, protein evolution and engineering, and the study of human disease genetics (Fowler and Fields, 2014).

Deep mutational scanning and related approaches have been applied to both protein-protein (antibody:antigen, RNA recognition motif:eIF4G1, computationally designed influenza hemagglutinin binders) and protein-peptide interactions (WW and PDZ) (Forsyth et al., 2013; Fowler et al., 2010; Fujino et al., 2012; McLaughlin et al., 2012; Melamed et al., 2013; 2015; Whitehead et al., 2012). Any binding assay that can screen modestly sized libraries, such as phage display, cell surface display, ribosome or mRNA display, or two-hybrid assays can be utilized to enrich the low complexity libraries necessary for this technique. Analysis of the correlation between binding affinity and enrichment in screening for binding is lacking, but small scale comparisons indicate a good correlation (Fowler et al., 2010; McLaughlin et al., 2012; Pál et al., 2006). The most comprehensive analysis compared dissociation constants measured by

fluorescence polarization to enrichment values from a bacterial two-hybrid screen for 86 PDZ domain mutants binding a peptide and found a good linear correlation over a range of K_D values of 0.1-200 μM (McLaughlin et al., 2012). However, when mutating an entire protein domain, other parameters such as folding and stability will also influence the data. Researchers have used both experimental (e.g. initial selection for folded variants by binding of an antibody) and computational strategies (e.g. integration of evolutionary conservation information) to identify mutations that affect the binding interface (Melamed et al., 2015; Pál et al., 2006).

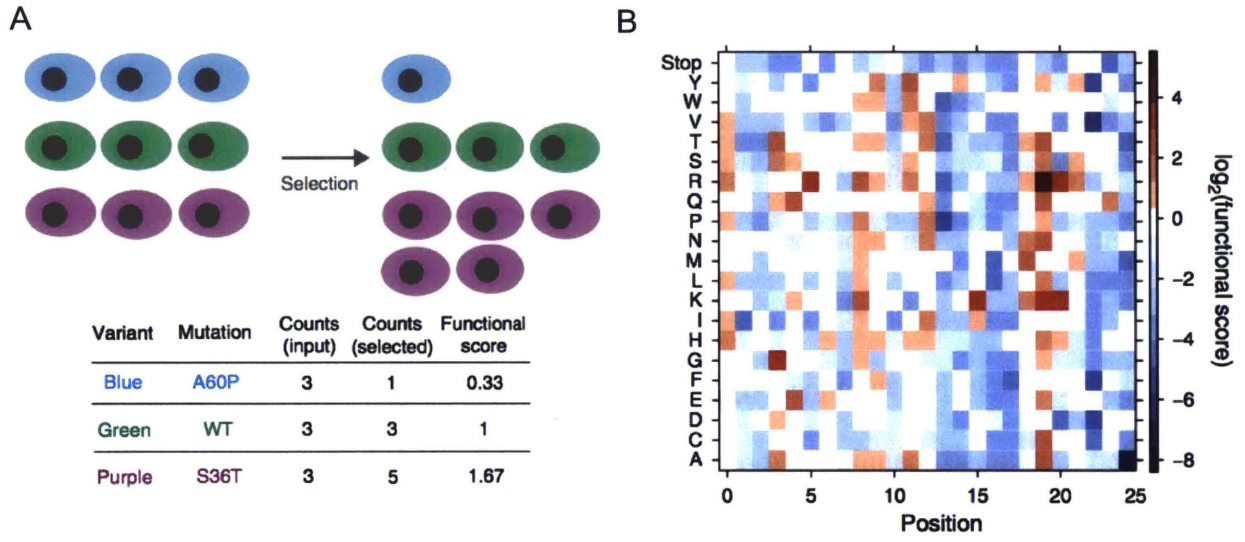


Figure 1.1. Deep mutational scanning provides a metric of function for mutations to all 20 amino acids at each protein position. A) Next generation sequencing is used to quantify the number of copies of each variant in the naïve library and the pool resulting from selection for some function of interest. B) The frequencies are converted to a functional score based on the frequencies of each variant in the naïve and selected pools, and this is used to create a heat map of protein mutational tolerance. Figures are from Fowler et al., 2014.

The ability to leverage enriched mutations to engineer tighter binding partners also demonstrates that enrichment can be a good proxy for affinity (Fujino et al., 2012; Tinberg et al., 2013; Whitehead et al., 2012). One powerful application of deep mutational scanning of protein-protein interfaces is to use the datasets to inform design of combinatorial libraries of enriched mutations. These libraries can then be screened for variants that successfully combine the affinity

enhancing effects of several mutations to greatly increase binding affinity. This approach has been used to improve the affinity of computationally designed binders of influenza hemagglutinin and antibodies that bind tumor necrosis factor-α receptor (Fujino et al., 2012; Whitehead et al., 2012).

To date, deep mutational scanning studies have mutated the protein side of protein-peptide interfaces. These data have been utilized to identify sequence determinants of binding specificity for peptide ligands. Deep mutational scanning of a PDZ domain was performed twice for binding to a cognate peptide ligand and a single point mutant of that ligand, and these datasets were compared to find PDZ domain positions that governed peptide binding preferences (McLaughlin et al., 2012). When deeper mutation is performed, such that double mutants are present, information on epistasis or covariation between positions can be inferred by comparing single and double mutant enrichment (Araya et al., 2012; McLaughlin et al., 2012; Melamed et al., 2015; Whitehead et al., 2012). This analysis has been incomplete, as it is difficult to get complete coverage of double mutants in folded domains by the current methods used for library assembly. Double, and even higher-order mutations should be easier to cover in short contiguous peptide segments. Currently, the degree of covariation between positions in different linear peptide motifs is not known. This question needs to be answered in order to create fully parameterized computational models for prediction and design of peptide binding partners.

Computational methods

Computational models of protein-peptide interaction specificity can build on experimental datasets and utilize features from sequence and structures of complexes. Peptide interaction motifs are often first identified on the sequence level in multiple sequence alignments of proteins of related function. Models of linear motifs such as position specific scoring matrices

(PSSMs), regular expressions, or hidden Markov models (HMMs) can be built based on lists of known examples and used to search for more proteins with a matching sequence (Dinkel et al., 2014; McLaughlin et al., 2011). Short linear motifs are difficult to establish without prior knowledge of binding sites, but several methods exist based on searching for motif over-representation in non-homologous sequences that are functionally related by some metric, such as binding to a common partner in yeast two-hybrid datasets (Neduva and Russell, 2006; Neduva et al., 2005; Palopoli et al., 2015). In recent years, large library or array experiments have generated long lists of natural and non-natural binders of PDZ, SH3, and Bcl-2 domains, among others (DeBartolo et al., 2012; Ernst et al., 2010; Tonikian et al., 2009; 2008; Xin et al., 2013). This increase in known binders has allowed the construction of powerful PSSMs and other models useful for scanning the proteome for more peptide binders of these domains (DeBartolo et al., 2014; Gfeller et al., 2011; Kim et al., 2012). Experimental generation of these datasets requires significant work, and the resulting models can only reflect the sequence space upon which they were built, limiting their utility for finding motifs in novel specificity classes. Though useful and fast, sequence-based methods for motif discovery and matching ignore an important source of information on interaction specificity—the physical interactions encoded in the complex structures.

Structure-based binding specificity prediction methods exist at varying levels of physical detail. All methods require a high resolution structure of an example protein-peptide complex and a set of experimentally validated peptide sequences upon which to build and test the model (London et al., 2013). I will discuss three examples of computational approaches spanning levels of physical detail and computational intensity including full physics-based models, a combined statistical-physical approach, and finally, a statistical approach based on distances in protein

complexes. As protein interaction modeling is an important goal in computational biology, these examples are only a sampling of the many published approaches.

At the most computationally intensive and physically detailed end of the spectrum lie models utilizing molecular dynamics simulations to model lists of complexes and molecular mechanics-generalized Born surface area (MM-GBSA) to compute binding energies. MM-GBSA is used to calculate molecular interaction energy components for many complexes, which can be used to train a support vector machine to create a general model of the binding preferences of a PRD. Wang and colleagues used this approach to create models of SH3, PDZ, PKA-AKAP, and chromodomain binding preferences, which they successfully used to identify peptide interactors from proteomes (Hou et al., 2012; 2011; 2009; Li et al., 2011; 2013). Because these models involve direct calculation of the energy contributions of each peptide and receptor residue, they can be used to identify the structural mechanisms behind different binding specificities, such as that of class I *versus* class II SH3 binders (Hou et al., 2012).

Rosetta FlexPepBind is a second structural method for modeling of peptide binding specificity that scores peptide sequences using the combined statistical and physical Rosetta energy function as implemented in the protocol Rosetta FlexPepDock (Raveh et al., 2010). London et al. used this method to model the binding specificity of farnesyl transferase to its peptide substrates, and Bcl-2 to BH3 peptides (London et al., 2012; 2011). They found good predictive performance by simple threading of peptide sequences on a complex structure followed by minimization and energy calculation. Treatment of peptide side-chain and backbone flexibility via minimization was necessary for good performance on a test set of known binding partners, and more extensive sampling of flexibility using Rosetta FlexPepDock improved performance, at the cost of increased computing time. The Rosetta FlexPepBind protocol

requires calibration on a set of known binding partners, and its performance is often improved with further parameterization based on known structural constraints of the protein-peptide system (London et al., 2011; 2012).

At a third and lowest level of complexity and computational intensity, lies STATIUM, a statistical model that relies only on a protein-peptide complex structure or homology model and a reference set of monomeric structures from the PDB (DeBartolo et al., 2012; 2014). STATIUM calculates the distances from the Cα and Cβ atoms of each peptide residue to all receptor side-chain atoms within a given distance cutoff. It then tabulates the frequency with which all 20 amino acids are found in similar residue-residue pair geometries in the reference database and calculates a score based on that frequency normalized by the frequency of that amino acid in the reference protein set (Figure 1.2). Because the residue-residue distances in the reference set are pre-tabulated, scoring a sequence on a complex structure is essentially a "look-up procedure", allowing scoring of $>10^6$ sequences per second. STATIUM has been used to scan the human proteome for novel BH3 peptide sequences that bind the anti-apoptotic Bcl-2 homologs, and shows good agreement with experiment-based models of Bcl-2:BH3 binding preferences (DeBartolo et al., 2012; 2014). STATIUM is, in theory, applicable to any protein-protein interface, though it has yet to be tested extensively on other protein-peptide families.

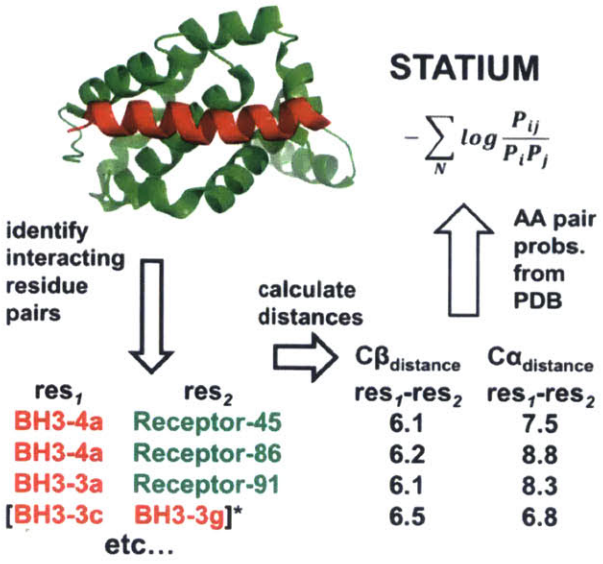


Figure 1.2. How a STATIUM model is built for a protein:peptide complex structure. Distances from each peptide Cα and Cβ to each receptor residue Cα and Cβ are calculated. The probability of finding each residue pair and distance are found in the reference PDB set and normalized against the frequency of each amino acid in the reference. The STATIUM score for a peptide is the sum of this value for each peptide residue. The more recent STATIUMsidechain program uses distances to all receptor side-chain heavy atoms. Figure is from DeBartolo et al., 2012.

Binding specificity mechanisms

A hierarchy of binding specificity mechanisms within PRD families

Detailed characterization of peptide binding for a few PRD families has uncovered mechanisms that differentiate the binding preferences of individual members within a PRD family. These differences can take the form of recognition of different motifs, variations on the same motif, or even recognition of completely different protein structures (e.g. binding the tertiary structure of another domain rather than a linear peptide). First, it will be helpful to understand the energetic basis of a motif. Stein and Aloy analyzed the complex structures of 30 different PRD families and evaluated the energetic contributions of peptide residues using the computational algorithm FoldX (Stein and Aloy, 2008). They found that residues in the motifs cited in the Eukaryotic Linear Motif (ELM) database contributed on average 80% of the binding

energy, whereas the more variable context residues only contributed 20%. Therefore, the sequence features that make up a peptide motif can be regarded as the hotspot residues, i.e. those that contribute the most binding energy. A more general finding that sequence conservation in bound peptides correlates with per-residue binding energy contribution also held true in that residues that were fixed as one amino acid in the motif versus restricted to a set of amino acids had higher energetic contributions on average (Stein and Aloy, 2008). Given this concept of a motif, I will discuss how binding specificity is encoded on different levels, starting from the most diverged and ending with examples of fine-scale differences in binding specificity between closely related members of the same PRD family.

The use of multiple motifs for specificity in large PRD families

Nature utilizes similar protein forms for diverse purposes. One example of this concept in the realm of protein-ligand binding is the pleckstrin homology fold. This same fold is recycled in domains that bind a wide variety of ligands including phospholipids (pleckstrin homology (PH) domains), phosphopeptides (PTB domains), and polyproline peptides (Ena/VASP homology 1 domains (EVH1)) (Pawson and Nash, 2003). Conversely, similar ligands (peptides containing polyproline type II helices) can bind similar interfaces (aromatic-rich) in domains with very different folds (SH3, WW, EVH1, GYF, and profilin) (Li, 2005). This illustrates that there are certain structural solutions to recognition that are repeated in the protein interactome due to both divergent and convergent evolution. However, a high degree of specificity can be encoded within a given structural form by use of different residues.

Within a PRD family, members can utilize highly diverged interface features to bind peptides or proteins of different structures. SH3 domains have multiplied in eukaryotes from 34 different SH3 domains in yeast to 958 in humans, according to the most recent estimates in the

SMART database (Letunic et al., 2015). Within such a large family, a diverse array of protein-binding modes has been realized over the course of evolution. Though the majority of SH3 domains bind peptides containing a PxxP motif, decades of work on the SH3 domains has uncovered examples that bind peptides with very different motifs (e.g. RxxK), peptides binding in 3_{10}- or α-helical conformations, and even interactions with tertiary structures of other domains (Figure 1.3A) (Li, 2005). The binding groove is defined by two loops of the SH3, the RT loop and the n-Src loop. Variations on the length and sequence content of these loops can change the shape and charge properties of the groove to favor peptides matching the RxxK versus the PxxP motif (Kaneko et al., 2011).

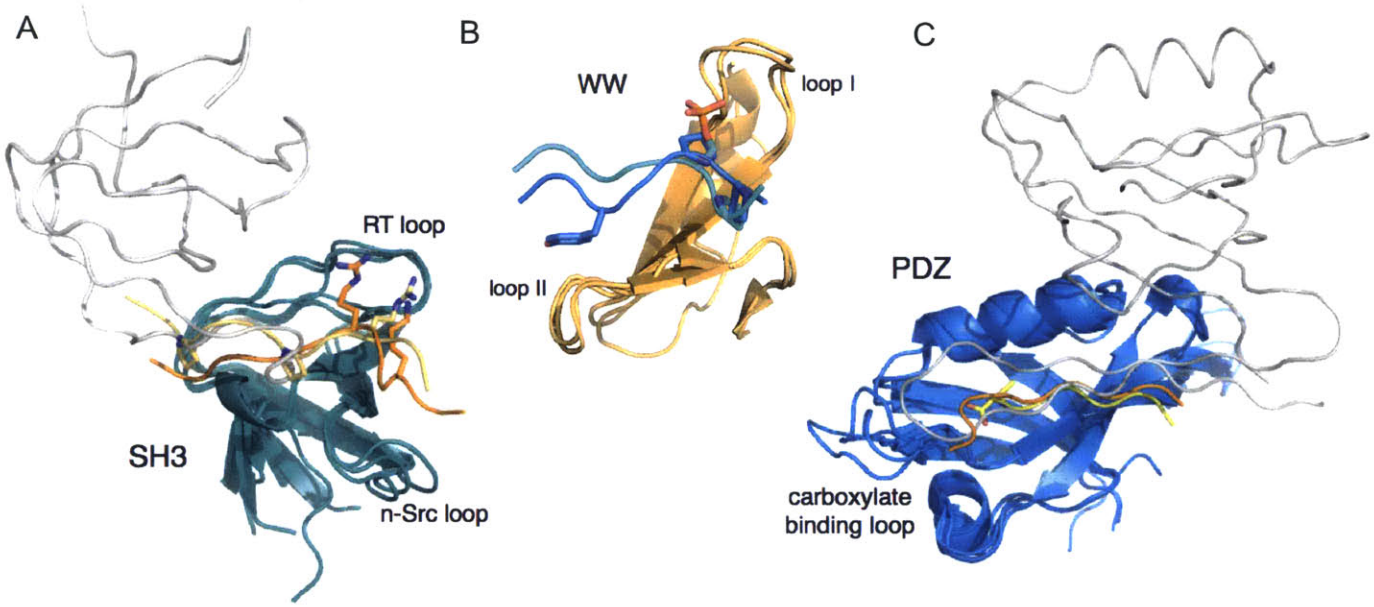


Figure 1.3. Three large PRD families that use different motifs for specificity. A) SH3 domains (teal) bound to a variety of partners: another SH3 domain (gray, 1GCQ (Nishida et al., 2001)), a peptide with a PxxPxR motif (yellow, 1GBQ (Wittekind et al., 1997)), and a peptide with an RxxK motif (orange, 1H3H (Liu et al., 2003)). Motif residues are shown in sticks. B) WW domains (orange) bound to two peptides of different motifs: pSP (cyan, 1FH8 (Verdecia et al., 2000)), PPxY (blue, 1EGH (Huang et al., 2000)). C) PDZ domains (blue) bound to: a C-terminal peptide with C-terminal residue shown in sticks (yellow, 1RZX (Peterson et al., 2004)), an internal peptide (orange, 1X8S (Penkert et al., 2004)), and a β-finger from another PDZ domain (gray, 1QAV (Hillier, 1999)). Loops that play a critical role in specificity are labeled for each domain.

WW domains have many similarities to SH3 domains, and are also a very common domain with an estimated 247 examples in the human proteome (Letunic et al., 2015). WW domains are named after the two tryptophans that form the binding groove for an "xP" motif. Several studies of WW domain binding specificity have grouped the binding preferences into 4-5 classes represented by different ligand motifs (Hu et al., 2004; Ingham et al., 2005). Two very different motifs are PPxY and p(S/T)P (phosphorylated serine or threonine) (Figure 1.3B). Like SH3 domains, there is a common platform for binding the "xP motif", and different WW domains utilize variation in two loops to dictate specificity (Verdecia et al., 2000). The p(S/T)P motif, first identified in binders of the Pin1 WW domain, is recognized by an arginine and serine on loop I that make direct contacts with the phosphorylated residue. In WW domains of the PPxY specificity class, a hydrophobic surface formed by loop II recognizes the tyrosine.

Another well-studied PRD family, the PDZ domain, also recognizes diverse motifs. PDZ domains predominantly bind to the C-termini of peptides, recognizing the C-terminal carboxylate via the backbone amides of the domain's carboxylate binding loop (Lee and Zheng, 2010). However, like SH3 domains, PDZ domains can form domain-domain interactions with other PDZ domains and can also bind other peptide conformations, including internal peptide sequences (Figure 1.3C) (Chang et al., 2011; Hillier, 1999; Penkert et al., 2004). Among PDZ domains that bind C-terminal peptides, many different specificity classes have been identified that recognize distinct motifs. By selecting phage display libraries of random peptides for binding to 82 human and *C. elegans* PDZ domains, Tonikian et al. identified 16 specificity classes, into which the specificity profiles of 90% of the profiled domains could be grouped (Tonikian et al., 2008). These 16 different motifs, represented as position weight matrices (PWMs), exhibited distinct preferences at up to 7 peptide positions leading up to the C-terminus.

To analyze which PDZ domain residues governed specificity at each peptide position, Tonikian et al. determined the specificity profiles of 91 single mutants of the Erbin PDZ domain, varying 10 different PDZ positions. They found that the preferences at some peptide positions were primarily influenced by PDZ residues in direct contact with the peptide side chain, while other peptide positions were also influenced through non-direct contacts. This demonstrates that peptide binding specificity reflects a complex combination of direct residue-residue interactions and underlying structural features.

The use of motif context residues for specificity in smaller PRD families

PRD family members can bind peptides that match the same motif but still exhibit distinct preferences for their interaction partners. In the analysis of protein-peptide structures performed by Stein and Aloy, they looked at specificity by performing a peptide-swapping experiment between structures of PRD families with at least 10 different domain-peptide pairs (cyclins, hormone receptors, MATH, PDZ, and class I and II SH3 domains) (Stein and Aloy, 2008). Even when controlling for similar peptide topology, they found that "context" peptide positions were more likely to have poor FoldX energies than motif positions when assessed in a non-cognate complex. Thus, PRD family members that can bind peptides of the same motif utilize the context residues for specificity. Context residues that are tolerated or even provide modest contributions to binding affinity in their cognate complexes can act as strong deterrents to binding non-cognate domains. This idea of "anti-motifs" present in the context residues has been observed in the binding specificity determinants of many PRD families (Liu et al., 2012). The large PRD families (SH2, SH3, PDZ, and WW) exhibit such context-dependent specificity mechanisms between closely related members. However, my discussion will focus on smaller PRD families, including Bcl-2, EVH1, and TRAF domains. Smaller PRD families often have

less diversity in motif preferences, and they are good examples of how fine-scale differences in motif context can have substantial effects on binding specificity.

Bcl-2 proteins control apoptosis through interactions with helical BH3 motifs. Pro-apoptotic Bcl-2 proteins can take one of three forms (Llambi et al., 2011). The effectors Bax and Bak homo-oligomerize to form pores in the outer mitochondrial membrane, which release death factors into the cytoplasm. This process is triggered by activator BH3-only proteins, such as Bid and Bim. The third class of pro-apoptotic proteins are the sensitizer BH3-only proteins, such as Bad, Noxa, Hrk, and Bik, which de-repress the function of the other two pro-apoptotic forms by binding and inhibiting the anti-apoptotic Bcl-2 homologs. Anti-apoptotic Bcl-2 proteins—including Bcl-2, Bcl-x_L, Bcl-w, Bfl-1, Mcl-1, and Bcl-b in humans—bind the BH3 motif of the BH3-only proteins and Bax and Bak to prevent apoptosis. The interaction between anti-apoptotic Bcl-2 homologs and BH3 motifs has been extensively studied because the anti-apoptotic homologs are widely overexpressed in human cancers to prevent apoptosis, making them important drug targets.

The BH3 motif, like other protein-peptide motifs, is often found in unstructured regions in BH3-only proteins, many of which are intrinsically disordered proteins (Rautureau et al., 2010). BH3 motifs are also found folded as α-helices in the anti-apoptotic Bcl-2 and pro-apoptotic effector homologs, and in these cases they are exposed for binding upon structural rearrangements of the folded Bcl-2 domains (Lee et al., 2014). The BH3 motif binds to anti-apoptotic Bcl-2 homologs as an amphipathic α-helix of ~25 residues with a general motif definition of ΦxxxΦxxΦ(GSA)DxΦ where "Φ" is a hydrophobic residue. Because the large majority of contacts are mediated by residues within and immediately surrounding the motif, Bcl-2:BH3 interactions are usually studied with BH3 peptides of 20-30 residues. A nomenclature

of a heptad repeat (a-g) is often used to describe the peptide positions, with the buried hydrophobic positions falling at positions a and d (see Figure 1.4A). Perhaps because this is a longer peptide interface than many short linear peptide-PRD interfaces that have been studied, and therefore has a larger number of peptide residues contributing to binding, there is substantial heterogeneity allowed at the hydrophobic positions (Dutta et al., 2010). The specificity of the six human anti-apoptotic Bcl-2 homologs for binding different BH3 sequences arises both from variations in the motif residues and the context residues. BH3-containing proteins, which include pro-apoptotic and pro-autophagic members, are differentially expressed and regulated in response to apoptotic or autophagic stimuli. Thus, the BH3-binding preferences of the anti-apoptotic Bcl-2 homologs dictate their relative importance in responses to these stimuli (Chen et al., 2005). I will first summarize the work that has been done to study the binding preferences of anti-apoptotic Bcl-2 homologs and then discuss examples of specificity mechanisms relying on variations of motif and context residues.

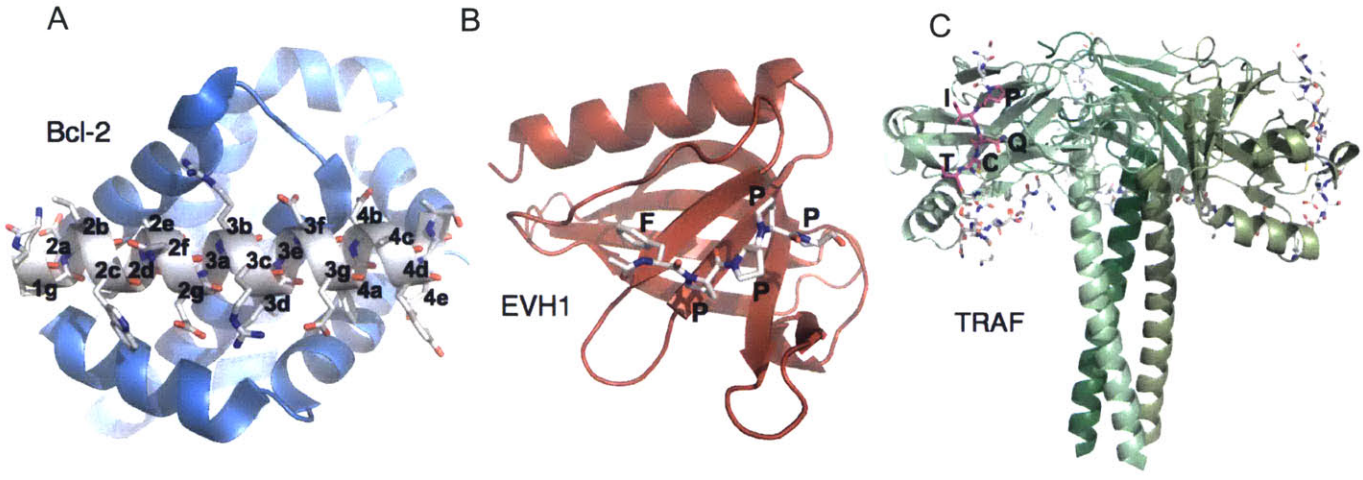


Figure 1.4. Three peptide recognition domain families utilizing motif context for specificity. A) A Bcl-2 homolog Mcl-1 (blue) bound to the Bim BH3 peptide (gray) is shown with BH3 positions labeled according to heptad register (Fire et al., 2010). B) The EVH1 domain of Mena is shown bound to one FPPPP motif of ActA (Prehoda et al., 1999). C) Trimerized TRAF3 is shown with monomers in shades of green. An extended peptide corresponding to the TANK TIM is in gray, with the 5-residue core site highlighted in magenta in one monomer (Li et al., 2002).

A number of high throughput approaches have been used to characterize the binding preferences of anti-apoptotic Bcl-2 homologs. All Bcl-2 homologs bind the BH3-only protein Bim with high affinity, and the Bim BH3 peptide has been used as a starting point to study mutational binding preferences of each Bcl-2 homolog. Dutta et al. synthesized peptides with 18 amino-acid variants (excluding cysteine and methionine) at 10 Bim BH3 positions on SPOT arrays, and measured binding of Mcl-1 and Bcl-x_L to the arrays (Dutta et al., 2010). Subsequent studies performed the same experiment with the other homologs Bcl-w, Bfl-1, and Bcl-2 (DeBartolo et al., 2012; London et al., 2012). These studies showed distinct preferences at subsets of peptide positions for each Bcl-2 homolog, including both motif and context positions.

Larger combinatorial libraries of random peptides or variants of Bim BH3 have been screened for binding to Mcl-1, Bcl-x_L, and Bfl-1. Lee et al. screened a random phage display library of 16-mer peptides for binding to Mcl-1 and largely recovered peptides that conform to the previously described BH3 motif (Lee et al., 2009a). Many of the recovered sequences showed specificity for binding Mcl-1 over the other human homologs Bcl-x_L, Bcl-2, and Bcl-w, despite the fact that the screen only selected for Mcl-1 binding and not selectivity. Dutta et al. screened a library of Bim BH3 variants by yeast surface display for peptides showing specificity to Mcl-1 or Bcl-x_L or binding to both homologs (Dutta et al., 2010). Comparison of sequence logos from the three screening paths revealed specificity determinants for Mcl-1 and Bcl-x_L. More recent yeast display screens by Dutta et al. have found further specificity determinants for Bcl-x_L and Bfl-1 (Dutta et al., 2013; 2014).

Bcl-2 homologs exhibit specificity differences at both motif and context positions in binding to BH3 peptides. One example of specificity from a motif residue is the preference for tyrosine or phenylalanine at the fourth hydrophobic position, 4a, exhibited by Bcl-x_L. Mutation

of 4a to smaller residues including valine or asparagine is tolerated by Mcl-1, but drastically decreases binding to Bcl-x_L (Dutta et al., 2010). This preference is encoded by the shape of the 4a binding pocket, which is enclosed in Bcl-x_L, but open to the solvent in Mcl-1 (Fire et al., 2010; Lee et al., 2009b). As another example, Bcl-x_L and Bcl-2 have high sequence similarity and both bind the BH3-only protein Bad (Certo et al., 2006). The mutation of tyrosine at 4e to lysine in a designed BH3 peptide does not decrease Bcl-x_L affinity, but reduces binding to Bcl-2 by 7-fold (Dutta et al., 2014). This position is outside the core BH3 motif, near the C-terminus of the peptide. In a structure of Bcl-2 bound to a Bax BH3 peptide, the methionine at 4e is directed into the groove and largely shielded from the solvent, while in a structure of Bcl-x_L bound to Bim, tyrosine at 4e is directed out from the groove with access to solvent (Ku et al., 2011; Lee et al., 2009b). The abundance of structural data for Bcl-2 homologs and the extensive mutational analysis of BH3 peptides has uncovered the structural basis of many other specificity mechanisms employed by the anti-apoptotic Bcl-2 members.

EVH1 (Ena/VASP homology 1) domains are present in four different families of proteins: Homer, WASP, SPRED, and Ena/VASP (Peterson and Volkman 2009). These proteins bind to polyproline motifs to regulate actin cytoskeleton dynamics. The three human Ena/VASP proteins, Mena, Evl, and VASP, bind to actin regulatory factors like zyxin and vinculin with motifs of FPxΦP, where Φ is a hydrophobic residue (Figure 1.4B). The pathogen *Listeria monocytogenes* expresses the protein ActA, which contains four FPPPP sites. Ball et al. mapped the chemical shift perturbations caused by a longer ActA peptide, SFEFPPPPTEDEL on the surface of VASP (Ball et al., 2000). They found that the C-terminal leucine contacted a hydrophobic patch outside the core polyproline-binding site. This exosite interaction increases the affinity of the ActA peptide for VASP. Similar hydrophobic positions are not observed near

the FPxΦP binding sites on zyxin, which do not bind as tightly as the ActA peptide. Thus, the pathogen uses a higher-affinity interaction to co-opt the host actin cytoskeleton for movement around the cell. It is not know whether any native Ena/VASP EVH1 binders form interactions outside of the core binding site, and Evl, Mena, and VASP have identical residues forming the binding site for the FPxΦP sequence. However, differences between the three homologs can be seen in the loops surrounding the identified exosite, which lies towards the C-terminus of the peptide (Ball et al., 2000; Fedorov et al., 1999; Prehoda et al., 1999). Notably, Mena forms a different type of interaction with the folded LIM3 domain of Tes, and this interface overlaps with the same exosite region bound on VASP by ActA (Boëda et al., 2007). Tes does not bind to Evl or VASP, indicating that differences between the homologs in this region have functional importance. It remains to be seen whether Evl, Mena, and VASP have differential binding preferences for their native binding partners containing FPxΦP sites, or whether variations in their functions should be attributed to differences in their other domains or aspects of their regulation (Gertler and Condeelis, 2011).

TRAF (tumor necrosis factor receptor-associated factor) proteins contain a peptide-binding interface on their MATH domains (meprin and traf homology). The MATH domain binds peptides from TNF receptors, as well as Toll-like receptors, other signaling receptors, and a variety of cytoplasmic signaling proteins (Figure 1.4C) (Xie, 2013). The MATH domain, also known as the TRAF-C domain, is at the C-terminus of TRAFs 1-6, which trimerize via the TRAF-N coiled-coil domain. N-terminal to the TRAF domain are zinc-finger domains, and TRAFs 2-6 also contain N-terminal RING domains, which play a regulatory role in many signaling pathways through their E3 ubiquitin ligase activity. TRAFs also act as adapter proteins between signaling receptors and downstream factors, which can interact with their coiled-coil

domain (e.g. cIAP1 and 2), or be recruited or activated by the K63-linked polyubiquitin chains created by the TRAF RING domains (Xie, 2013). TRAFs 1, 2, 3, and 5 are close homologs that are primarily found in vertebrates, while TRAFs 4 and 6 are more ancient proteins (Zapata et al., 2007). The binding preferences of the MATH domains reflect this evolutionary history in that TRAFs 1, 2, 3, and 5 share many of the same binding partners, recognizing a major ((P/S/A/T)x(Q/E)E) and minor (PxQxxD) motif (Ye et al., 1999). TRAF6, though it plays a role in many of the same pathways and binds to some of the same proteins, recognizes a different motif (PxExx(Aromatic/Acidic)) (Ye et al., 2002). Binding partners of TRAF4 have not been characterized in direct binding assays as peptides, and the current structures of TRAF4 are apo structures, but there is some evidence that it may compete with TRAF6 for binding the same sites on NOD2 and Act1 (Marinis et al., 2011; Zepp et al., 2012). The TRAF6 and TRAF1/2/3/5 motifs have obvious similarities, but the peptides bind in different orientations on the MATH domains. Structures of TRAF6 bound to peptides from CD40 and RANK show the peptide bound in a β-strand conformation nearly parallel to a MATH domain β-strand and able to form backbone hydrogen bonds (Ye et al., 2002). In contrast, structures of peptides bound to TRAFs 2 and 3 show the peptides in a polyproline type II helix conformation, bound at an angle of 40° in comparison with the TRAF6-bound peptides (Ni et al., 2000; Ye et al., 1999). This difference can be attributed at least in part to a change in the position of the pocket that binds the proline in each motif, which is determined by different positioning of the aromatic residues that form the pocket.

The MATH-peptide interactions of TRAFs are very weak; the tightest reported monomeric K_D, measured for TRAF3 binding to a peptide from the cytoplasmic protein TANK, is 24 μM, and many interactions are in the hundreds of micromolar (Li et al., 2002). The weak

nature of the monomeric interactions is an important feature of TRAF function, as pathways are only activated upon trimerization of their binding partners, such as the trimerization of TNFR2 by its ligand TNFα. Trimerization, or higher-order oligomerization, allows for avidity to recruit the trimeric TRAFs (Pullen et al., 1999b). Their weak nature makes the TRAF-peptide interactions challenging to study *in vitro*, and only a limited number of the affinities have been directly quantified. Therefore, relatively little is known about the binding determinants of TRAFs beyond the motifs that have been constructed from multiple sequence alignments of known binding partners.

The most detailed information on binding preferences comes from SPOT array analysis of the two TRAF binding sites in the cytoplasmic tail of the TNFR homolog CD40 (Pullen et al., 1999a). The analysis examined binding of TRAFs 1, 2, 3, and 6 to mutants (all 20 amino acids) of each site of the TRAF 1/2/3 and TRAF6 binding sites on CD40. The results primarily emphasized the lack of tolerance of mutation at the conserved motif positions. However, SPOT arrays performed with CD40 peptides of differing length indicated that TRAF3 may also bind CD40 sequence C-terminal to the core-binding site defined as the motif region. The increased affinity supplied by these exosite interactions relaxes the binding preferences of TRAF3 in the core peptide region. This observation did not hold true for TRAF2, suggesting that extra interactions outside of the core motif may be one specificity mechanism employed by TRAF3. Structures of TRAF3 with long peptides (>17 residues) from CD40, BAFF-R, TANK, and LTβR show extra interactions mediated by peptide regions beyond the core motif (Ely and Li, 2002; Li et al., 2002; Ni et al., 2000; 2004). All TRAF2 structures were solved with short peptides encompassing primarily the motif, so it is not known whether other TRAFs can utilize interactions outside the core motif for enhanced affinity or specificity. The use of peptide

residues beyond the motif regions is a common specificity strategy for PRDs and an aspect of protein-peptide binding specificity that is often poorly understood due to the convention of solving structures with peptides corresponding to only the minimal motif (Stein and Aloy, 2008).

Another example of specificity between close TRAF homologs was detailed in a recent paper on the binding of TRAF3 to the adapter protein Cardif (also known as Mavs) (Zhang et al., 2012). TRAF3, but not TRAF5, is recruited by Cardif upon intracellular detection of viral double stranded RNA, leading to the induction of type I interferons (Saha et al., 2006). TRAF5 cannot bind the TRAF interacting motif (TIM) in Cardif. Zhang et al. demonstrated that this specificity was due in part to interactions between a tyrosine hydroxyl on TRAF3 and the main chain carbonyls of two prolines present in Cardif C-terminal to the core motif. TRAF5 has a phenylalanine at the same position, and mutation of this residue to tyrosine allowed TRAF5 to bind to CARDIF. Thus, there is ample evidence that TRAF3 utilizes interactions beyond the core TIM to determine its binding specificity.

With many large families of PRDs and even more potential peptide binding partners, evolution has arrived at several methods of achieving the binding specificity necessary for proper functioning of cellular pathways. As described for SH3, PDZ, and WW domains, family members can utilize different loops and binding site residues to recognize very different motifs. Other families of proteins with overlapping functions, like the anti-apoptotic Bcl-2s, recognize variations on the motif and context residues to achieve different binding profiles. Context residues can include both the "x" residues between hotspot residues represented by the motif and more distant peptide sites that may bind exosites on the PRD surface, as is seen for TRAF3 and the EVH1 domains. This supports the idea of protein-protein interaction specificity as a continuum within sequence space. Large studies of PDZ, SH2, and WW domain binding

preferences have suggested that classifying PRDs according to the motifs they bind is not a productive exercise, as the number of classes rapidly multiplies, and there are always domains with binding preferences that do not fit neatly into a class (Ingham et al., 2005; Machida et al., 2007; Stiffler et al., 2007). This observation holds true on both specificity levels: specificity determined by different motifs, and specificity determined by fine-scale differences in context. With this nuanced view of protein-peptide interaction specificity determinants, the difficulty of accurately modeling a given protein's binding preferences increases. Exhaustive sampling of the peptide sequence space bound by each protein by experimental methods is impossible, especially when considering the many PRD families that have been only lightly characterized like TRAF and EVH1 domains. Therefore, computational models of binding preferences are increasingly important for the goal of a systems-level understanding of protein-protein interaction networks. Structure-based models offer the best prospects for comprehensive coverage of sequence space. However, structure-based computational models can only accurately model the portion of interactions captured in structures. Experimental binding datasets will remain important for many years to come, as we still require benchmarks to test and improve computational techniques.

Applications of specificity prediction

Interactome prediction

The applications of accurate binding models for interactome prediction and inhibitor design are of vital importance for our understanding of cellular pathways and our ability to regulate protein-peptide interactions for therapeutic purposes. Interactome prediction for protein-peptide interactions is generally performed by looking for potential peptide partners of given domains and requires a model for scoring peptides as binding partners. Such models can be data-based, or structure-based. Many studies that have generated large binding datasets have used the

positive hits to form position specific scoring matrices, or a related motif representation, and then screened the entire proteome for more hits that could match the motif. This approach was used in a study that combined information from peptide microarray and phage display libraries of PDZ binders via a support vector machine (SVM) (Hui and Bader, 2010). Several studies in recent years have used computational, structure-based models to predict new partners. In a follow up to the experiment-based SVM, Hui et al. created a structure-based SVM built on features encoded in PDZ:peptide structures (Hui et al., 2013). This model was found to have more power to identify sequences with features that were not in the training set, unlike the experiment-based model. Because these methods can generate very long lists of potential partners, extra criteria are often used to filter the results. Hou et al. built a scoring matrix from the calculated binding energy of point mutants of AKAP peptides binding to the RIIalpha subunit of protein kinase A (Hou et al., 2011). This scoring matrix was used to score all peptides from the SWISSPROT database, and high-scoring peptides were further filtered for helicity, predicted presence in transmembrane regions, and conservation in seven vertebrate species. The goal of this extra filtering was to reduce false positives and narrow the list of peptides to be tested experimentally, but it also resulted in removal of some known AKAP peptides from the list.

Another example that used less filtering and directly scored more peptides was employed by DeBartolo et al. to find new BH3 peptides that interact with human anti-apoptotic Bcl-2 homologs (DeBartolo et al., 2014). The only pre-filtering imposed in this study was to impose a loose BH3-like requirement on peptides found by sliding a 26-residue window through the human proteome; two positions were required to be a restricted set of amino acids, and the overall peptide was required to include at least 35% polar/charged residues. The peptides that fit these criteria were scored using both experimental models built from peptide SPOT arrays

($PSSM_{SPOT}$) and STATIUM, a structure-based model. Interestingly, retrospective analysis of confirmed binding partners showed that a combination of both the $PSSM_{SPOT}$ and STATIUM models gave the best predictive performance in identifying binders. Such studies identify many putative partners and provide *in vitro* validation for the top-scoring subset, thus supplying a large set of strong hypotheses for the cellular interactome. Further experimental validation with full-length proteins in cells will be required before the efficacy of these approaches can be judged. Accurate binding models also have the potential to predict the effect of disease-associated mutations on interactions. This approach has been used to predict the effects of non-synonymous SNPs on formation or disruption of PDZ binding sites on peptides in the human proteome (Gfeller et al., 2014).

Inhibitor design

Accurate binding models are also useful for designing peptide inhibitors of proteins of interest. Peptide inhibitors have several applications: interrogation of protein function, use as diagnostics, or use as therapeutics. Discovering peptide inhibitors can be easier than discovering small molecules because, as previously discussed, very large peptide libraries can be encoded genetically. Peptides also offer larger interfaces, and this makes them well suited to targeting protein-protein interfaces, which generally offer few of the deep pockets preferred for small molecule targets. The downsides of peptide inhibitors can include protease susceptibility and lack of cell membrane penetrance. These issues have been circumvented in recent years by modifications or peptidomimetic strategies including adding hydrocarbon staples to α-helical peptides, adding di-proline mimetics in peptides that bind as polypropline type II helices, cyclization, or using backbone modifications, among others (Bock et al., 2013; Huang et al., 2007; Opitz et al., 2015; Stewart et al., 2010). Before reagent optimization, a starting sequence

with the desired binding specificity is required. Because natural sequences often bind more than one homologous PRD, design and selection strategies are needed to create peptides with the desired binding profiles. Binding models and libraries have been used separately, and in conjunction for this purpose (Chen and Keating, 2012).

Anti-apoptotic Bcl-2 homologs are one family for which selective peptide inhibitors are in high demand. Stapled versions of natural BH3 peptide sequences have been used to effectively inhibit Bcl-2 family members in cancer models in cells and *in vivo* (Stewart et al., 2010; Walensky et al., 2004). Natural BH3 sequences are also used in BH3 profiling assays, which assess which anti-apoptotic Bcl-2 homolog a cancer cell line or primary tumor is dependent upon by comparing mitochondrial permeabilization response to *in vitro* BH3 binding profiles (Certo et al., 2006). For both applications, BH3 peptides that bind with high affinity and specificity to just one Bcl-2 homolog over the other five homologs are desired. Specificity is critical to reduce off-target effects and for clear diagnostic read-outs. A recent study by Dutta et al. used experiment-based models to design a combinatorial peptide library enriched in BH3 peptides exhibiting specificity for Bcl-x_L over the other four main Bcl-2 homologs (Dutta et al., 2014). This library was then screened by yeast surface display for peptides showing specificity and high affinity for Bcl-x_L. Resulting peptides, with some further rational mutagenesis, exhibited the potential to discriminate between Bcl-x_L and Bcl-2 dependence in cancer cell lines. The same approach of computational library design guided by experiment-based models, followed by screening was previously used to find peptides with specificity for Bfl-1 over the other Bcl-2 homologs (Dutta et al., 2013). These libraries were restricted to a subset of the positions that were sampled on SPOT arrays of the Bim BH3, and therefore did not sample all possible specificity determinants, most notably, the Y4e position, at which a specific mutation was introduced by rational design in

the Bcl-x_L-specific peptide. STATIUM, the purely structure-based model, can model all peptide positions that make close interactions with the Bcl-2 receptors. Notably, STATIUM exhibits potential for modeling specificity between Bcl-2 homologs (DeBartolo et al., 2012; 2014). Thus, future design cycles could combine these experimental and structure-based models for inhibitor design, as they were successfully combined for interactome prediction (DeBartolo et al., 2014).

Research approach

In my thesis work I expanded on the Keating lab's Bcl-2 work by designing specific peptide inhibitors of human Mcl-1 and two viral Bcl-2 homologs, and I began characterization of TRAF:TRAF interaction motif (TIM) binding determinants. My thesis project began with the goals of comparing the BH3 binding determinants of viral Bcl-2 homologs (KSBcl-2 and BHRF1) to those of the human homologs and designing specific inhibitors of the viral homologs. Initial yeast display libraries of BH3 peptides showed that KSBcl-2 has very similar binding preferences to those of the human homolog Mcl-1, which is an important cancer target. In chapter 2, I describe how I developed peptides from this initial screen as Mcl-1-specific inhibitors that showed greatly improved affinity and specificity over a natural BH3 peptide previously used to target Mcl-1. Our collaborators were able to use these peptides to identify Mcl-1-dependence in cancer cell lines using BH3 profiling assays. In chapter 3, I further explored the relationship between viral and human Bcl-2 homologs by comparing sequence, structure, and BH3 binding similarity. A second round of BH3 libraries were designed for viral Bcl-2 binding specificity using $PSSM_{SPOT}$ and STATIUM models for all viral and human homologs. These libraries were screened for specificity to KSBcl-2 or BHRF1 using a bacterial surface display system that I optimized for affinity-based screening of peptide libraries. Resulting peptides were specific for KSBcl-2 or BHRF1 over human homologs Bfl-1, Bcl-x_L,

Bcl-2, and Bcl-w, but still showed tight-to-moderate binding to Mcl-1. Further rational mutagenesis improved the specificity of the library peptides to yield reagents with high affinity and specificity for targeting KSBcl-2 and BHRF1, which are from human herpesviruses associated with cancer. Thus, I leveraged the growing body of Bcl-2 binding data and models, as well as library design and screening approaches to add to the repertoire of homolog-specific Bcl-2 inhibitors. In the process, I advanced our knowledge of binding determinants in this family and elucidated important experimental considerations for peptide specificity design.

Finally, in chapter 4, I describe our preliminary work on the interactions between TRAFs and TIMs. Beyond some basic motifs, the binding determinants for this system are not well determined. TRAF interactions present several challenges. They are very weak, and the extent of their interaction footprint on their peptide partners is not well known. I first optimized a display strategy for TIMs on the surface of *E. coli*. I then performed deep mutational scanning of all single and double point mutants for 3, 7-mer segments along two peptide TIMs from CD40 and TANK. Sequences enriched for binding the close homologs TRAF2, TRAF3, and TRAF5 show both commonalities and differences that yield several new hypotheses about TRAF binding preferences.

References

Akiva, E., Friedlander, G., Itzhaki, Z., and Margalit, H. (2012). A dynamic view of domain-motif interactions. PLoS Comput Biol *8*, e1002341.

Araya, C.L., Fowler, D.M., Chen, W., Muniez, I., Kelly, J.W., and Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci USA *109*, 16858–16863.

Ball, L.J., Kühne, R., Hoffmann, B., Häfner, A., Schmieder, P., Volkmer-Engert, R., Hof, M., Wahl, M., Schneider-Mergener, J., Walter, U., et al. (2000). Dual epitope recognition by the VASP EVH1 domain modulates polyproline ligand specificity and binding affinity. Embo J *19*, 4903–4914.

Bock, J.E., Gavenonis, J., and Kritzer, J.A. (2013). Getting in shape: controlling peptide bioactivity and bioavailability using conformational constraints. ACS Chem Biol *8*, 488–499.

Boersma, M., Sadowsky, J., and Tomita, Y. (2008). Hydrophile scanning as a complement to alanine scanning for exploring and manipulating protein-protein recognition: Application to the Bim BH3 domain. Protein Sci *17*, 1232-1240.

Boëda, B., Briggs, D.C., Higgins, T., Garvalov, B.K., Fadden, A.J., McDonald, N.Q., and Way, M. (2007). Tes, a specific Mena interacting partner, breaks the rules for EVH1 binding. Mol Cell *28*, 1071–1082.

Certo, M., Del Gaizo Moore, V., Nishino, M., Wei, G., Korsmeyer, S., Armstrong, S.A., and Letai, A. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. Cancer Cell *9*, 351–365.

Chang, B.H., Gujral, T.S., Karp, E.S., BuKhalid, R., Grantcharova, V.P., and MacBeath, G. (2011). A systematic family-wide investigation reveals that ~30% of mammalian PDZ domains engage in PDZ-PDZ interactions. Chem Biol *18*, 1143–1152.

Chen, L., Willis, S.N., Wei, A., Smith, B.J., Fletcher, J.I., Hinds, M.G., Colman, P.M., Day, C.L., Adams, J.M., and Huang, D.C.S. (2005). Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. Mol Cell *17*, 393–403.

Chen, T.S., and Keating, A.E. (2012). Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods. Protein Sci *21*, 949–963.

Chen, T.S., Palacios, H., and Keating, A.E. (2013). Structure-based redesign of the binding specificity of anti-apoptotic Bcl-x(L). J Mol Biol *425*, 171–185.

Danial, N.N. (2008). BAD: undertaker by night, candyman by day. Oncogene *27 Suppl 1*, S53–S70.

de Lichtenberg, U., Jensen, L.J., Brunak, S., and Bork, P. (2005). Dynamic complex formation

during the yeast cell cycle. Science *307*, 724–727.

DeBartolo, J., Dutta, S., Reich, L., and Keating, A.E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. J Mol Biol *422*, 124–144.

DeBartolo, J., Taipale, M., and Keating, A.E. (2014). Genome-wide prediction and validation of peptides that bind human prosurvival Bcl-2 proteins. PLoS Comput Biol *10*, e1003693.

Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res *42*, D259–D266.

Dutta, S., Chen, T.S., and Keating, A.E. (2013). Peptide ligands for pro-survival protein Bfl-1 from computationally guided library screening. ACS Chem Biol *8*, 778–788.

Dutta, S., Gullá, S., Chen, T.S., Fire, E., Grant, R.A., and Keating, A.E. (2010). Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. J Mol Biol *398*, 747–762.

Dutta, S., Ryan, J., Scott Chen, T., Kougentakis, C., Letai, A., and Keating, A.E. (2014). Potent and specific peptide inhibitors of human pro-survival protein Bcl-xL. J Mol Biol 427, 1241-1253.

Ely, K.R., and Li, C. (2002). Structurally adaptive hot spots at a protein interaction interface on TRAF3. J Mol Recognit *15*, 286–290.

Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P.M., Bader, G.D., and Sidhu, S.S. (2010). Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Molecular BioSystems *6*, 1782–1790.

Fedorov, A.A., Fedorov, E., Gertler, F., and Almo, S.C. (1999). Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. Nat Struct Biol *6*, 661–665.

Fire, E., Gullá, S., Grant, R., and Keating, A. (2010). Mcl-1-Bim complexes accommodate surprising point mutations via minor structural changes. Protein Sci *19*, 507-519.

Forsyth, C.M., Juan, V., Akamatsu, Y., DuBridge, R.B., Doan, M., Ivanov, A.V., Ma, Z., Polakoff, D., Razo, J., Wilson, K., et al. (2013). Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. MAbs *5*, 523–532.

Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat Methods *11*, 801–807.

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat Methods *7*, 741–746.

Fujino, Y., Fujita, R., Wada, K., Fujishige, K., Kanamori, T., Hunt, L., Shimizu, Y., and Ueda, T. (2012). Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. Biochem Biophys Res Commun *428*, 395–400.

Gertler, F., and Condeelis, J. (2011). Metastasis: tumor cells becoming MENAcing. Trends Cell Biol *21*, 81–90.

Gfeller, D., Butty, F., Wierzbicka, M., Verschueren, E., Vanhee, P., Huang, H., Ernst, A., Dar, N., Stagljar, I., Serrano, L., et al. (2011). The multiple-specificity landscape of modular peptide recognition domains. Mol Syst Biol *7*, 484.

Gfeller, D., Ernst, A., Jarvik, N., Sidhu, S.S., and Bader, G.D. (2014). Prediction and experimental characterization of nsSNPs altering human PDZ-binding motifs. PLoS ONE *9*, e94507.

Gold, M.G., Fowler, D.M., Means, C.K., Pawson, C.T., Stephany, J.J., Langeberg, L.K., Fields, S., and Scott, J.D. (2013). Engineering A-kinase anchoring protein (AKAP)-selective regulatory subunits of protein kinase A (PKA) through structure-based phage selection. J Biol Chem *288*, 17111–17121.

Hegde, S.R., Manimaran, P., and Mande, S.C. (2008). Dynamic changes in protein functional linkage networks revealed by integration with gene expression data. PLoS Comput Biol *4*, e1000237.

Hillier, B.J. (1999). Unexpected Modes of PDZ Domain Scaffolding Revealed by Structure of nNOS-Syntrophin Complex. Science *284*, 812–815.

Hou, T., Li, N., Li, Y., and Wang, W. (2012). Characterization of domain-peptide interaction interface: prediction of SH3 domain-mediated protein-protein interaction network in yeast by generic structure-based models. J. Proteome Res. *11*, 2982–2995.

Hou, T., Li, Y., and Wang, W. (2011). Prediction of peptides binding to the PKA RIIalpha subunit using a hierarchical strategy. Bioinformatics *27*, 1814–1821.

Hou, T., Xu, Z., Zhang, W., McLaughlin, W.A., Case, D.A., Xu, Y., and Wang, W. (2009). Characterization of domain-peptide interaction interface: a generic structure-based model to decipher the binding specificity of SH3 domains. Mol. Cell Proteomics *8*, 639–649.

Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., Carter, M., Chen, L., et al. (2004). A map of WW domain family interactions. Proteomics *4*, 643–655.

Huang, D., Tomita, Y., and Gellman, S. (2007). (α/β+ α)-Peptide Antagonists of BH3 Domain/Bcl-xL Recognition: Toward General Strategies for Foldamer-Based Inhibition of Protein− Protein Interactions. J. Am. Chem. Soc. 129, 139-154.

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., and Eck, M.J. (2000). Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. Nat Struct Biol *7*, 634–638.

Hui, S., and Bader, G.D. (2010). Proteome scanning to predict PDZ domain interactions using support vector machines. BMC Bioinformatics *11*, 507.

Hui, S., Xing, X., and Bader, G.D. (2013). Predicting PDZ domain mediated protein interactions from structure. BMC Bioinformatics *14*, 27.

Ingham, R.J., Colwill, K., Howard, C., Dettwiler, S., Lim, C.S.H., Yu, J., Hersi, K., Raaijmakers, J., Gish, G., Mbamalu, G., et al. (2005). WW domains provide a platform for the assembly of multiprotein networks. Mol Cell Biol *25*, 7092–7106.

Kaneko, T., Sidhu, S.S., and Li, S.S.C. (2011). Evolving specificity from variability for protein interaction domains. Trends Biochem Sci *36*, 183–190.

Kim, T., Tyndel, M.S., Huang, H., Sidhu, S.S., Bader, G.D., Gfeller, D., and Kim, P.M. (2012). MUSI: an integrated system for identifying multiple specificity from very large peptide or nucleic acid data sets. Nucleic Acids Res *40*, e47.

Komurov, K., and White, M. (2007). Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. Mol Syst Biol *3*, 110.

Ku, B., Liang, C., Jung, J.U., and Oh, B.-H. (2011). Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. Cell Res *21*, 627–641.

Lee, E.F., Dewson, G., Evangelista, M., Pettikiriarachchi, A., Gold, G.J., Zhu, H., Colman, P.M., and Fairlie, W.D. (2014). The functional differences between pro-survival and pro-apoptotic B cell lymphoma 2 (Bcl-2) proteins depend on structural differences in their Bcl-2 homology 3 (BH3) domains. J Biol Chem *289*, 36001–36017.

Lee, E.F., Fedorova, A., Zobel, K., Boyle, M.J., Yang, H., Perugini, M.A., Colman, P.M., Huang, D.C.S., Deshayes, K., and Fairlie, W.D. (2009a). Novel Bcl-2 homology-3 domain-like sequences identified from screening randomized peptide libraries for inhibitors of the pro-survival Bcl-2 proteins. J Biol Chem *284*, 31315–31326.

Lee, E.F., Sadowsky, J.D., Smith, B.J., Czabotar, P.E., Peterson-Kaufman, K.J., Colman, P.M., Gellman, S.H., and Fairlie, W.D. (2009b). High-resolution structural characterization of a helical alpha/beta-peptide foldamer bound to the anti-apoptotic protein Bcl-xL. Angew Chem Int Ed Engl *48*, 4318–4322.

Lee, H.-J., and Zheng, J.J. (2010). PDZ domains and their binding partners: structure, specificity, and modification. Cell Commun. Signal *8*, 8.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. Nucleic Acids Res *43*, D257–D260.

Levin, A.M., and Weiss, G.A. (2006). Optimizing the affinity and specificity of proteins with molecular display. Molecular BioSystems *2*, 49–57.

Li, C., Ni, C.-Z., Havert, M.L., Cabezas, E., He, J., Kaiser, D., Reed, J.C., Satterthwait, A.C., Cheng, G., and Ely, K.R. (2002). Downstream regulator TANK binds to the CD40 recognition site on TRAF3. Structure *10*, 403–411.

Li, N., Hou, T., Ding, B., and Wang, W. (2011). Characterization of PDZ domain-peptide interaction interface based on energetic patterns. Proteins *79*, 3208–3220.

Li, N., Stein, R.S.L., He, W., Komives, E., and Wang, W. (2013). Identification of methyllysine peptides binding to chromobox protein homolog 6 chromodomain in the human proteome. Mol. Cell Proteomics *12*, 2750–2760.

Li, S.S.C. (2005). Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. Biochem J *390*, 641–653.

Liu, B.A., Engelmann, B.W., and Nash, P.D. (2012). High-throughput analysis of peptide-binding modules. Proteomics *12*, 1527–1546.

Liu, Q., Berry, D., Nash, P., Pawson, T., McGlade, C.J., and Li, S.S.-C. (2003). Structural basis for specific binding of the Gads SH3 domain to an RxxK motif-containing SLP-76 peptide: a novel mode of peptide recognition. Mol Cell *11*, 471–481.

Llambi, F., Moldoveanu, T., Tait, S.W.G., Bouchier-Hayes, L., Temirov, J., McCormick, L.L., Dillon, C.P., and Green, D.R. (2011). A unified model of mammalian BCL-2 protein family interactions at the mitochondria. Mol Cell *44*, 517–531.

London, N., Gullá, S., Keating, A.E., and Schueler-Furman, O. (2012). In silico and in vitro elucidation of BH3 binding specificity toward Bcl-2. Biochemistry *51*, 5841–5850.

London, N., Lamphear, C.L., Hougland, J.L., Fierke, C.A., and Schueler-Furman, O. (2011). Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS Comput Biol *7*, e1002170.

London, N., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). The structural basis of peptide-protein binding strategies. Structure *18*, 188–199.

London, N., Raveh, B., and Schueler-Furman, O. (2013). Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. Curr Opin Struct Biol *23*, 894–902.

Löfblom, J. (2011). Bacterial display in combinatorial protein engineering. Biotechnol J *6*, 1115–1129.

Machida, K., Thompson, C.M., Dierck, K., Jablonowski, K., Kärkkäinen, S., Liu, B., Zhang, H., Nash, P.D., Newman, D.K., Nollau, P., et al. (2007). High-throughput phosphotyrosine profiling using SH2 domains. Mol Cell *26*, 899–915.

Marinis, J.M., Homer, C.R., McDonald, C., and Abbott, D.W. (2011). A Novel Motif in the Crohn's Disease Susceptibility Protein, NOD2, Allows TRAF4 to Down-regulate Innate Immune

Responses. J Biol Chem *286*, 1938–1950.

McLaughlin, R.N., Jr, Poelwijk, F.J., Raman, A., Gosal, W.S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–142.

McLaughlin, W.A., Hou, T., Taylor, S.S., and Wang, W. (2011). The identification of novel cyclic AMP-dependent protein kinase anchoring proteins using bioinformatic filters and peptide arrays. Protein Eng Des Sel *24*, 333–339.

Melamed, D., Young, D.L., Gamble, C.E., Miller, C.R., and Fields, S. (2013). Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA *19*, 1537–1551.

Melamed, D., Young, D.L., Miller, C.R., and Fields, S. (2015). Combining natural sequence variation with high throughput mutational data to reveal protein interaction sites. PLoS Genet. *11*, e1004918.

Neduva, V., and Russell, R.B. (2006). DILIMOT: discovery of linear motifs in proteins. Nucleic Acids Res *34*, W350–W355.

Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T.J., Lewis, J., Serrano, L., and Russell, R.B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol *3*, e405.

Ni, C.Z., Welsh, K., Leo, E., Chiou, C.K., Wu, H., Reed, J.C., and Ely, K.R. (2000). Molecular basis for CD40 signaling mediated by TRAF3. Proc Natl Acad Sci USA *97*, 10395–10399.

Ni, C.-Z., Oganesyan, G., Welsh, K., Zhu, X., Reed, J.C., Satterthwait, A.C., Cheng, G., and Ely, K.R. (2004). Key molecular contacts promote recognition of the BAFF receptor by TNF receptor-associated factor 3: implications for intracellular signaling regulation. J Immunol *173*, 7394–7400.

Nishida, M., Nagata, K., Hachimori, Y., Horiuchi, M., Ogura, K., Mandiyan, V., Schlessinger, J., and Inagaki, F. (2001). Novel recognition mode between Vav and Grb2 SH3 domains. Embo J *20*, 2995–3007.

Opitz, R., Müller, M., Reuter, C., Barone, M., Soicke, A., Roske, Y., Piotukh, K., Huy, P., Beerbaum, M., Wiesner, B., et al. (2015). A modular toolkit to inhibit proline-rich motif-mediated protein-protein interactions. Proc Natl Acad Sci USA *112*, 5011–5016.

Palopoli, N., Lythgow, K.T., and Edwards, R.J. (2015). QSLiMFinder: improved short linear motif prediction using specific query protein data. Bioinformatics, doi: 10.1093/bioinformatics/btv155.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445–452.

Pál, G., Kouadio, J.-L.K., Artis, D.R., Kossiakoff, A.A., and Sidhu, S.S. (2006). Comprehensive

and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. J Biol Chem *281*, 22378–22385.

Penkert, R.R., DiVittorio, H.M., and Prehoda, K.E. (2004). Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. Nature Struct & Mol Bio 10(10), 1122-1127.

Peterson, F.C., Penkert, R.R., Volkman, B.F., and Prehoda, K.E. (2004). Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. Mol Cell *13*, 665–676.

Peterson, F, and Volkman, B. (2009). Diversity of polyproline recognition by EVH1 domains. Front Biosci 14, 833-846.

Petsalaki, E., and Russell, R.B. (2008). Peptide-mediated interactions in biological systems: new discoveries and applications. Curr Opin Biotechnol *19*, 344–350.

Prehoda, K.E., Lee, D.J., and Lim, W.A. (1999). Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. Cell *97*, 471–480.

Pullen, S.S., Dang, T.T., Crute, J.J., and Kehry, M.R. (1999a). CD40 signaling through tumor necrosis factor receptor-associated factors (TRAFs). Binding site specificity and activation of downstream pathways by distinct TRAFs. J Biol Chem *274*, 14246–14254.

Pullen, S.S., Labadia, M.E., Ingraham, R.H., McWhirter, S.M., Everdeen, D.S., Alber, T., Crute, J.J., and Kehry, M.R. (1999b). High-affinity interactions of tumor necrosis factor receptor-associated factors (TRAFs) and CD40 require TRAF trimerization and CD40 multimerization. Biochemistry *38*, 10168–10177.

Rautureau, G.J.P., Day, C.L., and Hinds, M.G. (2010). Intrinsically Disordered Proteins in Bcl-2 Regulated Apoptosis. Int. J. Mol. Sci. 11, 1808-1824.

Raveh, B., London, N., and Schueler-Furman, O. (2010). Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins *78*, 2029–2040.

Reich, L.L., Dutta, S., and Keating, A.E. (2014). SORTCERY-A High-Throughput Method to Affinity Rank Peptide Ligands. J Mol Biol. 427(11), 2135-2150.

Saha, S.K., Pietras, E.M., He, J.Q., Kang, J.R., Liu, S.-Y., Oganesyan, G., Shahangian, A., Zarnegar, B., Shiba, T.L., Wang, Y., et al. (2006). Regulation of antiviral responses by a direct and specific interaction between TRAF3 and Cardif. Embo J *25*, 3257–3263.

Sakamaki, J.-I., Daitoku, H., Ueno, K., Hagiwara, A., Yamagata, K., and Fukamizu, A. (2011). Arginine methylation of BCL-2 antagonist of cell death (BAD) counteracts its phosphorylation and inactivation by Akt. Proc Natl Acad Sci USA *108*, 6085–6090.

Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics, doi:10.1534/genetics.115.175802.

Stein, A., and Aloy, P. (2008). Contextual specificity in peptide-mediated protein interactions. PLoS ONE *3*, e2524.

Stewart, M., Fire, E., and Keating, A. (2010). The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. Nat Chem Biol 6, 595-601.

Stiffler, M.A., Chen, J.R., Grantcharova, V.P., Lei, Y., Fuchs, D., Allen, J.E., Zaslavskaia, L.A., and MacBeath, G. (2007). PDZ domain binding selectivity is optimized across the mouse proteome. Science *317*, 364–369.

Tang, X., Wang, J., Liu, B., Li, M., Chen, G., and Pan, Y. (2011). A comparison of the functional modules identified from time course and static PPI network data. BMC Bioinformatics *12*, 339.

Tinberg, C.E., Khare, S.D., Dou, J., Doyle, L., Nelson, J.W., Schena, A., Jankowski, W., Kalodimos, C.G., Johnsson, K., Stoddard, B.L., et al. (2013). Computational design of ligand-binding proteins with high affinity and selectivity. Nature *501*, 212–216.

Tonikian, R., Xin, X., Toret, C.P., Gfeller, D., Landgraf, C., Panni, S., Paoluzi, S., Castagnoli, L., Currell, B., Seshagiri, S., et al. (2009). Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol *7*, e1000218.

Tonikian, R., Zhang, Y., Sazinsky, S.L., Currell, B., Yeh, J.-H., Reva, B., Held, H.A., Appleton, B.A., Evangelista, M., Wu, Y., et al. (2008). A specificity map for the PDZ domain family. PLoS Biol *6*, e239.

Traxlmayr, M.W., Hasenhindl, C., Hackl, M., Stadlmayr, G., Rybka, J.D., Borth, N., Grillari, J., Rüker, F., and Obinger, C. (2012). Construction of a stability landscape of the CH3 domain of human IgG1 by combining directed evolution with high throughput sequencing. J Mol Biol *423*, 397–412.

Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat Struct Biol *7*, 639–643.

Vinayagam, A., Stelzl, U., and Wanker, E.E. (2009). Repeated two-hybrid screening detects transient protein–protein interactions. Theor Chem Acc *125*, 613–619.

Walensky, L.D., Kung, A.L., Escher, I., Malia, T.J., Barbuto, S., Wright, R.D., Wagner, G., Verdine, G.L., and Korsmeyer, S.J. (2004). Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. Science *305*, 1466–1470.

Whitehead, T.A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S.J., De Mattos, C., Myers, C.A., Kamisetty, H., Blair, P., Wilson, I.A., et al. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol *30*, 543–548.

Wittekind, M., Mapelli, C., Lee, V., Goldfarb, V., Friedrichs, M.S., Meyers, C.A., and Mueller, L. (1997). Solution structure of the Grb2 N-terminal SH3 domain complexed with a ten-residue peptide derived from SOS: direct refinement against NOEs, J-couplings and 1H and 13C

chemical shifts. J Mol Biol *267*, 933–952.

Xie, P. (2013). TRAF molecules in cell signaling and in human diseases. J Mol Signal *8*, 7.

Xin, X., Gfeller, D., Cheng, J., Tonikian, R., Sun, L., Guo, A., Lopez, L., Pavlenco, A., Akintobi, A., Zhang, Y., et al. (2013). SH3 interactome conserves general function over specific form. Mol Syst Biol *9*, 652.

Yang, E., Zha, J., Jockel, J., Boise, L.H., Thompson, C.B., and Korsmeyer, S.J. (1995). Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. Cell *80*, 285–291.

Ye, H., Park, Y.C., Kreishman, M., Kieff, E., and Wu, H. (1999). The structural basis for the recognition of diverse receptor sequences by TRAF2. Mol Cell *4*, 321–330.

Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaia, M., Yim, M., et al. (2002). Distinct molecular mechanism for initiating TRAF6 signalling. Nature *418*, 443–447.

Zapata, J.M., Martínez-García, V., and Lefebvre, S. (2007). Phylogeny of the TRAF/MATH domain. Adv Exp Med Biol *597*, 1–24.

Zepp, J.A., Liu, C., Qian, W., Wu, L., Gulen, M.F., Kang, Z., and Li, X. (2012). Cutting edge: TNF receptor-associated factor 4 restricts IL-17-mediated pathology and signaling processes. J Immunol *189*, 33–37.

Zhang, P., Reichardt, A., Liang, H., Aliyari, R., Cheng, D., Wang, Y., Xu, F., Cheng, G., and Liu, Y. (2012). Single Amino Acid Substitutions Confer the Antiviral Activity of the TRAF3 Adaptor Protein onto TRAF5. Sci Signal *5*, ra81–ra81.

Chapter 2

Designed BH3 peptides with high affinity and specificity for targeting Mcl-1 in cells

Reproduced with permission from

Foight, G. W., Ryan, J. A., Gullá, S. V., Letai, A., and Keating, A. E. (2014) Designed BH3 Peptides with High Affinity and Specificity for Targeting Mcl-1 in Cells. *ACS Chem Biol 9*, 1962-1968.

G.W.F and A.E.K. designed and wrote the study. J.A.R. performed the BH3 profiling assays. S.V.G. made the yeast display library. G.W.F. did all remaining experiments.

Abstract

Mcl-1 is over-expressed in many cancers and can confer resistance to cell-death signaling in refractory disease. Molecules that specifically inhibit Mcl-1 hold potential for diagnosing and disrupting Mcl-1-dependent cell survival. We selected three peptides from a yeast-surface display library that showed moderate specificity and affinity for binding to Mcl-1 over Bfl-1, Bcl-x_L, Bcl-2, and Bcl-w. Specificity for Mcl-1 was improved by introducing threonine at peptide position 2e. The most specific peptide, MS1, bound Mcl-1 with 40-fold or greater specificity over four other human Bcl-2 paralogs. In BH3 profiling assays, MS1 caused depolarization in several human Mcl-1-dependent cell lines with EC_{50} values of ~3 μM, contrasted with EC_{50} values >100 μM for Bcl-2-, Bcl-x_L-, or Bfl-1-dependent cell lines. MS1 is at least 30-fold more potent in this assay than the previously used Mcl-1 targeting reagent NoxaA BH3. These peptides can be used to detect Mcl-1 dependency in cells and provide leads for developing Mcl-1 targeting therapeutics.

Introduction

Mcl-1 is one of the top ten most frequently amplified genes in cancers, and an important factor in resistance to chemotherapeutic agents (Beroukhim et al., 2010; Wei et al., 2008; Wertz et al., 2011). Despite this, no effective Mcl-1 inhibitors have yet been brought to the clinic. Mcl-1 is a member of the family of anti-apoptotic Bcl-2 homologs, which also includes Bcl-x_L, Bcl-2, Bcl-w, Bfl-1, and Bcl-b in humans. These anti-apoptotic Bcl-2 proteins block apoptosis by preventing the pro-apoptotic proteins Bax and Bak from homo-oligomerizing and creating pores in the mitochondrial outer membrane. The anti-apoptotic proteins can either bind directly to Bax and Bak, or bind the related pro-apoptotic BH3-only activator proteins (Bim, Bid, and Puma), preventing activation of Bax and Bak. Other BH3-only proteins, called sensitizers, antagonize anti-apoptotic function by binding competitively with Bax/Bak and activators (Letai et al., 2002).

Synthetic reagents that mimic BH3 sensitizers have high therapeutic potential. For example, the small-molecule BH3 mimetic ABT-263, or Navitoclax, is being tested in humans following extremely promising results in mice (Oltersdorf et al., 2005; Roberts et al., 2012; Rudin et al., 2012; Tse et al., 2008). However, ABT-263 only binds Bcl-2, Bcl-x_L, and Bcl-w, and cancers that also express Bfl-1 or Mcl-1 show resistance to the drug (Konopleva et al., 2006; Yecies et al., 2010). Furthermore, ABT-263 exhibits platelet toxicity, attributed to an inhibitory effect on Bcl-x_L (Roberts et al., 2012; Rudin et al., 2012). Identifying more selective molecules that can target the Bcl-2 family member(s) responsible for blocking apoptosis in a given cancer may be a better strategy for differentially killing cancer cells while avoiding off-target effects. In support of this, ABT-199, a selective Bcl-2 inhibitor, has recently exhibited high activity in chronic lymphocytic leukemia patients, with no off-target killing of platelets (Souers et al., 2013). Several small-molecule Mcl-1 inhibitors have been reported recently (Abulwerdi et al.,

2014; Cohen et al., 2012; Friberg et al., 2013; Kim et al., 2012; Rega et al., 2011; Tanaka et al., 2013). These compounds, some of which display specificity for Mcl-1 over other Bcl-2 family proteins have IC_{50} values in the nanomolar to micromolar range and represent progress towards achieving a tight, Mcl-1-specific inhibitor that can be considered for clinical development.

Another strategy for developing BH3-mimetic molecules is to more directly mimic natural sensitizers by modifying peptides corresponding to the alpha-helical BH3 regions of known Bcl-2 family members. Peptide inhibitors can have higher affinities and specificities than small molecules, due to their larger interaction interfaces, and it is possible to identify peptides with desired properties by screening large genetically encoded libraries. In the clinical diagnostic setting, BH3 peptides are useful in BH3 profiling assays, which assess similarities between the *in vitro* binding profiles and mitochondrial permeabilization patterns of different peptides to determine which anti-apoptotic family member a cancer is dependent upon for resistance to apoptosis (Certo et al., 2006; Deng et al., 2007). There is a need for peptides that have high affinity and specificity for individual Bcl-2 proteins to enhance the diagnostic power of these types of assays. Finally, there are promising precedents in which helical peptides have been modified to render them cell penetrating and highly protease resistant, indicating that peptides should also be considered as candidate therapeutics (Muppidi et al., 2012; Smith et al., 2013; Stewart et al., 2010; Walensky et al., 2004).

Thus far, several Mcl-1-specific peptides have been reported in the literature. Peptides with higher affinity for Mcl-1 (K_d <100 nM) and lower affinities (K_d >1 μM) for the other human Bcl-2 proteins have been selected from Bim BH3-based libraries or generated by targeted mutagenesis (Dutta et al., 2010; Lee et al., 2009a). A promising Mcl-1 specific peptide was derived from the BH3 region of Mcl-1 itself. Mcl-1 BH3 has a K_d of 245 nM for binding to Mcl-

1, but chemical modification by hydrocarbon stapling significantly improved affinity (K_d = 10 nM) (Stewart et al., 2010). The stapled peptide, called Mcl-1 $SAHB_D$, showed no binding to Bcl-2, Bcl-x_L, Bcl-w or Bfl-1 up to 1 μM and sensitized cells to caspase-dependent killing by TRAIL in culture at concentrations of 10-20 μM. A more potent molecule similar to Mcl-1 $SAHB_D$ would have great potential.

We have engineered three peptides, MS1, MS2, and MS3, based on the BH3 region of pro-apoptotic Bim that show high specificity and affinity for Mcl-1. These peptides use novel specificity mechanisms compared to previously designed peptide ligands targeting Mcl-1. The higher affinity Mcl-1 binding exhibited by the peptides makes them better Mcl-1-targeting reagents than the previously engineered Mcl-1-specific peptides or the natural BH3 NoxaA. All three peptides also exhibit high potency and specificity in BH3 profiling assays, demonstrating their usefulness in cellular assays.

Results and Discussion

Selection of Mcl-1-specific peptides

Three Mcl-1-specific peptides were discovered while screening a yeast-surface display library of Bim-BH3 variants. The original purpose of the library screening was to identify peptides that bound selectively to the viral Bcl-2 homolog Kaposi sarcoma herpesvirus Bcl-2 (KSBcl-2). Of the human Bcl-2 proteins, Mcl-1 is most similar to KSBcl-2 (29% sequence identity in the BH3 binding groove, compared to 16% sequence identity in the groove for KSBcl-2 and Bcl-x_L). These two proteins also exhibit similar binding profiles to native BH3-only proteins (Flanagan and Letai, 2008). Therefore, it was not unexpected that peptides that were selected for high-affinity binding to KSBcl-2 also bound to Mcl-1.

Briefly, a library designed to be enriched in KSBcl-2 binders was sorted for binding to KSBcl-2 for two rounds, followed by five rounds of competition sorts in which clones were selected that could bind to KSBcl-2 in the presence of 50- to 100-fold excess Mcl-1 and Bcl-x_L (see Methods). BH3 sequences from two clones (B3 and A12) were chosen for further study as soluble peptides. B3 was chosen because it had the highest frequency of recovery, and A12 was chosen based on the presence of I2dM, an untested substitution, and E2gG, which was shown to provide modest specificity against Bcl-w in previously published Bim substitution SPOT arrays (Table 2.1) (DeBartolo et al., 2012; Dutta et al., 2010; London et al., 2012). Synthetic peptides of 23 amino acids with the sequences of B3 and A12 and an N-terminal fluorescein were made and tested in solution for binding to KSBcl-2 and five human Bcl-2 family proteins (Table 2.4; see Table 2.2 for all sequences). These experiments showed that, although we identified peptides that bound to KSBcl-2 and Mcl-1 in preference to Bcl-2, Bfl-1, Bcl-w, and to a lesser extent Bcl-x_L, we could not discriminate KSBcl-2 from Mcl-1 binding. Because the peptides that we identified

bound tightly to Mcl-1 and showed good specificity for Mcl-1 over other human Bcl-2 family members, we chose to develop them further as Mcl-1 binders.

Table 2.1. Sequences of clones selected from the library

Clone name	Sequence [b]	Number of copies
	abcdefgabcdefgabcdefg	
B3	RPEIWLGQSLQRLGDEINAYYARR	61
G9	RPEIWLGQHLQRLGDEINAYYARR	9
A4	RPEIWLGQSLRRLGDELNAYYARR	9
A1	RPEIWLGQALQRLGDELNAYYARR	5
A2[a]	RPEIWLGQQLIRFGDSLNAYYARR	4
A12	RPEIWMGQGLRRLGDEINAYYARR[b]	1
F6	RPEIWLGQNLQRLGDEINAYYARR	1
G10	RPEIWIGQDLRRLGDEINAYYARR	1

[a] Sequence A2 bound to the antibody reagents used to detect binding in FACS.
[b] Interesting substitutions differing from clone B3, which motivated selection of clone A12 for further characterization, are underlined.

Table 2.2. Sequences of peptides used for fluorescence anisotropy and BH3 profiling assays.[a]

Peptides	Sequence
	2 3 4
	abcdefgabcdefgabcdefg
MS1	RPEIWMTQGLRRLGDEINAYYAR
MS2	RPEIWLTQSLQRLGDEINAYYAR
MS3	RPEIWLTQHLQRLGDEINAYYAR
A12	RPEIWMGQGLRRLGDEINAYYAR
B3	RPEIWLGQSLQRLGDEINAYYAR
G9	RPEIWLGQHLQRLGDEINAYYAR
NoxaA	AELPPEFAAQLRKIGDKVYC
Bim	RPEIWIAQELRRIGDEFNAYYAR
Bim_A2eT	RPEIWITQELRRIGDEFNAYYAR
Bim_A2eT_I2dM	RPEIWMTQELRRIGDEFNAYYAR
Bim_A2eT_E2gG	RPEIWITQGLRRIGDEFNAYYAR
Bim_A2eT_I3dL	RPEIWITQELRRLGDEFNAYYAR
Bim_A2eT_F4aI	RPEIWITQELRRIGDEINAYYAR

[a] The heptad convention used to refer to positions in the BH3 peptide is shown. Bim point mutant positions are underlined.

We sought to improve the Mcl-1 binding selectivity of peptides identified in screening using rational mutagenesis guided by prior studies. Wild-type Bim has an alanine at position 2e (see Table 2.2 for peptide position labels), and SPOT-array tests of Bim BH3 point mutants have shown that glycine at 2e, found in B3, A12, and G9 (a point mutant of B3 that was also identified in screening), is tolerated by or increases binding to all of the receptors. Threonine at 2e was identified using SPOT arrays as a mutation that could decrease binding to Bfl-1, Bcl-x_L, Bcl-2

and Bcl-w, while maintaining strong binding to Mcl-1 (DeBartolo et al., 2012; Dutta et al., 2010; London et al., 2012). The specificity of peptides corresponding to B3, A12 and G9 was greatly improved by replacing the glycine at the 2e position with a threonine, generating the MS1, MS2, and MS3 variants (for Mcl-1 specific), corresponding to the sequences of A12, B3, and G9 with a G2eT mutation, respectively. The notation used to refer to mutants lists the original residue, the peptide position and then the substitution.

MS1, MS2, and MS3 labeled with an N-terminal fluorescein were tested for binding to the five human Bcl-2 receptors in fluorescence anisotropy assays. As shown in Table 2.4 and Figure 2.1, all three peptides bound Mcl-1 with $K_d \leq 2$ nM. MS1 bound with $K_d > 1$ μM to the other four receptors. MS2 bound with micromolar affinity to Bcl-x_L, Bcl-2 and Bfl-1, and bound in the hundred-nanomolar range to Bcl-w. MS3 also displayed micromolar affinity for Bcl-x_L and Bcl-2, and several-hundred nanomolar affinity to Bcl-w and Bfl-1. In contrast, murine NoxaA, for which no binding up to 2500 nM for Bcl-x_L, Bcl-2, Bcl-w, or Bfl-1 is reported in the literature, bound Mcl-1 more weakly than the three designed peptides, with a K_d of 46 nM (Table 2.4) (Certo et al., 2006). NoxaA is the most Mcl-1 selective natural BH3, and a NoxaA BH3 peptide is routinely used in BH3 profiling assays to detect apoptotic resistance dependent upon Mcl-1 (Brunelle et al., 2009; Deng et al., 2007; Ryan et al., 2010). Compared to NoxaA, the three designed peptides MS1, MS2 and MS3 have high affinity for Mcl-1 and also show high specificity against Bcl-x_L, Bcl-2, Bcl-w, and Bfl-1.

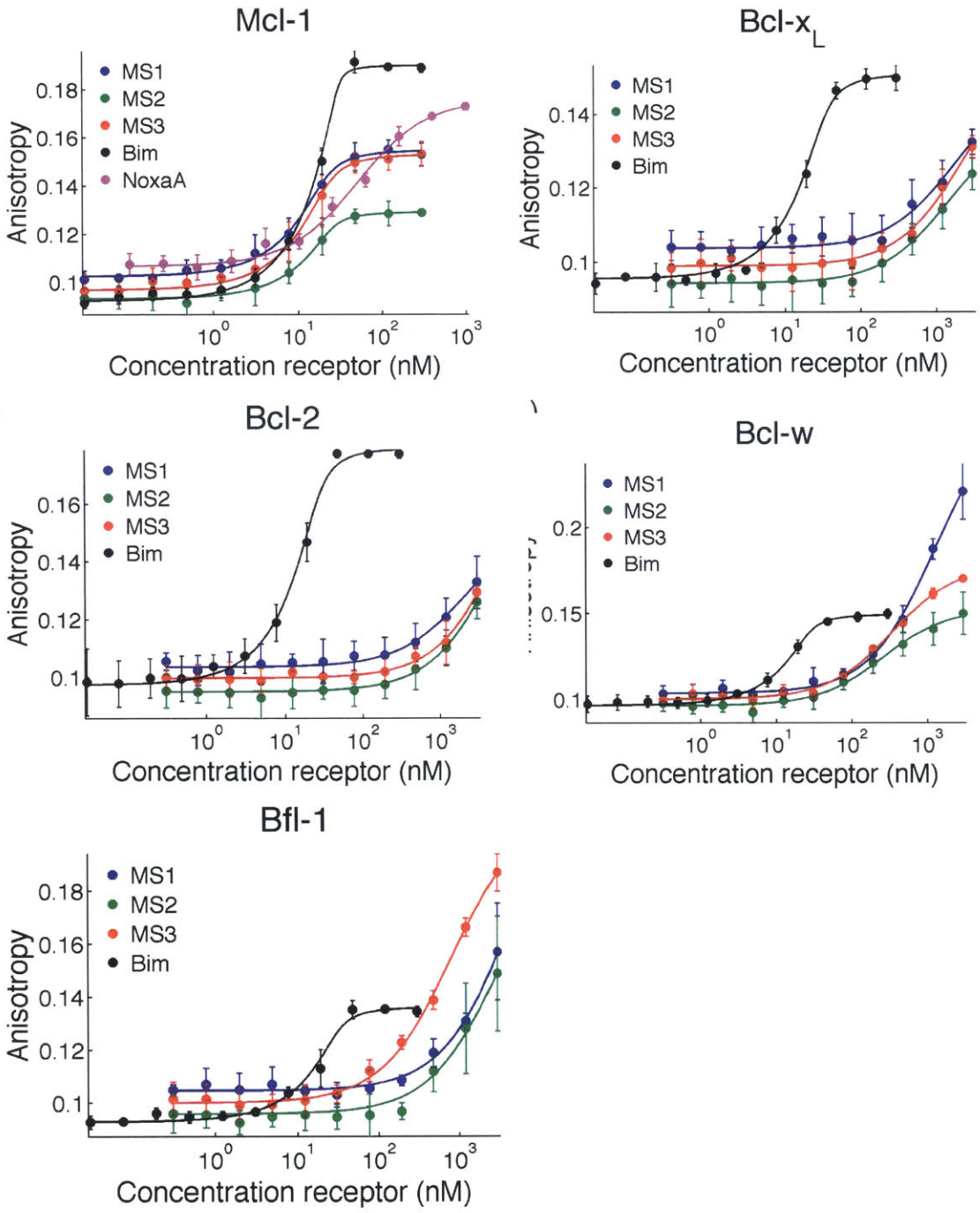


Figure 2.1. Mcl-1-specific peptides labeled with fluorescein binding to five human Bcl-2 homologs. a-e) Binding curves determined using fluorescence anisotropy changes for fluoresceinated peptides binding each Bcl-2 receptor. a) Mcl-1, b) Bcl-x_L, c) Bcl-2, d) Bcl-w, e) Bfl-1. Error bars represent the standard deviation of three replicates performed on different days. The variation in upper baselines presumably originates from different mobilities of the fluorescein dye in different complexes.

To assess the influence of the N-terminal fluorescein dye on binding, we tested a subset of unlabeled peptides, MS1, MS2, Bim, and NoxaA in competition with a fluorescently labeled Bim variant (Figure 2.2). The K_i values for Mcl-1 binding to MS1, MS2, and NoxaA were weaker than the K_d values determined using labeled peptides. The K_i for MS1 binding to Mcl-1 was between 8 and 24 nM, depending on the fitting protocol (see Methods and Table 2.3). The competition experiments indicated that MS1 is between ~40 and 190-fold specific for Mcl-1 over Bcl-w, the next-tightest binding family member. Competition experiments also confirmed that MS1 and MS2 are considerably tighter binders to Mcl-1 than is NoxaA; NoxaA binding to Mcl-1 was very weak and thus difficult to quantify with the competition assay.

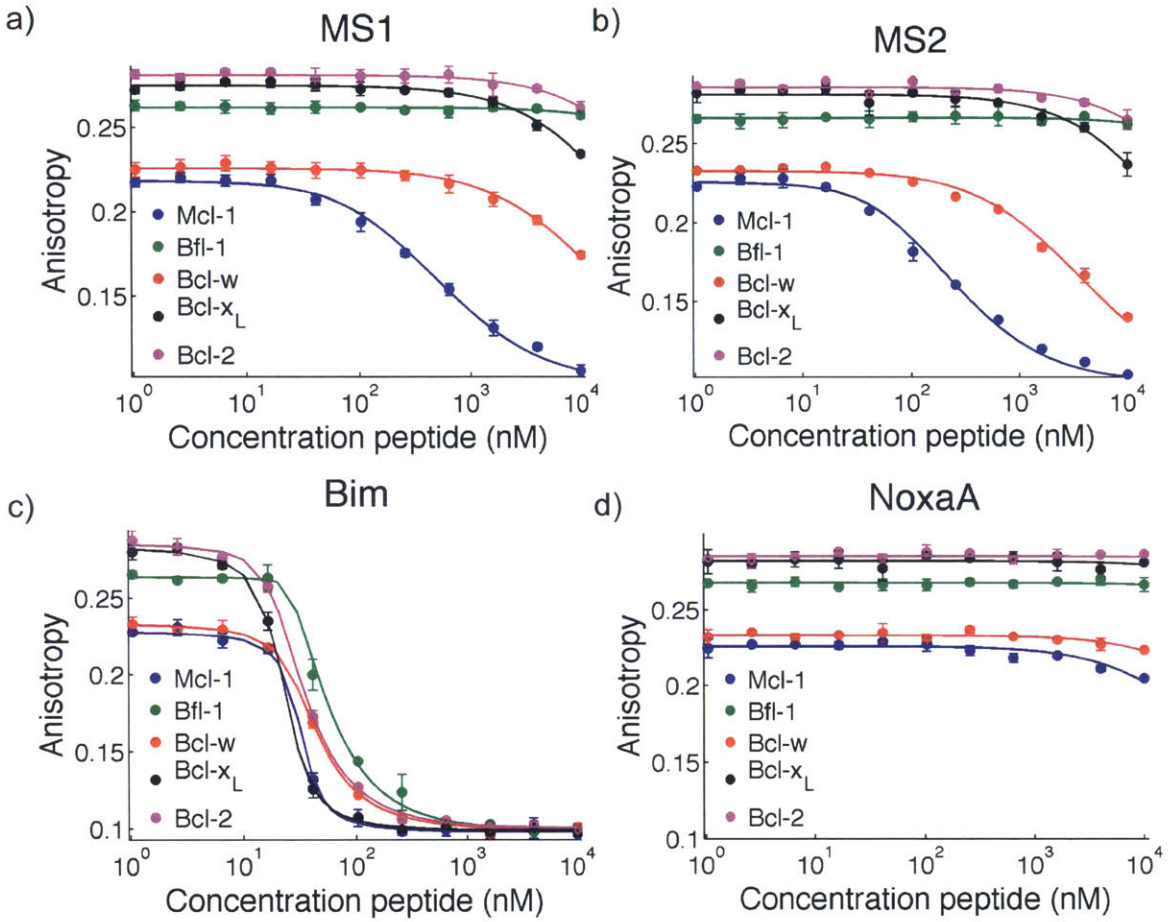


Figure 2.2. Competition fluorescence anisotropy binding experiments. Unlabeled peptides were mixed with fluoresceinated Bim BH3 and one of Mcl-1, Bfl-1, Bcl-w, Bcl-x_L or Bcl-2. a) MS1, b) MS2, c) NoxaA, d) Bim. Error bars are standard deviation of two replicates for MS2, NoxaA, and Bim, and at least 3 replicates for MS1. These curves represent the upper limit on the K_i, corresponding to the top row values in Table 2.3.

Table 2.3. K_i values (nM) for competition fluorescence anisotropy assays

	Receptor				
Peptide	Mcl-1	Bfl-1	Bcl-w	Bcl-x_L	Bcl-2
MS1[a]	24 (3.1)	N.B.[b]	990 (145)	1540 (820)	~ 2100[d]
	7.9 (0.68)	N.B.	1500 (190)	2400 (1700)	~2900
MS2	7.5 (1.4)	N.B.	230 (32)	~ 1300	~2000
	3.4 (0.31)	N.B.	360 (73)	~ 2100	~2700
Bim	< 1	0.089 (0.018)	0.62 (0.15)	< 1	0.14 (0.049)
	<1	<1	0.98 (0.25)	<1	0.19 (0.07)
NoxaA	~ 1700	N.B	~ 9900	N.B	N.B
	~ 670	N.B.	~14000	N.B.	N.B.
F-Bim[d]	1.9 (0.85)	0.16 (0.14)	2.7 (1.0)	1.4 (0.71)	0.91 (0.28)
	0.55 (0.82)	0.093 (0.32)	4.8 (1.7)	2.7 (1.2)	1.3 (0.52)

[a] The first row of K_i values listed for each peptide is from a fit using the top-row K_d values for fluoresceinated Bim (F-Bim) (resulting from a fit in which the F-Bim concentration was fixed); the second row lists K_i values fit using the bottom-row F-Bim K_d (resulting from a fit where the F-Bim K_d concentration was fit). All values are given with the 95% confidence interval in parentheses.

[b] N.B., no binding observed up to 10 μM peptide.

[c] Values reported as approximate had incomplete curves (see Figure 2.2), and therefore a high degree of uncertainty.

[d] F-Bim 18-mer $\mathbf{K_d}$ values, with the top row determined by fixing the concentration of F-Bim (possibly less accurate), and the bottom row determined by fitting the concentration (more accurate, but high degree of uncertainty given the many parameters fit).

Table 2.4. Affinities of native and designed BH3 peptides for six Bcl-2 homologs.

	Receptor					
Peptide	Mcl-1	Bfl-1	Bcl-w	Bcl-x_L	Bcl-2	KSBcl-2
MS1	1.9 ± 1.0[a]	5000 ± 3200	1300 ± 230	1600 ± 1000	2300 ± 1500	2.9 ± 0.68
MS2	1.5 ± 1.2	3100 ± 2300	250 ± 76	1400 ± 500	6200 ± 4100	< 1
MS3	2.0 ± 1.2	790 ± 140	340 ± 69	2300 ± 1000	> 3000	3.3 ± 1.6
A12	2.4 ± 2.3	22 ± 6.6	210 ± 110	9.8 ± 3.1	42 ± 8.9	4.0 ± 2.6
B3	< 1	26 ± 7.7	16 ± 6.2	4.2 ± 2.1	35 ± 11	1.43 ± 1.39
NoxaA	46 ± 11	N.D.	N.D.	N.D.	N.D.	N.D.
Bim 23-mer	< 1	< 1	1.75 ± 1.0	2.6 ± 1.9	1.9 ± 1.3	1.2 ± 0.79
Bim_ A2eT	< 1	31 ± 6.8	39 ± 9.4	17 ± 4.8	43 ± 12	1.8 ± 0.81
Bim_ A2eT_ I2dM	< 1	150 ± 69	260 ± 52	83 ± 53	210 ± 71	0.75 ± 0.37

Bim_ A2eT_ E2gG	1.6 ± 0.57	250 ± 63	94 ± 36	37 ± 7.0	150 ± 27	0.66 ± 0.31
Bim_ A2eT_ I3dL	<1	120 ± 27	12 ± 2.6	7.6 ± 3.8	37 ± 14	1.0 ± 0.38
Bim_ A2eT_ F4aI	<1	7.9 ± 2.1	16 ± 4.1	110 ± 62	150 ± 43	1.7 ± 0.80

[a] Dissociation constants for direct binding of fluoresceinated peptides to Bcl-2 proteins (in nM) with 95% confidence intervals. Values designated < 1 were too tight to be accurately fit. Values designated > 3000 were too weak to be accurately fit. N.D., not determined. See Table 1 for sequences of all peptides used. Binding data and fits are shown in Figure 2.1.

Specificity mechanisms

A peptide corresponding to the BH3 region of Bim binds very tightly to all receptors. To better understand the determinants of binding specificity for MS1, MS2 and MS3, we sought to identify residues in these peptides that differ from Bim and destabilize interactions with receptors other than Mcl-1. The 2eT mutation was vital in generating highly Mcl-1-specific peptides. This single point mutation in Bim (giving Bim_A2eT, Table 2.2) provides a 6-fold reduction in Bcl-x_L binding and over 20-fold reduction in Bcl-2, Bcl-w, and Bfl-1 binding in a peptide with the wild-type Bim background (Table 2.4). Likewise, introducing 2eT into library peptides A12 and B3 reduced binding to Bfl-1, Bcl-2, and Bcl-x_L ~100-fold, and gave a more moderate ~10-fold reduction in Bcl-w binding affinity. Thus, it is clear that threonine at position 2e is highly destabilizing for all human Bcl-2 receptors other than Mcl-1, in several different peptide contexts.

Position 2e is conserved as small (alanine, glycine, serine) in natural BH3 sequences. Mcl-1 can bind BH3 peptides with larger residues at position 2e, including Bim _A2eT and a peptide corresponding to the Mcl-1 BH3 region, which has a leucine at position 2e (Stewart et al., 2010). To look for possible reasons that the other Bcl-2 paralogs cannot accommodate threonine at position 2e, we compared structures of Bcl-x_L, Bfl-1, and Mcl-1 bound to the BH3 region of Bim (3FDL, 2VM6, and 2PQK, respectively) (Fire et al., 2010; Herman et al., 2008; Lee et al., 2009b). As shown in Figure 2.3a, helix 4 of Bfl-1 is closer to the peptide near the 2e position than is helix 4 in Mcl-1. Simple modeling of preferred threonine rotamers at 2e on static Mcl-1:Bim BH3 and Bfl-1:Bim BH3 structures illustrates that threonine is easily accommodated in the Mcl-1 structure in a helix-preferred rotomer, but clashes with helix-4 residues of Bfl-1 (Figure 2.3b) (Dunbrack and Cohen, 1997; Fire et al., 2010). In the Bcl-x_L:Bim BH3 structure,

the BH3 peptide is positioned slightly differently in the groove, resulting in Ala2e being oriented further into the groove than in the Mcl-1:Bim BH3 structure (Figure 2.3c). Rearrangement of Bcl-x_L helix 4 would likely be required to accommodate threonine, and such rearrangement could be disfavored if it led to disruption of a three-residue salt-bridge network that forms in the Bcl-x_L:Bim BH3 structure between Glu129 and Arg132 of Bcl-x_L (on helix 4) and Arg3b of Bim (Figure 2.3d). This network cannot be formed in a structure like that of Mcl-1:Bim BH3 or Mcl-1:Mcl-1 BH3 (in which position 2e is leucine), because the equivalent of Bcl-x_L residue 129 is farther away from peptide position 3b in this complex. The charged residues in Bcl-x_L that participate in salt-bridge formation are also conserved in Bcl-2, suggesting that a similar mechanism might operate to disfavor threonine or larger residues at 2e for that protein.

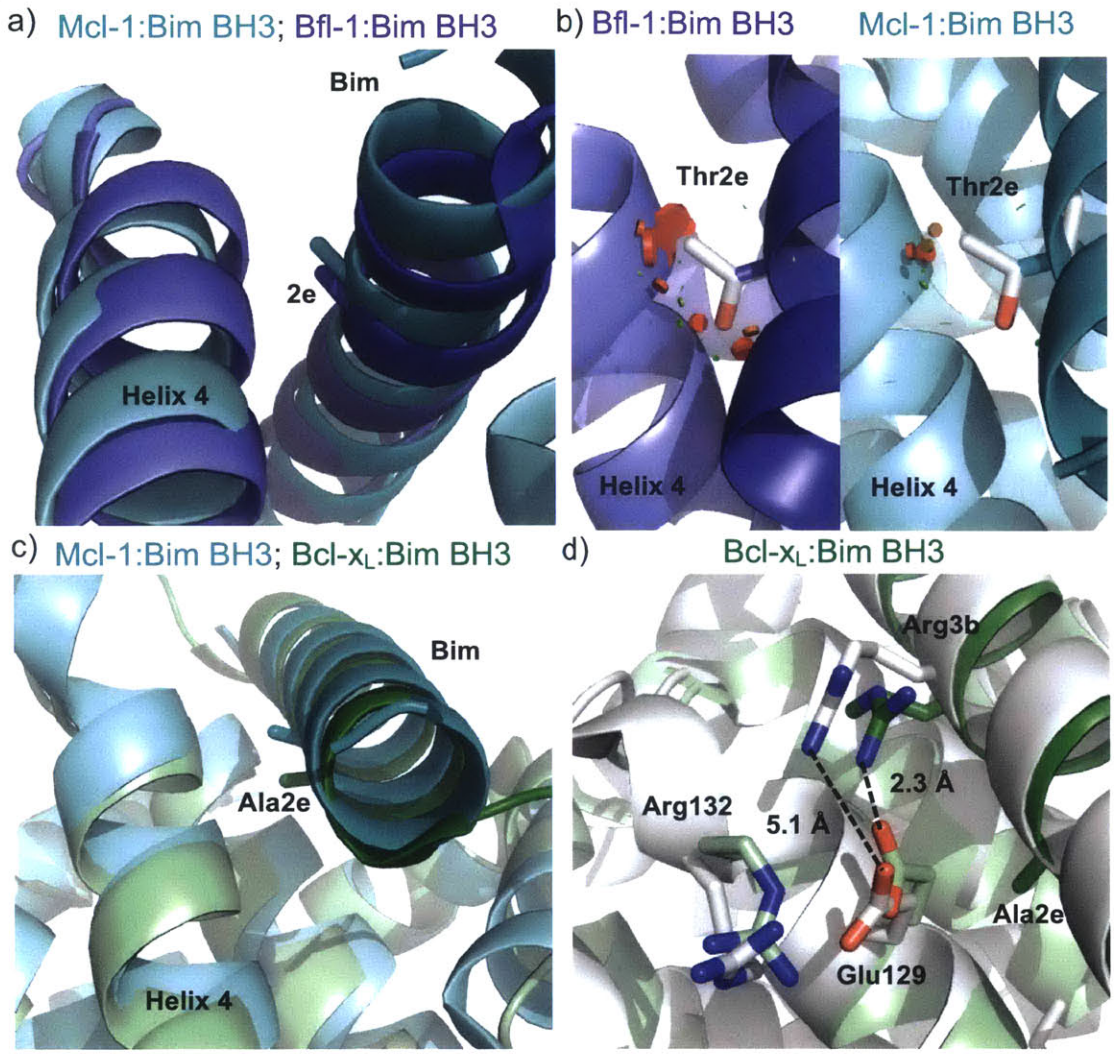


Figure 2.3. Comparison of Bim BH3 position 2e in structures of Bfl-1, Bcl-x_L, and Mcl-1. a) Helix 4 is closer to Ala2e in the Bfl-1:Bim BH3 structure (light:dark purple, 2VM6) than in the Mcl-1:Bim BH3 structure (light:dark cyan, 2PQK), as shown by aligning the structures in PyMOL.(Fire et al., 2010; Herman et al., 2008) b) Threonine modeled at position 2e clashes (red circles) with helix 4 in the Bfl-1 complex (left), but does not clash significantly in the Mcl-1 complex (right); mutations were modeled in PyMOL by selecting the most preferred backbone-dependent rotamer for Thr; all rotamers are predicted to clash on the Bfl-1 structure backbone). c) Ala2e is angled further into the BH3 binding groove in the Bcl-x_L:Bim BH3 structure (light:dark green, 3FDL) than in the Mcl-1:Bim BH3 structure (light:dark cyan, 2PQK).(Lee et al., 2009b) d) A salt bridge network that exists in the Bcl-x_L:Bim BH3 structure (green) between Glu129, Arg132, and Arg3b would be disrupted at the equivalent sites in Mcl-1:Bim BH3 structure 2PQK (white). Glutamate and arginine are shown modeled in the place of His252 and Ser255 in the 2PQK structure, with rotamers chosen to position the side chains as close as possible to the orientations in the 3FDL structure.

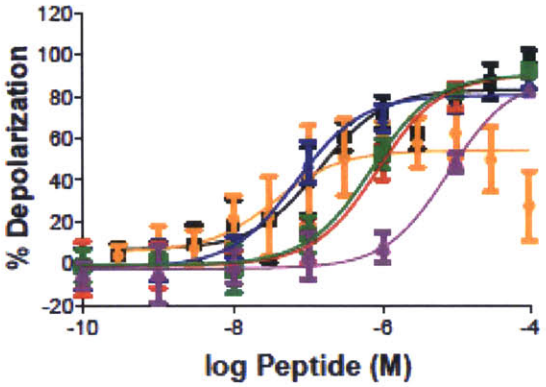
MS1, MS2, and MS3 all have different substitutions at 2g, which is a glutamate in wild-type Bim and is typically a medium-to-large residue in other known BH3 regions. MS1, our most selective peptide, has a glycine at this position, and mutating glutamate to glycine at position 2g in Bim_A2eT decreased binding to all receptors. The change in affinity for Mcl-1 could not be quantified, but affinities for Bfl-1, Bcl-2, Bcl-w, and Bcl-x_L were reduced an additional 2-8 fold compared to Bim_A2eT. Thus, glycine at 2g provides some of the negative design disfavoring interactions with off-target receptors, although at the cost of weakening binding to the Mcl-1 target (Table 2.4).

Three mutations in peptides MS1, MS2, and MS3 occur in positions that are usually conserved as hydrophobic in known BH3 motifs (positions 2d, 3d and 4a). When tested in the Bim_A2eT context, I2dM (found in MS1) provided a roughly 4-fold reduction in binding to Bcl-x_L, Bcl-2, Bcl-w, and Bfl-1 (Table 2.4). Notably, significant decreases in Bcl-w binding for Bim_A2eT_E2gG and Bim_A2eT_I2dM may explain why MS1 is more selective for Mcl-1 *vs.* Bcl-w than are MS2 and MS3, which have different mutations at 2g and 2d. Mutation I3dL reduced binding of Bim_A2eT to Bfl-1 by 4-fold, while this mutation increased binding slightly to Bcl-x_L, Bcl-2, and Bcl-w (Table 2.4). The F4aI mutation increased Bim_A2eT binding slightly to Bfl-1 and Bcl-w, but decreased binding by 6-fold and 3-fold to Bcl-x_L and Bcl-2, respectively (Table 2.4). Position 4a is a well-documented source of specificity for Mcl-1 binding (Dutta et al., 2010; Fire et al., 2010; Lee et al., 2008; Stewart et al., 2010). Mutagenesis studies and peptide library screens have demonstrated that Bcl-x_L binds preferentially to peptides that include a phenylalanine or tyrosine to fill the enclosed hydrophobic pocket near 4a, whereas Mcl-1 tolerates a wide variety of substitutions at this position (Dutta et al., 2010; Lee et al., 2008; Stewart et al., 2010).

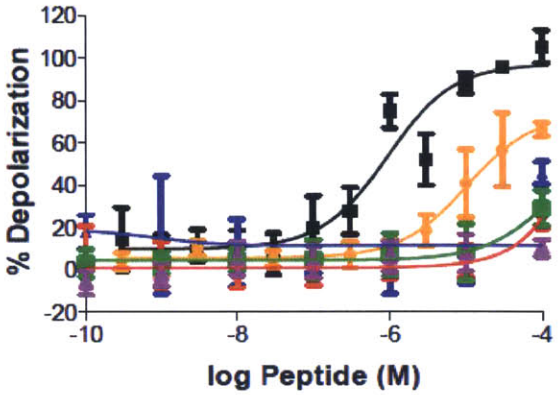
Cellular BH3 profiling assays

A whole-cell BH3 profiling assay was used to test the specificity of our Mcl-1-binding peptides in several cell lines with differing dependencies on Bcl-2, Mcl-1, Bcl-x_L, or Bfl-1 (Ryan et al., 2010; Ryan and Letai, 2013). In this assay, permeabilized cells were treated with increasing doses of BH3 peptides, and mitochondrial outer membrane permeabilization (MOMP) was monitored using the dye JC-1 (see Methods). EC_{50} values for BH3 profiling experiments involving peptides from this study are given in Figure 2.5 and Table 2.5, with full titration curves shown in Figure 2.4.

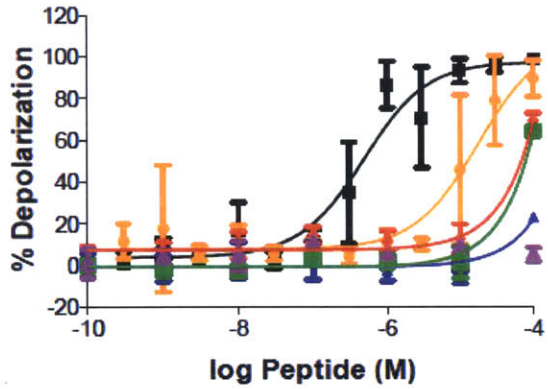
a) Mcl-1 2640 (Mcl-1 Dependent)



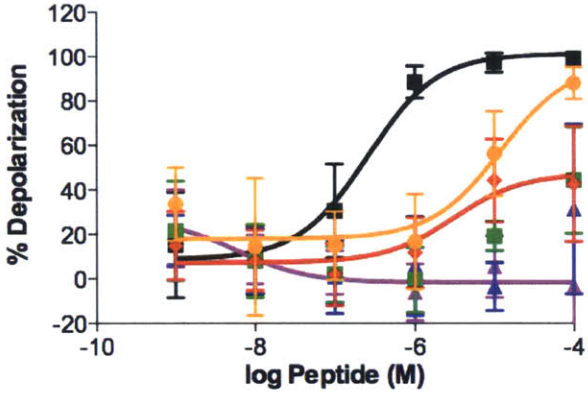
b) Bcl-2 2924 (Bcl-2 Dependent)



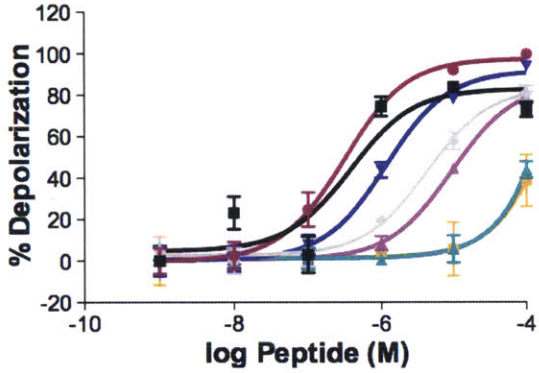
c) MDA-MB-231 (Bcl-x_L Dependent)



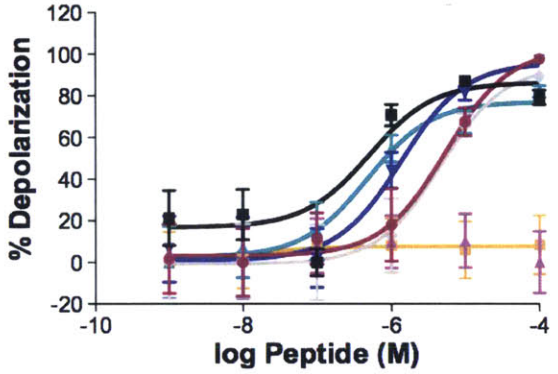
d) Pfeiffer (Bfl-1 dependent)



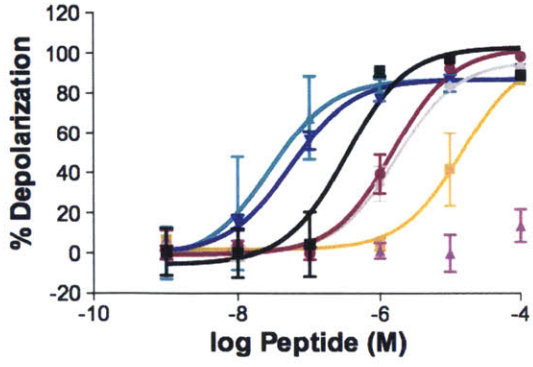
e) Mcl-1 2640 (Mcl-1 dependent)



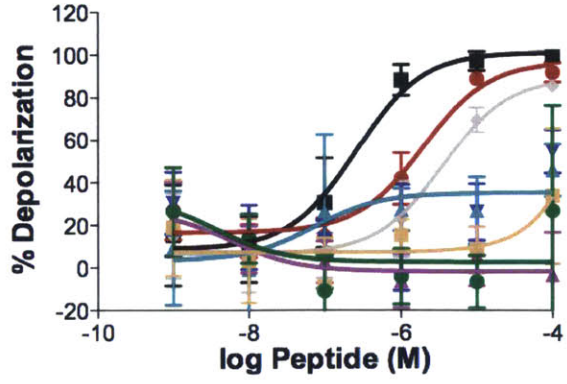
f) Bcl-2 2924 (Bcl-2 Dependent)



g) MDA-MB-231 (Bcl-x_L Dependent)



h) Pfeiffer (Bfl-1 dependent)



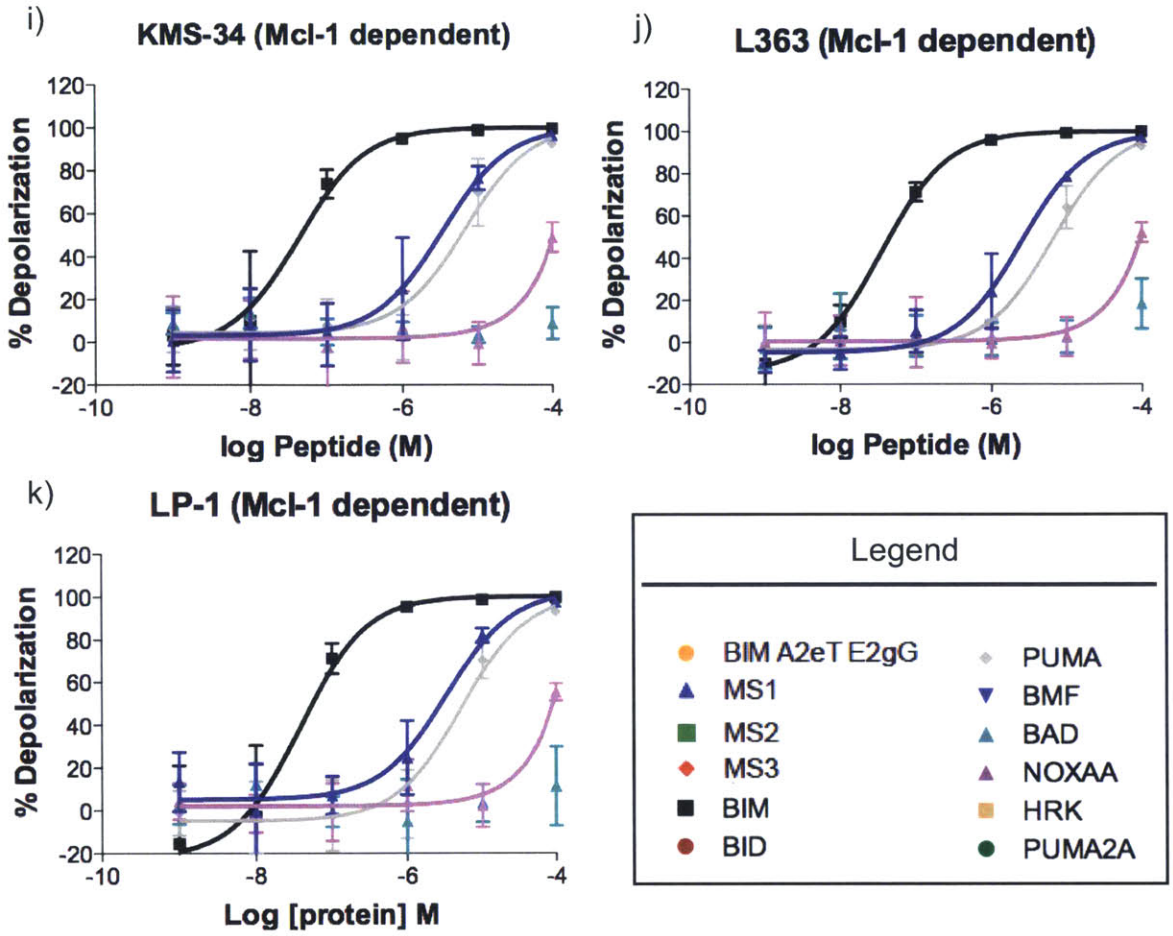


Figure 2.4. BH3 profiling of cell lines using engineered and native BH3 peptides. a-d) Mitochondrial depolarization response caused by the engineered peptides MS1, MS2, MS3, and Bim_A2eT_E2gG, with Bim and NoxaA shown for comparison. e-h) Mitochondrial depolarization response caused by native BH3 peptides. i-k) Mitochondrial depolarization response of three human, Mcl-1 dependent cell lines to MS1 and a subset of the native BH3 peptides. Error bars indicate the standard deviation over 3 or more replicates performed on 3 or more days.

Table 2.5. EC_{50} values (μM) for depolarization response [a]

Peptide	Dependency (cell line)						
	Mcl-1 (Mcl-1 /Myc 2640)	Bcl-2 (Bcl-2 /Myc 2924)	Bcl-xL (MDA-MB-231)	Bfl-1 (Pfeiffer)	Mcl-1 (KMS-34)	Mcl-1 (L363)	Mcl-1 (LP-1)
MS1	0.070 ± 0.021	>100[b]	>100	>100	3.3 ± 1.7	2.5 ± 0.83	3.3 ± 1.6
MS2	0.70 ± 0.18	>100	>100	>100	N.D.	N.D.	N.D.
MS3	0.86 ± 0.26	>100	>100	3.1 ± 2.6	N.D.	N.D.	N.D.
Bim	0.14 ± 0.068	0.94 ± 0.45	0.48 ± 0.21	0.26 ± 0.11	0.042 ± 0.020	0.035 ± 0.0095	0.038 ± 0.020
Bim_ A2eT_ E2gG	0.036 ± 0.026	11 ± 5.0	18 ± 11	12 ± 8.0	N.D.	N.D.	N.D.
NoxaA	20 ± 14	>100	>100	>100	>100	>100	>100
Puma	15 ± 10	5.3 ± 4.4	4.1 ± 2.6	3.5 ± 1.7	6.7 ± 3.2	6.6 ± 2.0	5.2 ± 2.3
Bad	>100	0.96 ± 0.57	0.020 ± 0.0091	0.073 ± 0.068	>100	>100	>100
Puma2A	>100	>100	>100	>100	N.D.	N.D.	N.D.

[a] Values are given in μM with 95% confidence intervals. N.D., not determined.
[b] Values designated as >100 were estimated by fixing the upper baseline as 100% depolarization.

Mcl-1/Myc 2640 is an engineered murine leukemia cell line overexpressing murine Mcl-1 and Myc, and Bcl-2/Myc 2924 is a similarly engineered cell line overexpressing human Bcl-2 (Brunelle et al., 2009). By Western blot and BH3 profiling, these cells exhibit Mcl-1 and Bcl-2 dependencies, respectively (Brunelle et al., 2009). MS1, MS2, and MS3 elicited potent mitochondrial depolarization responses in Mcl-1/Myc 2640, with EC_{50} values of 70 nM, 700 nM, and 860 nM, respectively. These peptides were much more potent than NoxaA in this assay (EC_{50} = 20 mM). Human and murine Mcl-1 are over 90% identical in the Bcl-2 domain, and

94% identical in the BH3 binding groove. Human multiple myeloma cell lines dependent upon Mcl-1 (as indicated by response to NoxaA and Bad) gave EC_{50} values of 2.5-3.3 mM for MS1, compared to EC_{50} values > 100 mM for NoxaA (Figure 2.4i-k and Table 2.5). Thus, multiple Mcl-1 dependent cell lines were much more sensitive to MS1 than to NoxaA.

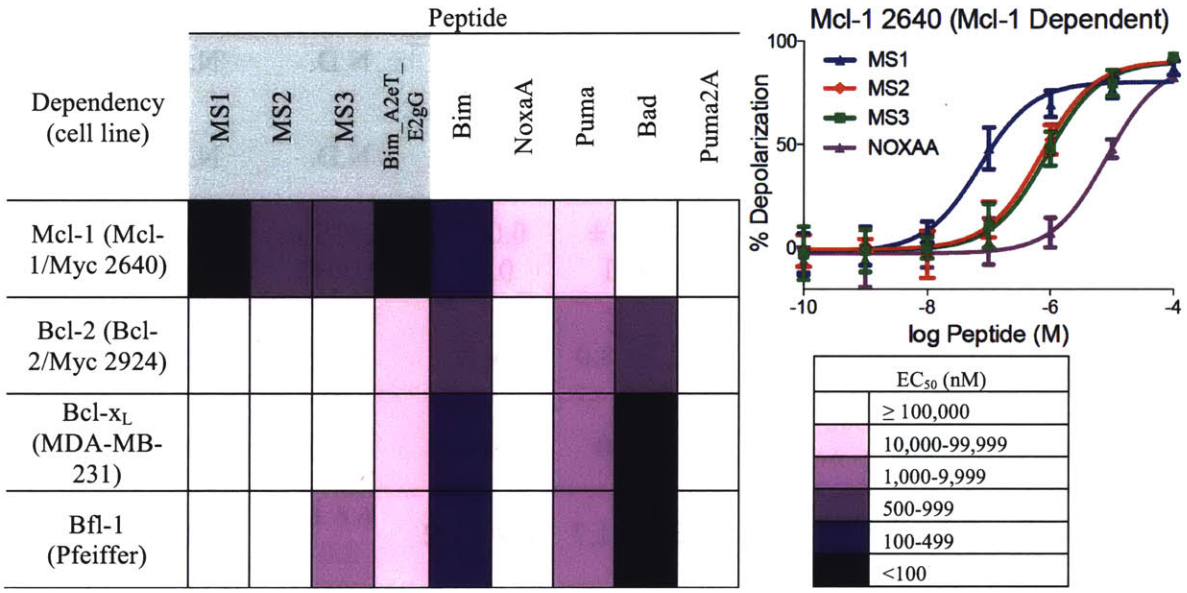


Figure 2.5. Heat map of the EC_{50} values (peptde concentration in nM) for mitochondrial depolarization induced by engineered and native BH3 peptides in four cell lines. Engineered peptide names are shaded in gray. Each cell line is dependent upon a single Bcl-2 family member: Mcl-1, Bcl-2, Bcl-x_L, or Bfl-1 as determined by BH3 profiling with native BH3-only peptides (Figure 2.4e-h). All experiments were performed at least three times. The inset shows an example titration curve; see Figure 2.4 and Table 2.5 for other curves and all EC_{50} values with 95% confidence intervals.

MS1 and MS2 were highly selective in BH3 profiling. In a Bcl-2 dependent line, EC_{50} values were >100 μM for MS1, MS2, MS3 and NoxaA. MDA-MB-231 is a human breast cancer cell line that has been shown to have a Bcl-x_L-dependent profile (Ryan et al., 2010). EC_{50} values for MS1, MS2, MS3 and NoxaA were over 100 μM for MDA-MB-231 cells. MS1 and MS2 showed EC_{50} values >100 μM in Pfeiffer, a lymphoma line with high Bfl-1 mRNA expression that has previously been shown to exhibit a Bfl-1 dependent BH3 profile (Deng et al., 2007).

MS3 gave a stronger response in Pfeiffer than MS1 or MS2, but MS3 also exhibited tighter Bfl-1 binding by fluorescence anisotropy (Table 2.4). Finally, Bim_A2eT_E2gG, which showed modest specificity for Mcl-1 by fluorescence anisotropy (Table 2.4), exhibited a strong depolarization response in Mcl-1/Myc 2640 and a depolarization response intermediate to that of Bim and MS1 in Bcl-2/Myc 2924, MDA-MB-231, and Pfeiffer. Thus, *in vitro* binding specificities are replicated in BH3 profiling assays in cell lines showing all of the currently identified dependencies on Bcl-x_L, Bcl-2, Mcl-1, and Bfl-1, as a Bcl-w-dependent cell line has not yet been identified or constructed.

The engineered peptides tested here were derived from Bim BH3, which is an activator BH3 peptide. Nevertheless, these peptides do not cause strong depolarization in cell lines that are not dependent upon Mcl-1, indicating that they act as sensitizers rather than activators in these assays. Depolarization activity was specific to the Bcl-2 pathway, because the peptides did not depolarize mitochondria in the Bax/Bak deficient cell line Su-DHL10 (Figure 2.6) (Deng et al., 2007).

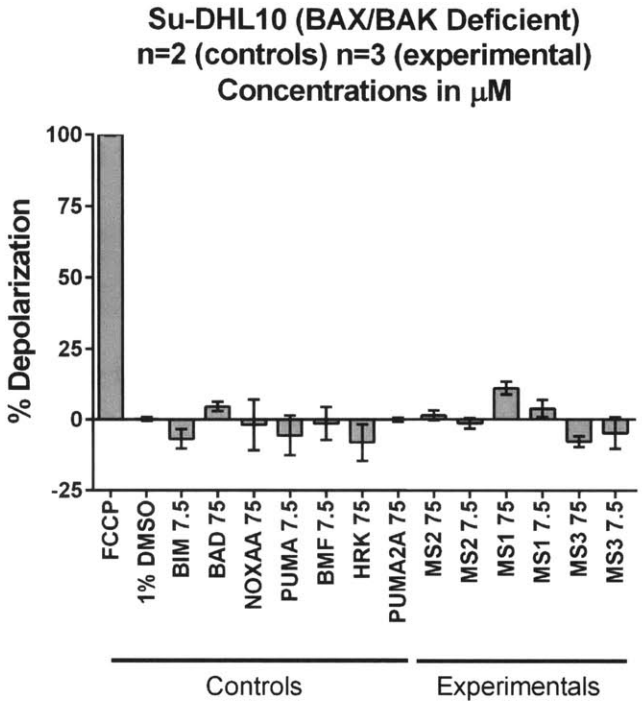


Figure 2.6. Engineered peptides do not cause mitochondrial membrane depolarization in Bax/Bak negative cells (Su-DHL10), showing that their method of action is dependent upon the Bax/Bak pathway. MS1, MS2, and MS3 were tested at 7.5 and 75 μM (concentration shown next to peptide name). Native BH3 peptides are shown for comparison. FCCP is a chemical uncoupler of oxidative phosphorylation that causes mitochondrial depolarization.

The three peptides presented here exhibit high affinity for Mcl-1 in solution binding assays. MS1, the most specific peptide, was at least 40-fold specific over Bcl-x_L, Bcl-2, Bcl-w, and Bfl-1 as assessed using biochemical competition binding experiments. We demonstrated the potency and specificity of these peptides as Mcl-1-specific reagents in BH3 profiling assays. Given that Mcl-1 overexpression is a key source of resistance to ABT-263, the ability to detect Mcl-1 activity will be important in designing treatment strategies (Konopleva et al., 2006; Yecies et al., 2010). The high affinity of the peptides for Mcl-1 also presents a good starting point for developing peptide-based therapeutics.

Methods

Library construction and sorting

A library of Bim BH3 domain variants was designed to be enriched in KSBcl-2-specific binders based on Bim BH3 substitution SPOT arrays (KSBcl-2 SPOT array to be published) (DeBartolo et al., 2012; Dutta et al., 2010; London et al., 2012). The library contained 4×10^5 protein variants of the Bim BH3 peptide with the following composition at each position (heptad register is defined in Figure 1): (AILPTV) at 2d, (AGPRST) at 2e, (ADEGHKNPQRST) at 2g, (FIST) at 3b, (FIST) at 3d, (NQGHKDERS) at 3g, (ADEFIKLMNSTVYStop) at 4a. Position 3b was intended to sample glutamine and arginine but was encoded as FIST instead due to an error in library construction. Only arginine and glutamine were selected in the screen at 3b, which is consistent with the wild-type residue of Bim being arginine. Interestingly, sequencing of expression-positive clones prior to screening showed that glutamine and arginine were encoded at 3b at low frequency. Other mutations not encoded in the theoretical library (e.g., M2d and L3d) that appeared in selected clones were also observed when sequencing expression-positive clones. Thus, deviations from the intended library appear to have been introduced early in the experimental protocol, likely during DNA synthesis, and favorable mutations were enriched during the highly stringent sorting protocol.

The library was constructed by PCR using a mutagenic forward primer. The forward primer encoding the library variants was 5' GGCCGTCCGGAAATTTGG VYY VSY CAG VVW CTA WYY CGT WYY GGC GAT VRK DHS AATGCGTATTATGCGCGTCGC 3' where W represents a mixture of A and T; K represents a mixture of G and T; D represents a mixture of A, G and T; V represents a mixture of A, C and G; H represents a mixture of A, C and T; R represents a mixture of A and G; Y represents a mixture of C and T; S represents a mixture

of C and G. The library was constructed as a fusion to Aga2p in the pCTCON2 plasmid and transformed into yeast strain EBY100 as previously reported (Dutta et al., 2010).

The library was sorted by FACS over seven rounds to identify clones that bound specifically to c-myc-tagged KSBcl-2 in preference to untagged Mcl-1 and His_6-Bcl-x_L. The first sort selected the top 3% of cells binding to 100 nM c-myc-KSBcl-2; the second sort selected the top 5% of cells binding to 10 nM c-myc-KSBcl-2. Sorts 3-5 collected the top 0.5-6% of cells binding to 10 nM c-myc-KSBcl-2 in the presence of 500 nM unlabeled Mcl-1 and 500 nM His_6-Bcl-x_L. Sorts 6 and 7 collected the top 0.3% and 1.0%, respectively, of cells binding to 10 nM c-myc-KSBcl-2 with 1 μM unlabeled Mcl-1 and 1 μM His_6-Bcl-x_L. The c-myc-tagged KSBcl-2 construct was constructed from the C-terminally truncated KSBcl-2 construct of Flanagan and Letai (residues 1-150 of KSBcl-2, with the additional non-native residues GRIVTD at the C-terminus) cloned into the pSVM vector and purified as described for other c-myc-tagged Bcl-2 constructs (DeBartolo et al., 2012; Dutta et al., 2010; Flanagan and Letai, 2008; London et al., 2012). Unlabeled human Mcl-1 (172-327) in the pSV282 vector was purified as described (Dutta et al., 2010; Fu et al., 2007). His_6-tagged Bcl-x_L was purified as described (Chen et al., 2013). Library cells were grown overnight in galactose-containing media (SG+CAA). Enough cells to oversample the library diversity by at least 10-fold on average were split into aliquots of ~10^6 cells/well in Millipore MultiScreen 0.45 μm-pore filter plates and incubated with 100 μL pre-mixed Bcl-2 receptors for 1 hour at room temperature in 50 mM Tris pH 8.0, 100 mM NaCl (TBS). Wells were washed once with 100 μL TBS followed by 100 μL TBS plus 1 mg/mL bovine serum albumin (BSS). Cells were incubated with 20 μL primary antibody mix (1:100 dilution in BSS) for 15 minutes at 4 °C and washed twice with 100 μL BSS. The antibody step and washes were repeated with the secondary antibodies. Cells were resuspended in BSS for

sorting on a BD FacsAria or a Beckman Coulter MoFlo and collected in SD+CAA media; sorted cells recovered at 30 °C for ~48 hours. For the first 6 rounds of sorting, a rabbit anti-c-myc primary antibody (Sigma) and anti-rabbit-PE secondary antibody (Sigma) were used to detect binding, and a mouse anti-FLAG primary (Sigma) and anti-mouse-APC secondary (BD Pharmigen) were used to detect a FLAG tag expressed on the yeast surface display construct for expression control. To minimize selection of antibody binders, the seventh sort used rabbit anti-FLAG, mouse anti-c-myc, anti-rabbit-FITC, and anti-mouse-PE antibodies (all from Sigma).

Fluorescence anisotropy assays

The Bim variant peptides were synthesized by the MIT Biopolymers Laboratory and were 23-mers as shown in Table 2.2 with N-terminal 5/6-fluorescein amidite and C-terminal amidation. The crude synthesis product was validated to contain primarily the peptide of interest by mass spectrometry and purified by HPLC on a C18 column using a linear gradient of acetonitrile. FITC-AHA-NOXAA peptide was produced by the Tufts University Physiology core facility, purified by HPLC, and validated by mass spectrometry. Bcl-2 receptors were c-myc-tagged variants without the C-terminal trans-membrane region, and without the N-terminal domain of Mcl-1, expressed and purified as in Dutta et al. 2010 (Dutta et al., 2010). Direct fluorescence anisotropy assays were performed in a buffer of 25 mM Tris pH 7.8, 50 mM NaCl, 1 mM EDTA, 0.001% Triton-X-100, 5% v/v DMSO. A twelve-point dilution of receptor protein was added to the wells of a Corning 96-well, black, polystyrene, non-binding surface plate, followed by the fluoresceinated peptide for a total volume of 120 μL with a final peptide concentration of 10 nM. Plates were incubated for 2 hours before being read on a SpectraMax M5 plate reader (Molecular Devices) at 25 °C. Equilibration was checked by comparing curves read at 2 hours and 24 hours. Most curves exhibited no change so the 2-hour data were used. The

exceptions were the Bfl-1 curves for peptides MS3 and MS2 for which the 24-hour curves were used.

Fitting protocol for fluorescence anisotropy curves

Fluorescence anisotropy data for direct binding experiments were fit to the direct binding model below (derived from Roehrl et al., 2004) using Matlab, where l is the lower baseline, u is the upper baseline, c is the concentration of the peptide, and x is the concentration of the receptor.

$$Anisotropy = l + \frac{(u-l)}{2c}\left(c + x + K_D - \sqrt{(c + x + K_D)^2 - 4cK_D}\right)$$

The lower baseline was fixed as the average of the anisotropy from the two lowest receptor concentrations for each curve. To determine a value for the concentration of the peptide, the parameters c, u, and K_d were fit for the curves with lower and upper baselines, and then the average concentration of the peptide fit for those curves was used as the concentration for all receptor curves for that peptide. This was done because the raw fluorescence values observed when collecting the data appeared to vary over a 2-fold range. Small variation in peptide concentration is an inevitable product of setting up these assays, and fixing the concentration at 10 nM gave poorer fit quality. Curves were fit from averaged data from at least three replicates done over three days, and the error bars represent the standard deviation from these measurements. 95% confidence intervals were found using the confint function in Matlab. As expected, curves with no upper baselines had larger confidence intervals. For very weak curves, K_d was listed as >3000, which was the highest receptor concentration used. Likewise, for very tight curves for which we could not get an accurate fit, we reported the value as <1, which is the approximate limit of what we can fit.

Competition fluorescence anisotropy assays were performed by titrating 23-mer unlabeled peptides over a concentration range of 0-10 μM. Unlabeled peptides (synthesized and purified as described for the fluoresceinated peptides) were N-acetylated and C-amidated, with sequences as given in Table 2.2. An 18-mer N-terminally fluoresceinated Bim (IWIAQELRRIGDEFNAYY) (F-Bim) was used as the competitor peptide, and K_d values were determined for this peptide as described for the direct fluorescence anisotropy assays (with 4 replicate measurements). The K_d values were fit as described above, and also by an alternate method in which we fit the concentration of fluoresceinated Bim, K_d, and upper baseline. This resulted in better fits but a larger degree of uncertainty. Both K_d values were used to fit K_i values for the competition curves, yielding the range of K_i values reported in Table 2.3. In the competition assay, receptor concentrations were 50 nM, and fluoresceinated Bim was at 25 nM, in a final volume of 100 μL. Plates were mixed and incubated at 25 °C for 2 hours before reading as for the direct assays. Experiments were done in duplicate (triplicate or more for MS1). Competition fluorescence anisotropy data were fit to a complete competitive binding model (equation 17 in Roehrl et al., 2004) using Matlab. For the competition curves, the lower baselines were fixed to the average of the Bim lower baselines, which were well determined. All curves should reach the same lower baseline, as they all used the same fluoresceinated Bim. In addition, the receptor concentration was fixed as 50 nM, and the fluoresceinated Bim concentration, upper baseline, and K_i were fit. Competition curves were fit to the average of at least two replicates.

Cellular BH3 profiling assays

Assay plates were produced by serial dilution of each peptide from 200 μM to 0.2 nM using ten-fold dilutions in DTEB (Derived from Trehalose Experimental Buffer: 1 35 mM trehalose, 50 mM KCl, 20 μM EDTA, 20 μM EGTA, 0.1% BSA, 5 mM succinate, 10 mM HEPES-KOH pH

7.5) containing 0.005% w/v digitonin, 10 mM 2-mercaptoethanol, 2 mM JC-1, and 20 mg/mL oligomycin. Triplicate wells for each peptide were made for each cell line by adding 15 mL of the peptide dilutions to each well of a black, untreated 384-well plate. Control wells containing no peptide or 20 mM FCCP (carbonyl cyanide-4(trifluoromethoxy) phenylhydrazone, a chemical uncoupler of oxidative phosphorylation) were included for zero and complete depolarization, respectively. Multiple plates were produced from the same stock and frozen at –80 °C for later use. Frozen plates were brought to room temperature prior to use, and cells were suspended in DTEB at a density of 1.34×10^6 cells/mL, and 15 mL of cell suspension was added to each well of the dilution series to yield wells ranging from 0.1 nM-100 mM peptide and 20000 cells/well. Fluorescence of JC-1 aggregates was measured at 590 nm with 545 nm excitation on a Tecan Safire2 at 5 minute intervals for 3 hours. The area under each signal-*vs.*-time curve was calculated and normalized to the untreated and FCCP values to produce the percent depolarization. Curves were plotted as the log [peptide] vs. percent depolarization, with sigmoidal dose-response curves fitted using Graphpad PRISM 6. For curves without an upper baseline, an upper limit on the EC_{50} was estimated by fitting the curve with the upper baseline fixed at 100% depolarization, as this was the upper limit reached by most curves with a complete upper baseline.

Acknowledgements

We thank Koch Institute Biopolymers and Proteomics Facility for peptide synthesis, and the Koch Institute Flow Cytometry Core at MIT for assistance with cell sorting and analysis. This work was supported by NIGMS awards R01-GM084181 and P50-GM068762, and G.W. Foight was supported by an NSF Graduate Research Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Abulwerdi, F.A., Liao, C., Mady, A.S., Gavin, J., Shen, C., Cierpicki, T., Stuckey, J.A., Showalter, H.D.H., and Nikolovska-Coleska, Z. (2014). 3-Substituted-N-(4-Hydroxynaphthalen-1-yl)arylsulfonamides as a Novel Class of Selective Mcl-1 Inhibitors: Structure-Based Design, Synthesis, SAR, and Biological Evaluation. J Med Chem *57*, 4111–4133.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Brunelle, J.K., Ryan, J., Yecies, D., Opferman, J.T., and Letai, A. (2009). MCL-1-dependent leukemia cells are more sensitive to chemotherapy than BCL-2-dependent counterparts. J Cell Biol *187*, 429–442.

Certo, M., Del Gaizo Moore, V., Nishino, M., Wei, G., Korsmeyer, S., Armstrong, S.A., and Letai, A. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. Cancer Cell *9*, 351–365.

Chen, T.S., Palacios, H., and Keating, A.E. (2013). Structure-based redesign of the binding specificity of anti-apoptotic Bcl-x(L). J Mol Biol *425*, 171–185.

Cohen, N.A., Stewart, M.L., Gavathiotis, E., Tepper, J.L., Bruekner, S.R., Koss, B., Opferman, J.T., and Walensky, L.D. (2012). A competitive stapled peptide screen identifies a selective small molecule that overcomes MCL-1-dependent leukemia cell survival. Chem Biol *19*, 1175–1186.

DeBartolo, J., Dutta, S., Reich, L., and Keating, A.E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. J Mol Biol *422*, 124–144.

Deng, J., Carlson, N., Takeyama, K., Dal Cin, P., Shipp, M., and Letai, A. (2007). BH3 profiling identifies three distinct classes of apoptotic blocks to predict response to ABT-737 and conventional chemotherapeutic agents. Cancer Cell *12*, 171–185.

Dunbrack, R.L., Jr., and Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Science *6*, 1661–1681.

Dutta, S., Gullá, S., Chen, T.S., Fire, E., Grant, R.A., and Keating, A.E. (2010). Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. J Mol Biol *398*, 747–762.

Fire, E., Gullá, S., Grant, R., and Keating, A. (2010). Mcl-1-Bim complexes accommodate surprising point mutations via minor structural changes. Protein Science 19, 507-519.

Flanagan, A.M., and Letai, A. (2008). BH3 domains define selective inhibitory interactions with BHRF-1 and KSHV BCL-2. Cell Death Differ *15*, 580–588.

Friberg, A., Vigil, D., Zhao, B., Daniels, R.N., Burke, J.P., Garcia-Barrantes, P.M., Camper, D., Chauder, B.A., Lee, T., Olejniczak, E.T., et al. (2013). Discovery of potent myeloid cell

leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. J Med Chem *56*, 15–30.

Fu, X., Apgar, J.R., and Keating, A.E. (2007). Modeling backbone flexibility to achieve sequence diversity: the design of novel alpha-helical ligands for Bcl-xL. J Mol Biol *371*, 1099–1117.

Herman, M.D., Nyman, T., Welin, M., Lehtiö, L., Flodin, S., Trésaugues, L., Kotenyova, T., Flores, A., and Nordlund, P. (2008). Completing the family portrait of the anti-apoptotic Bcl-2 proteins: crystal structure of human Bfl-1 in complex with Bim. FEBS Lett *582*, 3590–3594.

Kim, Y.B., Balasis, M.E., Doi, K., Berndt, N., DuBoulay, C., Hu, C.-C.A., Guida, W., Wang, H.-G., Sebti, S.M., and Del Valle, J.R. (2012). Synthesis and evaluation of substituted hexahydronaphthalenes as novel inhibitors of the Mcl-1/BimBH3 interaction. Bioorg Med Chem Lett *22*, 5961–5965.

Konopleva, M., Contractor, R., Tsao, T., Samudio, I., Ruvolo, P.P., Kitada, S., Deng, X., Zhai, D., Shi, Y.-X., Sneed, T., et al. (2006). Mechanisms of apoptosis sensitivity and resistance to the BH3 mimetic ABT-737 in acute myeloid leukemia. Cancer Cell *10*, 375–388.

Lee, E., Czabotar, P., and Van Delft, M. (2008). A novel BH3 ligand that selectively targets Mcl-1 reveals that apoptosis can proceed without Mcl-1 degradation. J Cell Biol *180*, 341–355.

Lee, E.F., Fedorova, A., Zobel, K., Boyle, M.J., Yang, H., Perugini, M.A., Colman, P.M., Huang, D.C.S., Deshayes, K., and Fairlie, W.D. (2009a). Novel Bcl-2 homology-3 domain-like sequences identified from screening randomized peptide libraries for inhibitors of the pro-survival Bcl-2 proteins. J Biol Chem *284*, 31315–31326.

Lee, E.F., Sadowsky, J.D., Smith, B.J., Czabotar, P.E., Peterson-Kaufman, K.J., Colman, P.M., Gellman, S.H., and Fairlie, W.D. (2009b). High-resolution structural characterization of a helical alpha/beta-peptide foldamer bound to the anti-apoptotic protein Bcl-xL. Angew Chem Int Ed Engl *48*, 4318–4322.

Letai, A., Bassik, M.C., Walensky, L.D., Sorcinelli, M.D., Weiler, S., and Korsmeyer, S.J. (2002). Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as prototype cancer therapeutics. Cancer Cell *2*, 183–192.

London, N., Gullá, S., Keating, A.E., and Schueler-Furman, O. (2012). In silico and in vitro elucidation of BH3 binding specificity toward Bcl-2. Biochemistry *51*, 5841–5850.

Muppidi, A., Doi, K., Edwardraja, S., Drake, E.J., Gulick, A.M., Wang, H.-G., and Lin, Q. (2012). Rational design of proteolytically stable, cell-permeable peptide-based selective Mcl-1 inhibitors. J Am Chem Soc *134*, 14734–14737.

Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., Bruncko, M., Deckwerth, T. L., Dinges, J., Hajduk, P. J., Joseph, M. K., Kitada, S., Korsmeyer, S. J., Kunzer, A. R., Letai, A., Li, C., Mitten, M. J., Nettesheim, D. G., Ng, S., Nimmer, P. M., O'Connor, J. M., Oleksijew, A., Petros, A. M., Reed, J. C., Shen, W., Tahir, S. K., Thompson, C.

B., Tomaselli, K. J., Wang, B., Wendt, M. D., Zhang, H., Fesik, S. W., and Rosenberg, S. H. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature 435*, 677–681.

Rega, M.F., Wu, B., Wei, J., Zhang, Z., Cellitti, J.F., and Pellecchia, M. (2011). SAR by interligand nuclear overhauser effects (ILOEs) based discovery of acylsulfonamide compounds active against Bcl-x(L) and Mcl-1. J Med Chem *54*, 6000–6013.

Roberts, A.W., Seymour, J.F., Brown, J.R., Wierda, W.G., Kipps, T.J., Khaw, S.L., Carney, D.A., He, S.Z., Huang, D.C.S., Xiong, H., et al. (2012). Substantial susceptibility of chronic lymphocytic leukemia to BCL2 inhibition: results of a phase I study of navitoclax in patients with relapsed or refractory disease. J Clin Oncol *30*, 488–496.

Roehrl, M.H.A., Wang, J.Y., and Wagner, G. (2004). A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization. Biochemistry *43*, 16056–16066.

Rudin, C.M., Hann, C.L., Garon, E.B., Ribeiro de Oliveira, M., Bonomi, P.D., Camidge, D.R., Chu, Q., Giaccone, G., Khaira, D., Ramalingam, S.S., et al. (2012). Phase II study of single-agent navitoclax (ABT-263) and biomarker correlates in patients with relapsed small cell lung cancer. Clin Cancer Res *18*, 3163–3169.

Ryan, J.A., Brunelle, J.K., and Letai, A. (2010). Heightened mitochondrial priming is the basis for apoptotic hypersensitivity of CD4+ CD8+ thymocytes. Proc Natl Acad Sci USA *107*, 12895–12900.

Ryan, J., and Letai, A. (2013). BH3 profiling in whole cells by fluorimeter or FACS. Methods (San Diego, Calif) *61*, 156–164.

Smith, B.J., Lee, E.F., Checco, J.W., Evangelista, M., Gellman, S.H., and Fairlie, W.D. (2013). Structure-guided rational design of α/β-peptide foldamers with high affinity for BCL-2 family prosurvival proteins. Chembiochem *14*, 1564–1572.

Souers, A.J., Leverson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature Medicine *19*, 202–208.

Stewart, M., Fire, E., and Keating, A. (2010). The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. Nat Chem Biol *6*, 595-601.

Tanaka, Y., Aikawa, K., Nishida, G., Homma, M., Sogabe, S., Igaki, S., Hayano, Y., Sameshima, T., Miyahisa, I., Kawamoto, T., et al. (2013). Discovery of potent Mcl-1/Bcl-xL dual inhibitors by using a hybridization strategy based on structural analysis of target proteins. J Med Chem *56*, 9635–9645.

Tse, C., Shoemaker, A.R., Adickes, J., Anderson, M.G., Chen, J., Jin, S., Johnson, E.F., Marsh, K.C., Mitten, M.J., Nimmer, P., et al. (2008). ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. Cancer Res *68*, 3421–3428.

Walensky, L.D., Kung, A.L., Escher, I., Malia, T.J., Barbuto, S., Wright, R.D., Wagner, G., Verdine, G.L., and Korsmeyer, S.J. (2004). Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. Science *305*, 1466–1470.

Wei, S.-H., Dong, K., Lin, F., Wang, X., Li, B., Shen, J.-J., Zhang, Q., Wang, R., and Zhang, H.-Z. (2008). Inducing apoptosis and enhancing chemosensitivity to gemcitabine via RNA interference targeting Mcl-1 gene in pancreatic carcinoma cell. Cancer Chemother Pharmacol *62*, 1055–1064.

Wertz, I.E., Kusam, S., Lam, C., Okamoto, T., Sandoval, W., Anderson, D.J., Helgason, E., Ernst, J.A., Eby, M., Liu, J., et al. (2011). Sensitivity to antitubulin chemotherapeutics is regulated by MCL1 and FBW7. Nature *471*, 110–114.

Yecies, D., Carlson, N., Deng, J., and Letai, A. (2010). Acquired resistance to ABT-737 in lymphoma cells that up-regulate MCL-1 and BFL-1. Blood *115*, 3304-3313.

Chapter 3

Locating herpesvirus Bcl-2 homologs in the specificity landscape of anti-apoptotic Bcl-2 proteins

Foight, G.W. & Keating, A.E., Locating herpesvirus Bcl-2 homologs in the specificity landscape of anti-apoptotic Bcl-2 proteins. Journal of Molecular Biology (2015), doi: 10.1016/j.jmb.2015.05.015

Abstract

Viral homologs of the anti-apoptotic Bcl-2 proteins are highly diverged from their mammalian counterparts, yet they perform overlapping functions by binding and inhibiting BH3 motif-containing proteins. We investigated the BH3 binding properties of the herpesvirus Bcl-2 homologs KSBcl-2, BHRF1, and M11, as they relate to those of the human Bcl-2 homologs Mcl-1, Bfl-1, Bcl-w, Bcl-x_L, and Bcl-2. Analysis of the sequence and structure of the BH3 binding grooves showed that, despite low sequence identity, M11 has structural similarities to Bcl-x_L, Bcl-2, and Bcl-w. BHRF1 and KSBcl-2 are more structurally similar to Mcl-1 than to the other human proteins. Binding to human BH3-like peptides showed that KSBcl-2 has similar specificity to Mcl-1, and BHRF1 has a restricted binding profile; M11 binding preferences are distinct from those of Bcl-x_L, Bcl-2 and Bcl-w. Because KSBcl-2 and BHRF1 are from human herpesviruses associated with malignancies, we screened computationally designed BH3 peptide libraries using bacterial surface display to identify selective binders of KSBcl-2 or BHRF1. The resulting peptides bound to KSBcl-2 and BHRF1 in preference to Bfl-1, Bcl-w, Bcl-x_L, and Bcl-2, but showed only modest specificity over Mcl-1. Rational mutagenesis increased specificity against Mcl-1, resulting in a peptide with a dissociation constant of 2.9 nM for binding to KSBcl-2 and >1000-fold specificity over human Bcl-2 proteins, and a peptide with >70-fold specificity for BHRF1. In addition to providing new insights into viral Bcl-2 binding specificity, this study will inform future work analyzing the interaction properties of homologous binding domains and designing specific protein interaction partners.

Introduction

Many proteins function by binding selectively to other proteins. Within homologous families of protein interaction domains, members can have overlapping yet distinct functional specificities that are determined by factors such as expression pattern, subcellular localization, turnover, and intrinsic biochemical properties. The Bcl-2 family of proteins regulates apoptosis using selective interactions between its pro- and anti-apoptotic members. In these regulatory complexes, a helix formed by the BH3 (Bcl-2 homology 3) motif of a pro-apoptotic family member binds into a groove on the surface of a globular anti-apoptotic receptor. In humans, there are five main anti-apoptotic Bcl-2 family receptor proteins: Bcl-x_L, Bcl-2, Bcl-w, Mcl-1, and Bfl-1. Three classes of pro-apoptotic Bcl-2 family proteins can engage these receptors via BH3 docking (Chipuk et al., 2010; Leber et al., 2007). First, multi-Bcl-2 homology motif effector proteins Bax and Bak oligomerize in the outer mitochondrial membrane to promote apoptosis; binding of anti-apoptotic proteins to the helical BH3 motifs in Bak or Bax inhibits this process and can block cell death. Second, pro-apoptotic activator BH3-only proteins such as Bid and Bim trigger the oligomerization of Bak and Bax, and this activity is suppressed by anti-apoptotic receptors binding to activator BH3 motifs. Finally, pro-apoptotic BH3-only sensitizer proteins contain BH3 motifs that selectively bind and inhibit subsets of anti-apoptotic receptors (Letai et al., 2002). Thus, sensitizer proteins promote apoptosis in a manner that depends on the complement of anti-apoptotic proteins expressed in a particular cell. Competitive binding of the three classes of BH3-containing proteins to anti-apoptotic proteins is a key mechanism for regulating apoptosis (Llambi et al., 2011; Yang et al., 1995).

More than 15 viral homologs of anti-apoptotic Bcl-2 proteins have been identified in large double-stranded DNA viruses including adenoviruses, herpesviruses, and poxviruses

(Galluzzi et al., 2008; Polster et al., 2004). Kaposi's sarcoma-associated herpesvirus (KSHV or human herpesvirus 8) and Epstein-Barr herpesvirus (EBV or human herpesvirus 4) both express viral Bcl-2 homologs. Current understanding of the function of viral Bcl-2 proteins in infection and virulence is still evolving, and the roles of these proteins likely vary from virus to virus. Both EBV and KSHV latently infect humans, and most carriers present no symptoms. However, both viruses can give rise to malignancies in immunocompromised hosts, such as patients on immunosuppression therapy following organ transplant or AIDS patients. A mutant EBV found in about 15% of Burkitt lymphomas engenders greatly enhanced apoptosis resistance to infected B cells, which was attributed to increased latent cycle expression of the EBV Bcl-2 protein BHRF1 (Xiaofei et al., 2009). This led to the discovery that BHRF1 is expressed at low levels in wild-type EBV-infected, latent Burkitt lymphoma cells and is important in preventing apoptosis triggered by aberrant cell proliferation signals from constitutive c-myc expression (Xiaofei et al., 2009; Watanabe et al., 2010). The function of KSBcl-2, the KSHV viral Bcl-2, in oncogenesis or infection is less clear, but this protein may also act to counteract apoptosis driven by cell proliferation signals, in this case by a viral cyclin (Ojala et al., 1999; 2000). KSBcl-2 has been shown to be important for the initial stages of lytic reactivation from latent infection (Gelgor et al., 2015). The murine gamma herpesvirus γHV68 expresses Bcl-2 homolog M11. Although γHV68 is not associated with cancer in mice, it has been proposed as a useful model system for studying KSHV and EBV, which do not infect mice, and for which cell lines have limited utility as models of viral infection (Hardwick and Bellows, 2003). M11 has been shown to be important for persistent replication and virulence during chronic infection (Gangappa et al., 2002).

Despite playing a similar anti-apoptotic role, the three herpesvirus Bcl-2 proteins are not equivalent. KSHV and γHV68 are gamma-2-herpesviruses, and the synteny of the Bcl-2 genes in

these two viruses is similar, suggesting that their viral Bcl-2 proteins KSBcl-2 and M11 may share evolutionary origins (Cheng et al., 1997). However, EBV is a gamma-1-herpesvirus, and its Bcl-2 homolog, BHRF1, is located in a different region in the genome. This suggests that ancestral BHRF1 may have been acquired in a separate horizontal gene transfer event, possibly originating from a different mammalian Bcl-2 homolog than KSBcl-2 and M11 (Cheng et al., 1997; Letai et al., 2002). A key functional difference between the viral Bcl-2 homologs is that although KSBcl-2 and M11 are anti-autophagic, this is not true of BHRF1; EBV upregulates autophagy for non-cell death purposes (Galluzzi et al., 2008; Polster et al., 2004; Taylor and Blackbourn, 2011).

Viral Bcl-2 proteins could conceivably function by mimicking a specific mammalian homolog. Alternatively, they could share roles played by several mammalian homologs, or even have distinct functions. The *in vivo* function of viral Bcl-2 homologs, and how it compares to that of their human counterparts, has not been extensively characterized. But some clues can be gleaned by looking at viral effects on the cell. Herpesvirus gene products can negatively regulate human Bcl-2 and Bcl-x_L, suggesting that the viral Bcl-2 homologs may need to compensate for the decreased activity of these human homologs. For example, EBV transcription factor BZLF1 downregulates the cellular protein CD74, resulting in T-cell evasion and decreased expression of Bcl-2 and Bcl-x_L in B lymphoblastoid cell lines (Xiaofei et al., 2009; Lantner et al., 2007; Zuo et al., 2011). An EBV-infected cell line was nevertheless recently shown to be dependent upon Bcl-x_L for resistance to apoptosis, but as BHRF1 expression was not detected in this cell line, its role relative to human Bcl-2 homologs remains unclear (Cojohari et al., 2015; Xiaofei et al., 2009; Watanabe et al., 2010). In the KSHV-infected cell line Bcbl-1, KSBcl-2 is expressed at low levels and Mcl-1 at high levels. Bcbl-1 cells exhibited a response to a panel of BH3 peptides

indicative of a dependence upon both Mcl-1 and KSBcl-2 for protection from apoptosis (Cojohari et al., 2015; Ojala et al., 1999; 2000). KSHV also downregulates Bcl-2 activity by expression of a viral cyclin that directs cellular CDK6 to phosphorylate and inactivate Bcl-2. This may be advantageous for the virus because human Bcl-2 can impair cell cycle progression and be converted into a pro-apoptotic form by caspase cleavage (Gelgor et al., 2015; Hardwick, 2000; Ojala et al., 2000). KSBcl-2 and M11 can also fulfill the anti-autophagic roles of Bcl-2 and Bcl-x_L by binding Beclin-1 (Hardwick and Bellows, 2003; Ku et al., 2008; Pattingre et al., 2005). These findings illustrate that in addition to filling the anti-apoptotic niche, it may be advantageous for herpesviruses to use their Bcl-2 homologs to fulfill additional human Bcl-2 roles (e.g., in autophagy), but not others (e.g. pro-apoptotic and cell cycle regulatory roles). The functional analogies between human and viral Bcl-2 homologs, and how any similarities or differences relate to BH3 binding profiles, remain to be elucidated.

The mechanistic details of protection from apoptosis rely on which pro-apoptotic Bcl-2 family members each anti-apoptotic Bcl-2 homolog binds. The BH3 interaction preferences of the human anti-apoptotic Bcl-2 proteins have been extensively studied, with particular attention focused on the large differences between Bcl-x_L and Mcl-1 (Dutta et al., 2010b; Fire et al., 2010; Foight et al., 2014; Gangappa et al., 2002; Lee et al., 2009b; 2008; Zhang et al., 2012). BH3 motif binding is often tested using peptides ~20 residues in length, here referred to as BH3 peptides. Bim, Bid, and Puma BH3 peptides all bind to the five main anti-apoptotic Bcl-2 proteins, but sensitizer BH3 peptides such as Bad and Noxa are selective for different sets of anti-apoptotic receptors. Notably, Bad binds tightly to Bcl-x_L, Bcl-2, and Bcl-w, but not Mcl-1, whereas Noxa preferentially binds Mcl-1 (Certo et al., 2006; Chen et al., 2005; Cheng et al., 1997). This distinction has long been used to group Bcl-x_L, Bcl-2, and Bcl-w into a common

specificity class and Mcl-1 into its own class. Bfl-1 is sometimes grouped into a class with Mcl-1, based on not binding to Bad and binding weakly to Noxa, Bik, and Hrk. However, human Bfl-1 does not bind two murine Noxa variants, distinguishing it from Mcl-1, which does bind these proteins (Certo et al., 2006; Chen et al., 2005; Vogler, 2012). Viral protein BHRF1 has been shown to have a limited BH3 binding profile, binding only Bim, Bid, and Puma out of a set of 10 mammalian BH3 peptides tested (Flanagan and Letai, 2008). KSBcl-2 and M11 have more permissive binding and exhibit BH3 binding profiles more similar to that of Mcl-1 in that they show moderate binding to Noxa, but only very weak binding to Bad (Flanagan and Letai, 2008; Huang et al., 2003; Ku et al., 2008; Sinha et al., 2008). Further comparison of the binding specificities of viral and human Bcl-2 proteins may shed light on how viral Bcl-2 functions compare to human Bcl-2 functions, as BH3 binding specificity is a crucial determinant of anti-apoptotic Bcl-2 activity.

Studying the binding preferences of viral and human Bcl-2 homologs can also illuminate differences that could be exploited to design specific protein interaction inhibitors. Specific inhibitors of the viral Bcl-2 homologs would be useful for basic research and potentially for therapy. Early peptidic and small molecule inhibitors targeted at human Bcl-x_L achieved inter-class specificity in that they showed no binding to Mcl-1, but did not distinguish between members of the Bcl-x_L, Bcl-2, and Bcl-w class (Dutta et al., 2010b; Oltersdorf et al., 2005). Targeting Bcl-2 and Bcl-x_L in cancers with Navitoclax, the clinical form of the small-molecule inhibitor ABT-737, led to thrombocytopenia due to the fact that platelets are dependent upon Bcl-x_L for protection from apoptosis (Roberts et al., 2012; Rudin et al., 2012). A Bcl-2-specific small molecule, ABT-199, is achieving better success in clinical trials due to reduced off-target effects (Souers et al., 2013). This highlights the importance of specific targeting of Bcl-2

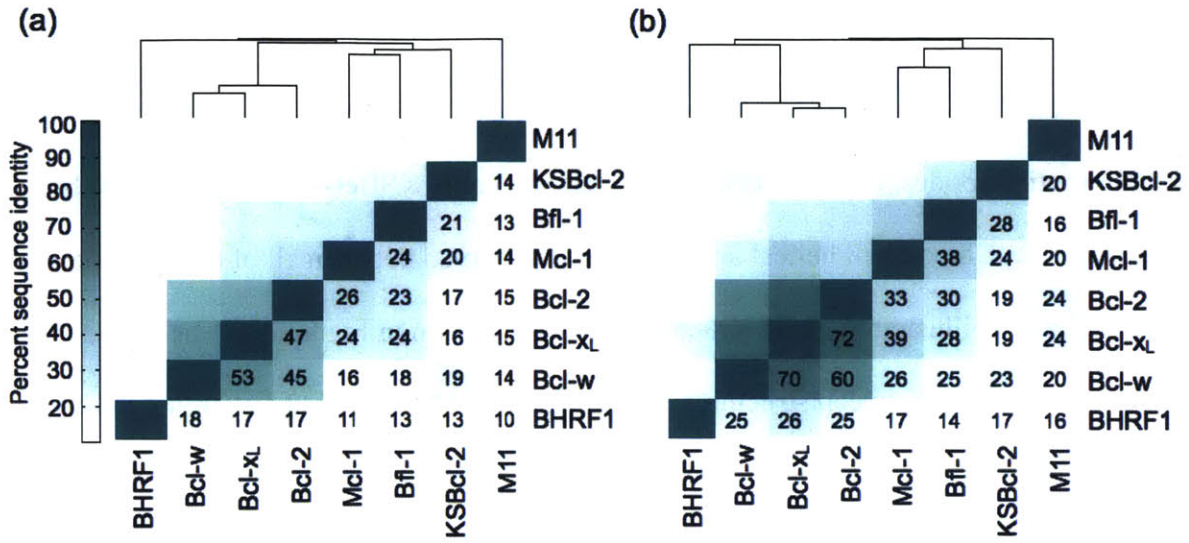
homologs for cancer therapy. Significant progress has been made identifying peptide inhibitors of individual human Bcl-2 family members including Mcl-1, Bcl-x_L and Bfl-1 (Dutta et al., 2013; 2014; Foight et al., 2014). For viral Bcl-2 homologs, a computationally designed protein, BINDI, binds BHRF1 with picomolar affinity and >180-fold specificity over the human Bcl-2 homologs (Procko et al., 2014). BINDI is a 14 kDa protein that incorporates a BH3 helix but gains much of its specificity from contacts outside of the BH3 binding groove. When attached to an antibody-targeted intracellular delivery carrier, BINDI reduced tumor growth in xenograft mouse models of EBV-positive human lymphoma, supporting BHRF1 as a candidate therapeutic target. A peptide variant of the BH3 motif of Beclin-1 was recently engineered that binds selectively to M11 over Bcl-x_L, but the peptide had weak affinity for M11 (K_D = 6.4 μM) (Su et al., 2014).

In this paper we report comparisons of viral and human Bcl-2 homologs at the levels of sequence, structure, and binding similarity in order to identify which homologs are most similar. Drawing on experimental data and structural models, we then interrogated the boundaries of this similarity by designing BH3 peptide libraries and screening them for selective binders of KSBcl-2 and BHRF1. Analysis of the peptides identified in the screen shed light on sequence and structural determinants of the BH3 binding preferences of viral and human homologs. Further mutation of library peptides provided molecules with >1000-fold specificity for binding KSBcl-2 and >70-fold specificity for binding BHRF1 over all of the human Bcl-2 homologs. These peptides could serve as reagents to probe and inhibit the function of viral Bcl-2 homologs in viral pathogenesis.

Results

Comparison of eight Bcl-2 homologs based on sequence identity, structure, and binding preferences

Protein sequence identity is often used to infer functional similarity. We compared the sequences of 8 anti-apoptotic Bcl-2 family homologs across the entire Bcl-2 domain (without the C-terminal trans-membrane helix or the N-terminal PEST domain of Mcl-1) (Figure 3.1). The viral Bcl-2 homologs KSBcl-2, BHRF1, and M11 have very low sequence identity to the human homologs and to each other (10-21%, Figure 3.1a). To assess the similarity of receptors in the regions that most directly influence BH3 helix binding, we computed the sequence identity over residues that line the BH3 binding groove (Figure 3.1b). Residues that were included are within 7 Å of a peptide residue in at least one BH3-bound structure, as described in the Materials and Methods. Sequence identity in the binding groove is higher than in the overall Bcl-2 domain, and Bcl-2, Bcl-x_L and Bcl-w form a subgroup sharing high sequence identity in this region ($\geq$ 60%). Human homologs Bfl-1 and Mcl-1 share at most 30% and 39% binding-groove identity with the Bcl-x_L/2/w set, respectively, and share 38% identity with each other. The human Bcl-2 homologs with highest sequence identity to the viral Bcl-2 proteins, in the binding groove, are Bfl-1 and Mcl-1 for KSBcl-2 (28%, and 24% identity, respectively), Bcl-x_L/2/w for BHRF1 (25-26% identity), and Bcl-x_L and Bcl-2 for M11 (24% identity). Thus, the similarities between Mcl-1, Bfl-1, and the viral homologs are much less significant than those between Bcl-x_L, Bcl-2, and Bcl-w, although Mcl-1, Bfl-1 and KSBcl-2 do cluster together at a lower similarity threshold.



(c)

```
                helix 1
Bcl-xL    ------MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEG----TESEMETPSAINGNPSWHLADSPA
Bcl-2     MAHAGRTGYDNREIVMKYIHYKLSQRGYEWDAGDVGAAPPGAAPAPGIFSSQPGHTPHPAASRDP---VARTSP
Bcl-w     -MATPASAPDTRALVADFVGYKLRQKGYVCGAGP-----------------------------------GEGPA
M11       -MSHKKSGTYWATLITAFLKTVSKVEELDCV-------------------------------------------
Bfl-1     --MTDCEFGYIYRLAQDYLQCVL-QIP-------------------------------QPGSGP---------
Mcl-1     ----DDLYRQSLEIISRYLREQATG---------------------------------AKDTKP---MGRS--
KSBcl-2   -MDEDVLPGEVLAIEGIFMACGLNEPEYLYHP-----------------------------------------
BHRF1     -----MAYSTREILLALCIRDSRVHGNGTLHPVLELAARETPLRLSPEDT-----------------------

                              helix 2        helix3             helix 4
Bcl-xL    VNGATAHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHIT-PGTAYQSFEQVVNELF-R---D
Bcl-2     LQTPAAPGAAAGPALSPVPPVVHLTLRQAGDDFSRRYRRDFAEMSSQLHLT-PFTARGRFATVVEELF-R---D
Bcl-w     ------------------ADPLHQAMRAAGDEFETRFRRTFSDLAAQLHVT-PGSAQQRFTQVSDELF-Q---G
M11       ---------------DSAVLVDVSKIITLTQEFRRHYDSVYRADYGPALK----NWKRDLSKLFTSLF-V---D
Bfl-1     -------------------SKTSRVLQNVAFSVQKEVEKNLKSCLDNVNVVSVDTARTLFNQVMEKEF-E---D
Mcl-1     ---------------GATSRKALETLRRVGDGVQRNHETAFQGMLRKLDIKNEDD-VKSLSRVMIHVF-S---D
KSBcl-2   ---------------------LLSPIKLYITGLMRDKESLFEAMLANVRFH----STTGINQLGLSML-QVSGD
BHRF1     ---------------------VVLRYHVLLEEIIERNSETFTETWNRFITHTEH-VDLDFNSVFLEIFHR---G
                                ....... ...............    ... ....... .   .
             helix 5                    helix 6                helix 7   helix 8
Bcl-xL    -GVNWGRIVAFFSFGGALCVESVDKE---MQVLVSRIAAWMATYLNDHLEP-WIQENGGWD-TFVE-LYGN
Bcl-2     -GVNWGRIVAFFEFGGVMCVESVNRE---MSPLVDNIALWMTEYLNRHLHT-WIQDNGGWD-AFVE-LYGP
Bcl-w     -GPNWGRLVAFFVFGAALCAESVNKE---MEPLVGQVQEWMVAYLETRLAD-WIHSSGGWA-EFTA-LYGD
M11       -VINSGRIVGFFDVGRYVCEEVLCPGS--WTEDHELLNDCMTHFFIENNLMNHFPLEDIFL-AQRK-FQTT
Bfl-1     GIINWGRIVTIFAFEGILIKKLLRQQIAPDVDTYKEISYFVAEFIMNNTG-EWIRQNGGWENGFVK-KFEP
Mcl-1     GVTNWGRIVTLISFGAFVAKHLKTIN---QESCIEPLAESITDVLVRT-KRDWLVKQRGWD-GFVE-FFHV
KSBcl-2   GNMNWGRALAILTFGSFVAQKLSNE-----PHLRDFALAVLPVYAYEAIGPQWFRARGGWR-GLKA-YCTQ
BHRF1     -DPSLGRALAWMAWCMHACRTLCCNQSTPYYVVDLSV-RGMLEASE-GLDG-WIHQQGGWS-TLIEDNIPG
           . .... ... ..                                          .     .........
```

Figure 3.1. Sequence comparison of 5 human and 3 viral Bcl-2 homologs. (a) Percent sequence identity over the entire Bcl-2 domain (without the C-terminal trans-membrane helices or the N-terminal PEST domain of Mcl-1). (b) Percent sequence identity for residues in the BH3 binding groove (see Materials and Methods). (c) Sequence alignment used for the sequence identity

calculations. Dots under residues denote residues in the BH3 binding groove. Motifs discussed in the Results section are colored and/or underlined.

Abundant structural data exist for complexes of anti-apoptotic Bcl-2 proteins bound to short BH3 motif peptides; of the proteins discussed here, only KSBcl-2 lacks a structure bound to a BH3 peptide. The overall helical architecture is conserved between all of the anti-apoptotic Bcl-2 homologs, as illustrated in the structure alignment shown in Figure 3.2. The Cα RMSD varies from ~0.8-3.2 Å between structures of different Bcl-2 homologs. To examine the chemical similarity of the different binding grooves, we performed analyses using SiteMAP on receptor-peptide complex structures or homology models for all eight Bcl-2 homologs. SiteMAP, originally designed for identifying small molecule binding sites, creates maps of hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor binding potential proximal to the protein surface (Halgren, 2009; 2007). Our process for defining a similarity score based on the intersection of SiteMAP maps from different receptors is summarized in Figure 3.2 and explained in greater detail in the Materials and Methods. A summary of the similarities between all pairs of maps, computed as the sum of the intersections of physicochemical property maps for pairs of proteins, is shown in Figure 3.3, where protein structures are clustered according to their intersection score profiles. Where available, we ran the SiteMAP analysis on multiple structures for each Bcl-2 homolog, including structures of the receptor with different BH3 peptides bound. Notably, different structures of the same receptor bound to different peptides clustered together, showing that this analysis is robust to small changes in conformation. To further test dependence on small changes in Bcl-2 conformation, we performed a restrained minimization without the peptide present for one structure of each homolog before running SiteMAP. The minimized structure grouped with the other structures of that homolog, though in some cases more distantly.

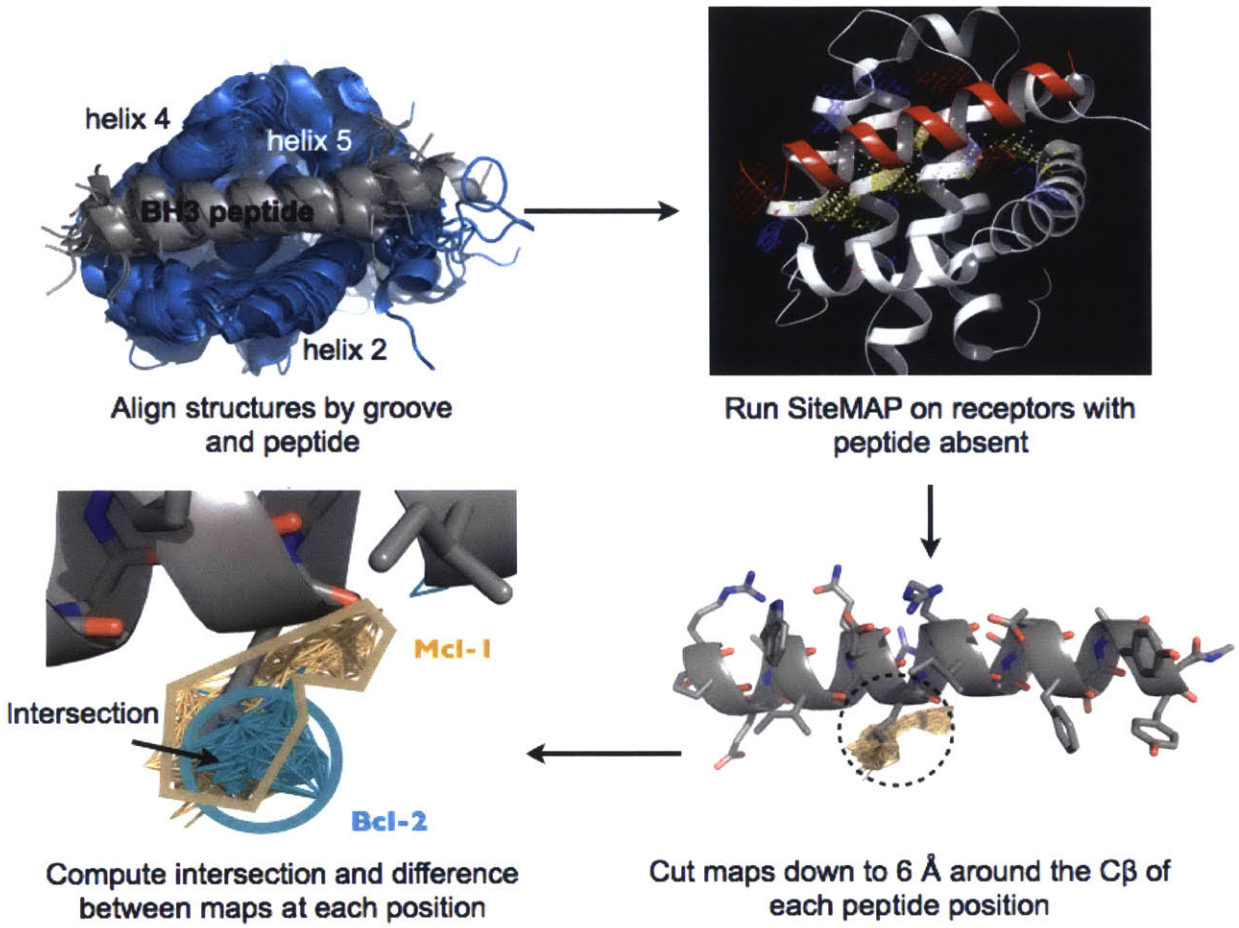


Figure 3.2. Schematic describing the SiteMAP analysis process. Bcl-2:BH3 complex structures or homology models were aligned to the Mcl-1:Bim (Fire et al., 2010) structure in PyMOL based on helix 2, helices 4-5, and the peptide. The image at the upper left includes all structures and homology models used, with the exception of the minimized structures (see Materials and Methods for a complete list of structures and references). SiteMAP was run on the receptors with the peptide removed. The peptide is shown here to illustrate the binding groove. The Mcl-1 structure peptide was then used as a common ligand for all receptors, and a 6 Å radius sphere around the Cβ of each peptide position was used to compare the SiteMAPs around each peptide position. To compute a similarity metric, the number of SiteMAP points of a given type (hydrophobic, acceptor, donor) found within 1 Å of a point of the same type in another receptor's map was counted to give an intersection score. The similarity scores presented in Figure 3.3 are the sum of the intersections of the hydrogen bond acceptor, donor, and hydrophobic maps.

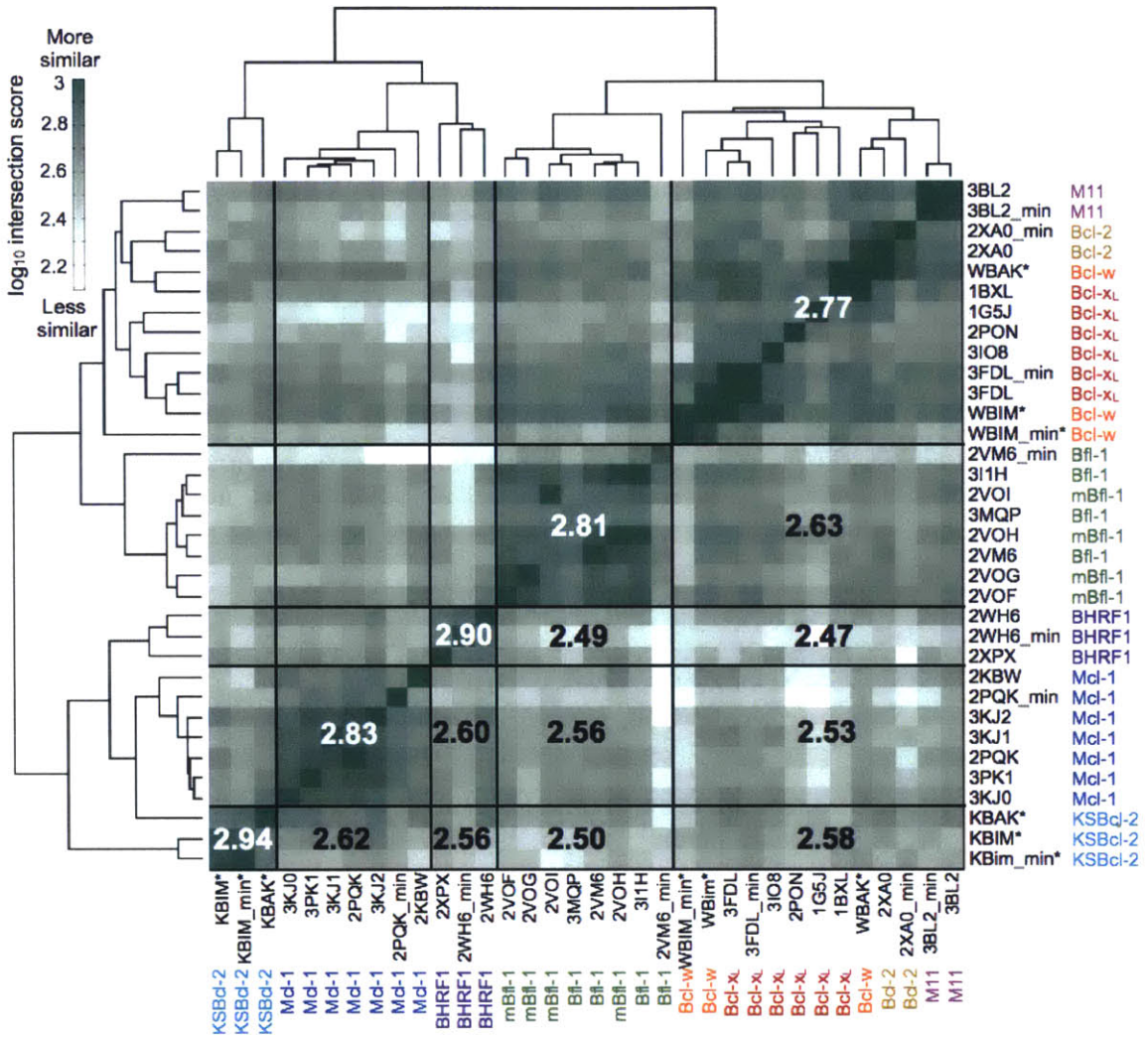


Figure 3.3. Comparison of the physicochemical characteristics of human and viral Bcl-2 homolog BH3 binding groove structures using SiteMAP. The intersection score is a measure of similarity that takes into account the hydrophobic and hydrogen-bond donor/acceptor characteristics of regions of the binding site (see Materials and Methods for details of the metric). The structures of eight Bcl-2 homologs were clustered according to the similarity of their intersection score profiles. Boxed regions partition receptors or groups of receptors; numbers are the average $\log_{10}$(intersection score) over the indicated box and its symmetry-related box across the diagonal. PDB IDs are given, with the receptor name next to them. All mammalian receptors are the human homologs except for several murine Bfl-1 structures ("mBfl-1"). The suffix "min" denotes a structure relaxed without the peptide bound. Asterisks denote homology models.

Consistent with the sequence-based analysis, SiteMAP results show that Bcl-x_L, Bcl-2, and Bcl-w form a tight cluster. M11 also joins this group, despite having much lower sequence identity with these proteins than they share with each other. Viral proteins KSBcl-2 and BHRF1

clearly cluster with Mcl-1. But surprisingly, given its higher binding-groove sequence identity to Mcl-1 than to Bcl-x_L/2/w, Bfl-1 clusters more closely to the Bcl-x_L/2/w/M11 group. Averaged similarity scores (shown in bold for boxed sets of structures in Figure 3.2) show that Mcl-1 and KSBcl-2 are most alike in binding-groove chemical structure, despite the fact that Mcl-1 is more similar to Bcl-x_L and Bfl-1 than to KSBcl-2 by sequence identity. The KSBcl-2 and Bcl-w maps are based on homology models, which are heavily influenced by the homologs with which they share the greatest sequence similarity, which are Bfl-1 and Mcl-1 for KSBcl-2 and Bcl-x_L for Bcl-w. Therefore, the high similarity between these structures may be somewhat artificial, but it is consistent with other trends reported in this paper. While this manuscript was in preparation, a structure of a mutant Bcl-w bound to a peptide corresponding to its own (mutated) BH3 domain was published (Lee et al., 2014). The Cα RMSD over helices 2-8 between our Bcl-w:Bak BH3 homology model and this Bcl-w:Bcl-w BH3 structure is 1.2 Å, validating our Bcl-w model.

Binding of the viral and human Bcl-2 homologs to a small set of functionally validated human BH3 peptides has been previously reported and is summarized in Figure 3.4a (Certo et al., 2006; Flanagan and Letai, 2008; Ku et al., 2008). Recently, DeBartolo et al. reported dissociation constants for the five human Bcl-2 homologs binding to 36 new, computationally identified candidate BH3 peptides from the human proteome (DeBartolo et al., 2014). In Figure 3.4b, we compare the binding patterns of the three viral Bcl-2 homologs to those of the human homologs for these 36 BH3-like peptides. The dissociation constants were determined by fluorescence anisotropy binding experiments. Hierarchical clustering of K_D values by Bcl-2 receptor shows similar trends for binding of the previously reported and functionally validated vs. new BH3 peptides. Bcl-x_L, Bcl-2, and Bcl-w, which have high sequence identity and structural similarity with each other, cluster together as expected. BHRF1 and Bfl-1 are grouped

together based on binding tightly to only a small number of peptides. KSBcl-2 and Mcl-1 cluster together in Figure 3.4b and are in the same group in Figure 3.4a, although there are a few notable differences in the binding profiles of these two proteins. In Figure 3.4b, peptides in the clusters marked with asterisks exhibit moderate-to-tight binding to KSBcl-2 and also to Bcl-x_L/2/w, but weaker binding to Mcl-1. We return to this point below.

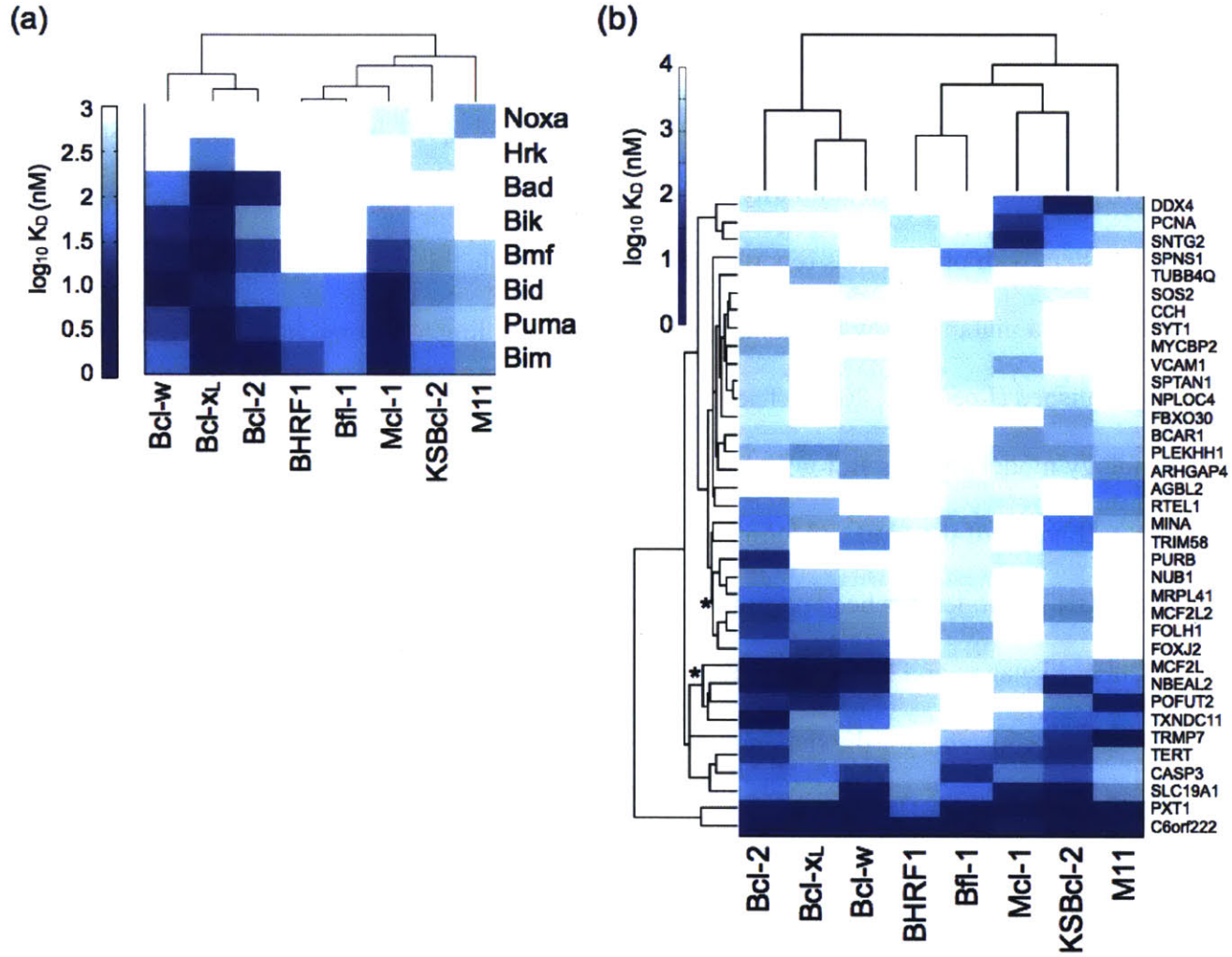


Figure 3.4. Comparison of BH3 peptide binding profiles for eight Bcl-2 homologs. (a) Interactions with functionally validated BH3 peptides, using data from the literature. (b) Interactions with 36 BH3-like peptides identified from the human proteome. Receptors are clustered based on correlation of their binding profiles, and peptides are clustered by Euclidean distance. Dissociation constants for the human proteins in panel B were taken from DeBartolo et al. Asterisks denote peptide clusters with tighter binding to KSBcl-2 than to Mcl-1. The heat map indicates affinity measured by fluorescence anisotropy as $\log_{10}$ (K_D in nM), with white indicating no detectable binding up to 3000 nM. K_D values with 95% confidence intervals are given in Table 1 for the 36 BH3-like peptides.

Table 3.1. Dissociation constants (nM) and 95% confidence intervals for BH3-like peptides in human proteins.

Gene ID	Bcl-xL	Bcl-w	Bcl-2	Bfl-1	Mcl-1	BHRF1	KSBcl-2	M11
PXT1	9±7[a]	5.5±0.4	14±2	<1	<1	140±77	<1[b]	<1
C6orf222	5±1	13±0	<1	<1	24±8	<1	<1	1.9±0.80
MCF2L	6±2	7±2	6±1	X 4100	X 4000	>3000	x2300	x1100
TXNDC11	1300± 200	280± 70	9±1	>10^5	2100± 500	x5200	140±59	150±38
NBEAL2	10±1	30±5	17±3	>10^4	2600± 900	x6600	31±4.6	250± 110
SLC19A1	1100± 200	18±2	238±9	190± 30	26±4	x840	8.9±4.6	x740
SNTG2	X4000	X 6525	X 3045	X 4068	29±14	>3000	210±41	x2000
POFUT2	40±2	120± 80	60± 50	>10^5	>10^4	>3000	x620	23±3.3
CASP3	280± 80	70±10	140±4	44±1	500± 200	x1400	x110	x2400
TERT	800± 200	900± 150	50± 10	101±8	80±40	x1300	49±18	x1300
PURB	>10^4	>10^4	40± 20	X 4665	x4000	>3000	x2400	>3000
MCF2L2	290± 20	1120± 40	60±1	X 3367	>10^5	>3000	600±300	>3000
PCNA	X 10,000	>10^4	>10^5	X 14,200	62±6	x2600	150±75	x4900
FOLH1	500± 100	900± 100	70±4	1300± 200	>10^5	>3000	x2400	>3000

FOXJ2	80±10	110± 40	200± 100	>3367	5100± 700	>3000	x4000	>3000
TRMP7	680± 30	>10^4	130± 10	2100± 800	570± 80	x9300	59±5.6	3.4±1.7
DDX4	X4000	X4350	X3045	>10^5	140± 10	>3000	14±3.6	x1300
MRPL41	1310± 60	5200± 2500	190± 60	X4665	>10^5	x5700	x1300	>3000
MINA	1010± 70	1200± 200	270± 50	670± 90	>10^5	x3700	160±33	x1300
SPNS1	3100± 200	6500± 4700	1320± 50	320± 70	640± 170	x9400	x2700	>3000
RTEL1	1400± 300	>10^4	450± 150	X4000	X4000	>3000	>3000	x600
TRIM58	>10^5	460± 210	1030± 60	X4665	X6071	>3000	240±120	>3000
NUB1	2700± 1000	X4000	564±9	X4000	X 10,000	>3000	x2600	>3000
PLEKHH1	620± 120	1000± 200	1500± 200	X5680	760± 150	>3000	x850	x1900
ARHGAP4	3700± 1100	760± 90	>10^5	X5595	4600± 3300	>3000	x3800	810± 560
BCAR1	2800± 800	2500± 700	3000± 1000	>10^4	930± 70	>3000	x1500	x2600
TUBB4Q	1120± 60	2500± 200	>10^4	X4665	>10^5	>3000	>3000	>3000
MYCBP2	>10^5	>10^5	1180± 50	X3367	X4000	>3000	>3000	>3000
VCAM1	X7000	X4350	X3045	X3368	1200± 200	>3000	>3000	>3000
NPLOC4	>10^5	X4500	2700± 1600	X4665	X4000	>3000	x3500	>3000

SPTAN1	X7000	X4350	X3045	X3370	4000± 2000	>3000	x4200	>3000
FBXO30	>10^5	4300± 2400	4000± 600	>10^4	>10^5	>3000	x1200	x4700
SOS2	X6900	4900± 3500	>10^4	>10^4	X4000	>3000	x4700	>3000
CCH	>10^4	>10^4	>10^4	X9615	X4000	>3000	>3000	>3000
AGBL2	>10^4	>10^4	>10^5	X4665	X5025	>3000	>3000	210± 150
SYT1	>10^5	X4350	>10^5	X3367	X4000	>3000	x8000	>3000

[a] Values for the human Bcl-2 homologs are from DeBartolo et al., 2014.[5] See reference for derivation of 95% confidence intervals and approximate K_D values for these receptors.
[b] For the viral Bcl-2 homologs, values designated >3000 showed no binding up to 3000 nM. 3000 nM was the highest concentration of receptor used. Values designated as <1 were too tight to fit accurately. Values marked with 'x' are approximate values for weak binders that gave an incomplete upper baseline, resulting in a large 95% confidence interval.

Natural BH3 sequences have high variability, which makes it difficult to parse the determinants that underlie the observed binding patterns. To compare binding preferences in a more interpretable sequence space, we performed Bim BH3 substitution SPOT array analysis for KSBcl-2 and BHRF1. The SPOT array assay is used to test binding to hundreds of membrane-immobilized peptides in parallel, and has proven useful in published analyses of Bcl-2 family binding (DeBartolo et al., 2012; Dutta et al., 2010a; London et al., 2012). Our viral Bcl-2 SPOT arrays included peptides with 18 individual point mutations (excluding cysteine and methionine) at a set of 10 positions in Bim BH3 that were previously analyzed for human Bcl-2 protein binding (Figure 3.5). We refer to BH3 peptide positions using a repeating-heptad nomenclature, with positions labeled a-g, which is shown for the full Bim BH3 sequence in Figure 3.6c. The viral Bcl-2 arrays additionally included peptides with mutations in 3 positions in the N-terminal region of Bim, and 2 positions in the C-terminal region. However, varying these terminal

positions gave only small changes in binding signal in this assay, not readily distinguished from noise; we do not interpret observed differences at these positions here.

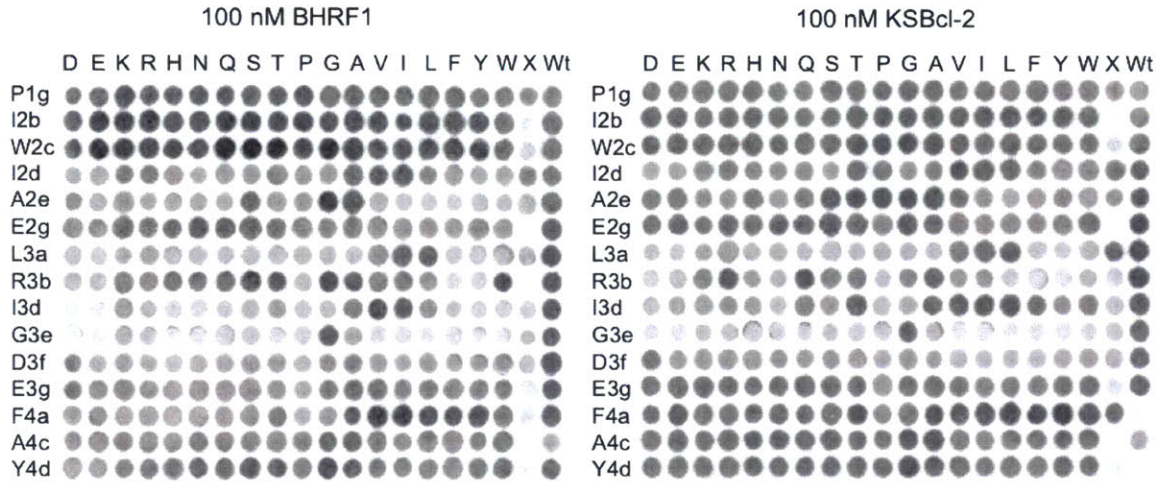


Figure 3.5. Bim BH3 and point-mutant peptides on SPOT arrays binding to 100 nM BHRF1 or KSBcl-2. Each row is labeled with the wild-type Bim residue and the position that was varied, and columns are labeled with the substitution. The column labeled "X" included peptides with the following BH3 sequences, from top to bottom: Bad, Bid, Bmf, Hrk, Noxa, Bik, Bak, Bax, Bcl-x_L, Bcl-w, Bcl-2, Mcl-1, Bfl-1, BHRF1, KSBcl-2.

The peptide binding similarity of different Bcl-2 proteins can be assessed using the correlation of SPOT signals between receptors for a common set of 180 Bim BH3 variants tested for five human and two viral proteins. Previous analyses showed that of the five human proteins tested by SPOT analysis, only the Bcl-x_L/w/2 arrays showed a high correlation of binding patterns (DeBartolo et al., 2012). The correlation of the KSBcl-2 SPOT array signals with SPOT data for most human homologs was moderate (Pearson R = 0.71-0.80), with Bcl-x_L having slightly lower similarity (R = 0.66). The BHRF1 SPOT array data were moderately correlated with data for KSBcl-2 and Mcl-1 (R = 0.75 and 0.71, respectively), but correlations with the other homologs were lower. Notably, despite exhibiting a similarly restrictive binding profile for natural BH3 peptides, Bfl-1 and BHRF1 show relatively low correlation between their SPOT

array signals (R = 0.66), suggesting they arrive at their restrictive binding profiles through different mechanisms. Differences in positional preferences that can be dissected with the use of the SPOT data are discussed below.

Peptide libraries targeting KSBcl-2 and BHRF1 specificity

Peptide design using library screening has repeatedly led to molecules that discriminate between binding to Bcl-x_L/2/w vs. Mcl-1 or Bfl-1 by large margins.(Dutta et al., 2010b; Lee et al., 2009b; Zhang et al., 2012) It has been more difficult to design specificity between members of more closely homologous groups (e.g., Bcl-x_L vs. Bcl-2/w or Bfl-1 vs. Mcl-1) (Dutta et al., 2013; 2014). Above, we showed that KSBcl-2 and BHRF1 have BH3 binding patterns most similar to those of Mcl-1 and Bfl-1, respectively. We interrogated the boundaries of this similarity by designing and screening Bim BH3-based libraries for peptides that could bind KSBcl-2 or BHRF1 selectively (Figure 3.6). To increase our chances of success, we used experimental binding data and structure-based models of BH3 peptide binding preferences to design libraries enriched in mutations predicted to provide specificity for KSBcl-2 or BHRF1. The libraries were based on mutating the Bim BH3 motif, for which we have rich binding data from prior work. Specifically, we designed the libraries using mutational data from SPOT arrays and predictions made using STATIUM, a statistical potential that can evaluate mutations based on analysis of human or viral Bcl-2 crystal structures or homology models (DeBartolo et al., 2012; 2014). We also used Illumina sequencing data from a yeast surface display library of Bim BH3 variants previously screened for KSBcl-2 binding (Foight et al., 2014). The computationally assisted library design process (described in further detail in the Materials and Methods) focused on including residues that were tolerated by the viral Bcl-2 proteins and that weakened binding to Mcl-1 and, for the BHRF1 library, Bfl-1. Despite moderate sequence

identity between KSBcl-2 and Bfl-1, we did not consider Bfl-1 to be a competitor difficult to discriminate against, because we have previously discovered mutations that favor KSBcl-2 over Bfl-1 binding, and we included these in the library (Foight et al., 2014). We also prioritized inclusion of a few mutations known to disfavor Bcl-x_L, Bcl-2, and Bcl-w binding (e.g. non-aromatic residues at position 4a) (Dutta et al., 2010b; Fire et al., 2010).

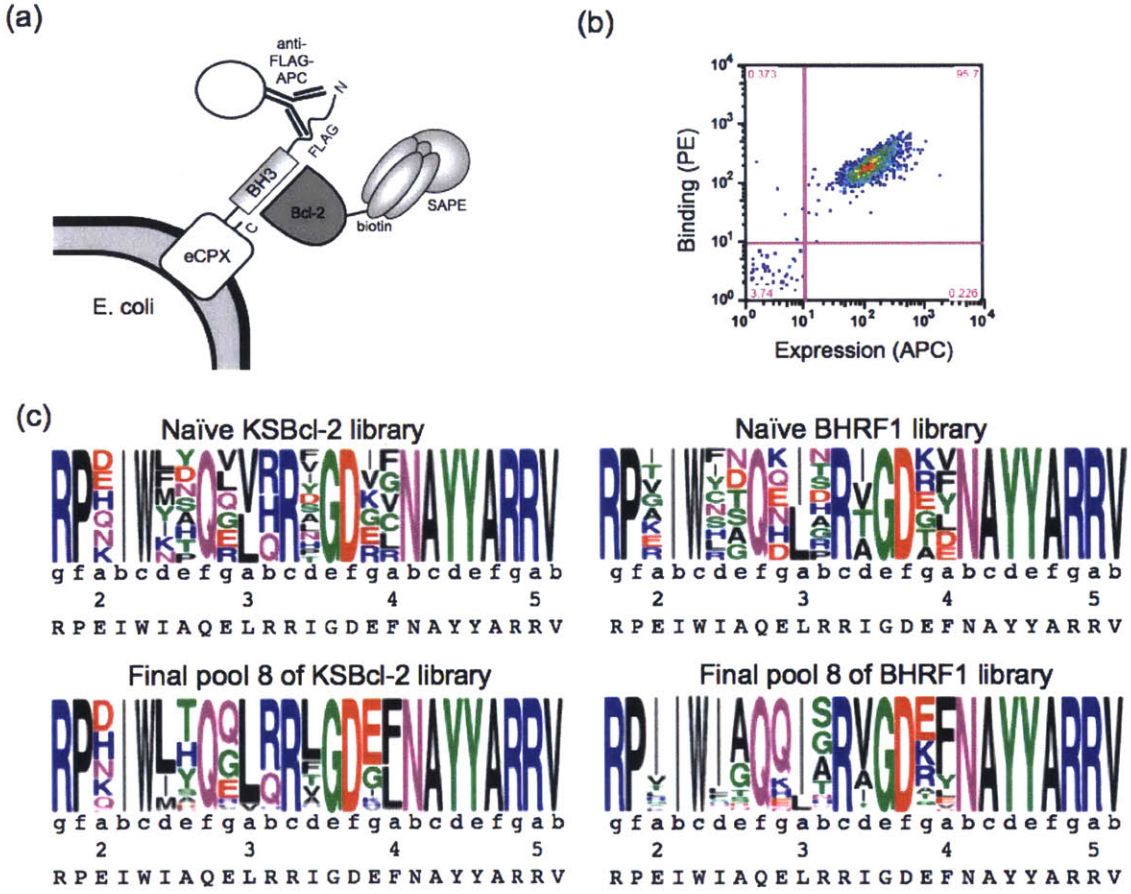


Figure 3.6. Bacterial surface display screen for selective binders of KSBcl-2 and BHRF1. (a) The libraries were displayed on the N-terminus of eCPX. A FLAG tag was displayed N-terminal to the peptide, and peptide expression was detected by an anti-FLAG antibody conjugated to APC. Streptavidin-phycoerythrin (SAPE) was used to detect binding of biotinylated Bcl-2 proteins. (b) Representative FACS plot of wild-type Bim BH3 displayed on the surface of *E. coli* binding to 5 nM biotinylated KSBcl-2. Binding, as reported by PE fluorescence, is plotted as a function of expression reported by APC fluorescence. The lower-left quadrant includes non-expressing cells. (c) Sequence logos built from deep sequencing of the naïve and eighth (final) library pools, using unique sequences. Heptad positions are indicated, and the wild-type Bim

residue is given at each position below the logo. Pool 8 sequences included in these logos were filtered to include only those sequences also present in pools 6 and 7.

Nine positions in Bim were mutated, and the theoretical libraries designed using our optimization protocol are shown in Figure 3.6c. The heptad register of Bim is given below each position. We chose to vary the four conserved hydrophobic positions 2d, 3a, 3d, and 4a, because the Bcl-2 homologs show strong, differential preferences for mutations at these positions (Boersma et al., 2008; Dutta et al., 2010b). We also included variation at position 2a, because the structural environments around this position in the viral receptors exhibit differences from the environment in Mcl-1. Specifically, Mcl-1 has an arginine that forms a salt bridge with glutamate at 2a in Bim that is lacking in the viral receptors (Fire et al., 2010). The four remaining positions chosen for design, 2e, 2g, 3b, and 3g, are generally occupied by small or polar/charged residues in native BH3 sequences, as they are on the sides of the BH3 helix and are more solvent exposed. The library designs are presented, along with an indication of which residues were included for affinity or specificity, in Tables 3.2 and 3.3.

Table 3.2 KSBcl-2 library design.

Position	Permitted residues	Required residues						Residues encoded
		Specificity vs.					Affinity to KSBcl-2	
		Mcl-1	Bfl-1	Bcl-x_L	Bcl-2	Bcl-w		
E2a	EDRHNAFM	H	H				EH	DEHKNQ[a] (VAM)
I2d	IRKHNQGLF YMSTAV	FYLM	FYM	FYLM	FYL M	FLM	IFYML	FIKLMNY stop (WWS)
A2e	ADEHKRNQS TGY	DHT	DHT	DHT	DHT	DHT	ADHT	ADHNPST Y (NMC)
E2g	ESGDRHNQT AVILFYWM	GL		GL	GL	G	EGL	EGLQRV (SDG)
L3a	LVIMR	V	V	V	V	V	LV	LV (STG)
R3b	RKHQSTAG	HQ	HQ	HQ	HQ	H	RHQ	HQR (CRK)
I3d	IKRHNQSTG AVLFYW	NLA	LA	A	LA	NL	INLA	ADFHILN PSTVY (NHC)
E3g	EKRHNQSTG AVIL	V		V	V		EV	EGIKRV (RDA)
F4a	FDEKRHNQS TGAVILYW	G	G	G	G		FG	CFGLRV (BKT)

[a] Disruptive residues encoded due to codon choice are underlined. None were enriched in the library screening.

Table 3.3. BHRF1 library design.

Position	Permitted residues	Required residues						Residues encoded
		Specificity vs.					Affinity to BHRF1	
		Mcl-1	Bfl-1	Bcl-x_L	Bcl-2	Bcl-w		
E2a	EDKRHNQTG AVILFYWM	I	I	I			EI	AEGIKRTV[a] (RNA)
I2d	IKRTAVLMY	RY	RY	RY	RY	RY	IRY	CFHILNRSY (HDC)
A2e	ADSG	DSG	D	DS	D	DG	ADSG	ADGNST (RVC)
E2g	EDKRHNQST GAVILFYWM	K	K	K	K		EK	DEHKNQ (VAM)
L3a	LVIFMR	I	I	I	I	I	LI	IL (MTT)
R3b	RKHNQSTGA W	HN	HN	HN	HN	HN	RHN	ADGHNPRST (VVC)

I3d	IAVLFTK	AV	A	AV	A		IAV	AITV (RYC)
E3g	EDKRNQSTG AWHV	AK		AK	AK	K	EAK	AEGKRT (RVG)
F4a	FENQSTAVIL YWG	E	E	E	E	E	FE	DEFLVY stop (KWK)

[a] Disruptive residues included due to codon choice are underlined. None were enriched in the library screening.

The eCPX *E. coli* surface display system has been used previously to display and sort BH3 peptide libraries (Zhang and Link, 2011; Zhang et al., 2012). We optimized it for affinity-based sorting by adding a FLAG-tag N-terminal to the BH3 peptide that can be used to detect the expression level of each library member (Figure 3.6a). A representative fluorescence activated cell sorting (FACS) plot illustrating the correlation between peptide cell-surface expression and binding to KSBcl-2 for a single clone is shown in Figure 3.6b. We sorted the designed libraries by FACS in a series of positive, negative, and competition sorts to identify high affinity and selective binders of the viral Bcl-2 proteins, as illustrated Figure 3.7 and detailed in the Materials and Methods. Human homologs Mcl-1, Bfl-1, and Bcl-w were used as competitors in the KSBcl-2 library sorting, and Mcl-1 and Bfl-1 were used for the BHRF1 library. Pools of clones isolated from rounds of sorting are designated KL1-KL8 for KSBcl-2 and BL1-BL8 or BL5'-BL8' for BHRF1 (see Methods and Figure 3.7). FACS analysis was performed on the naïve and final pools for both libraries (Figure 3.8). This analysis showed that although the naïve libraries contained more Mcl-1 binders than viral Bcl-2 binders, the final sorted library pools consisted of tight viral Bcl-2 binders with modest margins of specificity against Mcl-1, and larger margins of specificity against Bfl-1, Bcl-w, and Bcl-2.

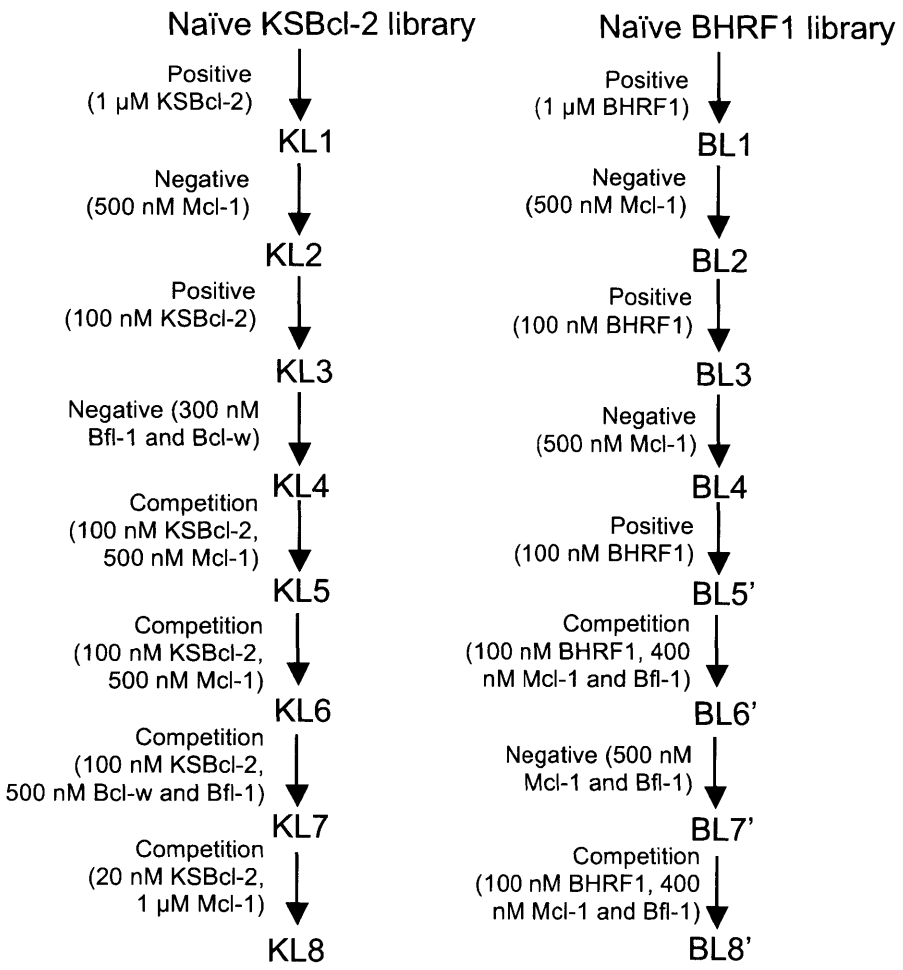


Figure 3.7. Library sorting scheme. The scheme presented here for BHRF1 is that used for deep sequencing. The BL6 and BL8 BHRF1 conventionally sequenced clones came from a different scheme detailed in the Methods.

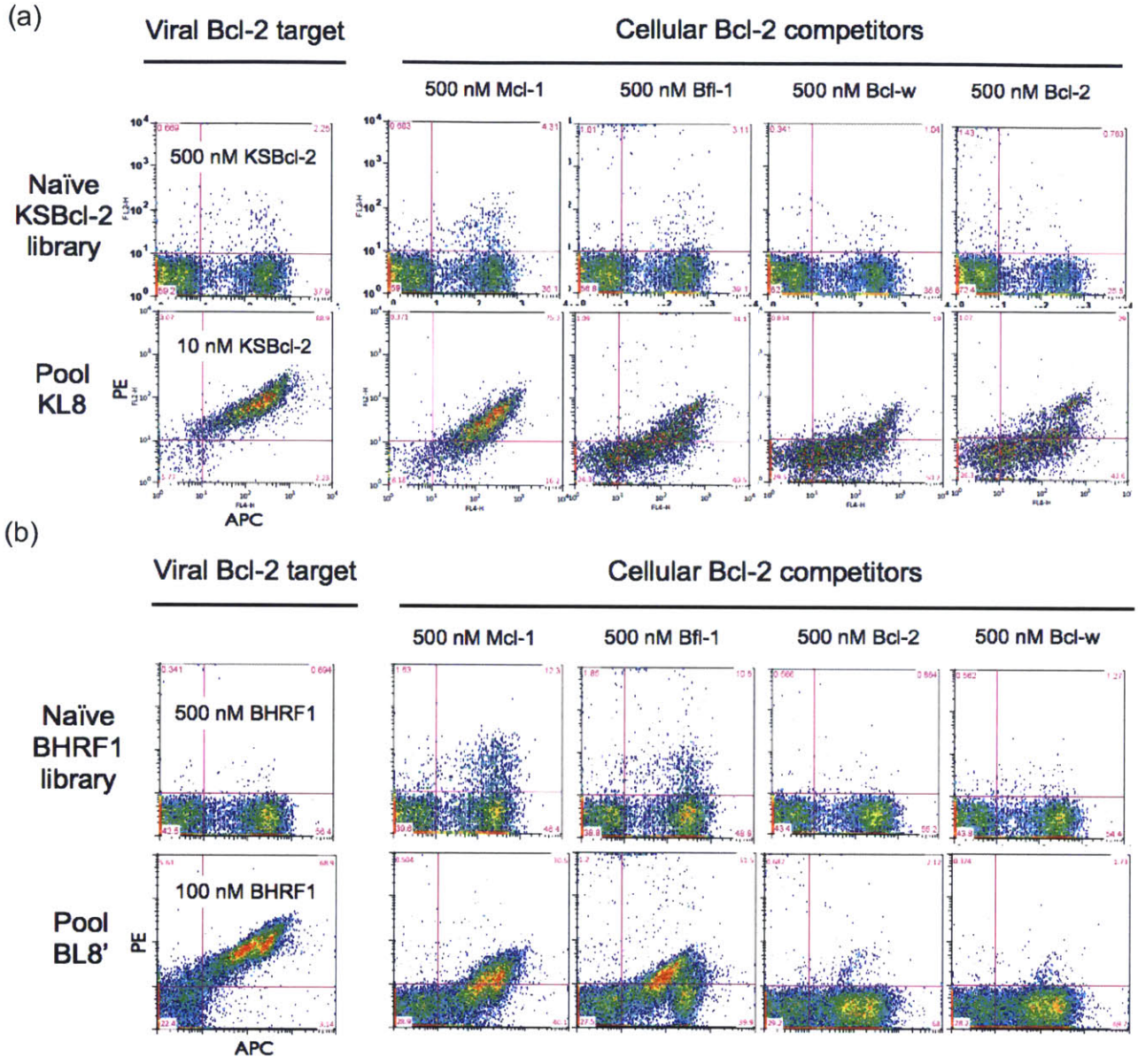


Figure 3.8. FACS plots showing the binding of the viral Bcl-2 protein and four human Bcl-2 proteins to the (a) KSBcl-2 naïve and final sorted library (KL8), and (b) the BHRF1 naïve and final sorted library (BL8').

To follow the progress of sorting and to isolate clones for further analysis, we sequenced clones from later library pools. From the KL6 and KL8 library pools, we sequenced 24 and 20 clones, respectively, and found a total of 16 unique sequences. Among four BHRF1 library pools from two different sorting schemes (see Materials and Methods), we obtained 32 unique

sequences. Peptide sequences and the frequencies with which they were found in the library pools are given in Table 3.4. To provide more sequences for analysis of library design success and mutation enrichment trends, the naïve libraries (KL0 and BL0) and pools 1, 3, 5, 6, 7, and 8 were sequenced using Illumina technology for each library. The sequence filtering process and number of reads for each pool are described in the Materials and Methods and in Table 3.5. Sequence data show that the library screens resulted in distinct residue preferences at select positions, as shown by logos constructed from unique, filtered sequences in the eighth (final) pools (Figure 3.6c). Many of the trends observed support the predictions of the SPOT arrays ($PSSM_{SPOT}$ models) and the STATIUM models. For example, at position 2a, the wild-type glutamate residue was selectively lost from the KSBcl-2 library, while the other five possible mutations were maintained. And in the BHRF1 library, isoleucine was enriched at 2a. Position 2a was not mutated on the SPOT arrays, but the STATIUM models predicted that most mutations from glutamate at 2a – including isoleucine – would provide specificity for binding either of the viral Bcl-2 receptors over Mcl-1, and our library results validate the STATIUM predictions. In the BHRF1 library screen, there was a strong preference for isoleucine over leucine at 2d and 3a. The SPOT arrays indicated that BHRF1 strongly prefers the wild-type isoleucine at 2d, and can tolerate both isoleucine and leucine at 3a. The preference for isoleucine over leucine at 3a in the BHRF1 library results (Figure 3.6c) may arise in part because this substitution provides specificity over some of the human receptors, as described in a later section. At peptide positions along the sides of the BH3 binding groove, such as 2e and 3b, the SPOT arrays also accurately reflected preferences. For example, the SPOT arrays indicated that KSBcl-2 is more tolerant of a variety of residues at 2e than are Bcl-x_L, Bcl-2, and Bfl-1, whereas

BHRF1 has greater tolerance at 3b than this same group of receptors, and this was reflected in the library results.

Table 3.4. Conventionally sequenced clones and their frequencies

Clone name	Sequence	Number of copies
KL6-1	RPDIWLYQQLRRLGDEFNAYYARRV	25
KL6-5	RPHIWNTQVLRRLGDELNAYYARRV	1
KL6-7	RPQIWIHQGLQRLGDGLNAYYARRV	1
KL6-8	RPNIWITQELRHLGDRFNAYYARRV	1
KL6-9	RPHIWLTQGLRRLGDEFNAYYARRV	2
KL6-10	RPHIWISQGLRRLGDEFNAYYARRV	1
KL6-14	RPKIWITQGLQRFGDELNAYYARRV	1
KL6-16	RPNIWIPQGLQRLGDRLNAYYARRV	1
KL6-17	RPHIWLTQQLRRTGDGFNAYYARRV	1
KL6-18	RPKIWLPQGLQRLGDELNAYYARRV	2
KL6-23	RPKIWLYQQLHRLGDELNAYYARRV	1
KL8-6	RPQIWLTQQLRRLGDEFNAYYARRV	1
KL8-7	RPNIWLTQRLQRLGDEFNAYYARRV	1
KL8-16	RPNIWLTQQLQRLGDEFNAYYARRV	1
KL8-18	RPNIWITQGLRRLGDEFNAYYARRV	3
KL8-19	RPHIWMHQGLRRLGDELNAYYARRV	1
BL6'-9	RPIIWIAQQISRVGDAENAYYARRV	1[a]
BL6'-10	RPIIWIAQQIARVGDRFNAYYARRV	2
BL6'-15	RPIIWIAQQIGRVGDRENAYYARRV	1
BL6'-16	RPIIWIAQQISRAGDEFNAYYARRV	1
BL7'-2	RPAIWSNQDLARVGDRLNAYYARRV	1
BL7'-6	RPIIWIAQQIARVGDGFNAYYARRV	1
BL7'-16	RPIIWIAQQIGRVGDKENAYYARRV	1
BL7'-21	RPVIWIAQQISRAGDEFNAYYARRV	1
BL7'-23	RPIIWIAQQISRVGDKFNAYYARRV	1
BL6-1	RPVIWHGQHISRVGDAENAYYARRV	2
BL6-3	RPIIWIGQQIARVGDRFNAYYARRV	2
BL6-4	RPGIWLSQKIDRVGDRYNAYYARRV	1
BL6-6	RPVIWIAQQIGRVGDKENAYYARRV	1
BL6-7	RPGIWYGQKIGRVGDGENAYYARRV	1
BL6-8	RPIIWIGQQISRIGDKFNAYYARRV	1
BL6-9	RPIIWIAQQISRVGDGFNAYYARRV	1
BL6-11	RPIIWIAQQITRVGDEFNAYYARRV	3
BL6-14	RPTIWYSQELSRIGDKYNAYYARRV	1
BL6-15	RPIIWIAQQITRVGDRFNAYYARRV	1
BL6-16	RPVIWIAQQISRVGDRFNAYYARRV	1
BL6-18	RPTIWIGQQIARAGDRFNAYYARRV	1
BL6-19	RPIIWIAQQIGRVGDEFNAYYARRV	19

BL6-20	RPIIWIAQQISRVGDRFNAYYARRV	7
BL6-22	RPIIWIGQEISRVGDRENAYYARRV	1
BL6-23	RPIIWIAQKIARVGDEFNAYYARRV	2
BL6-24	RPVIWIAQQIARVGDKFNAYYARRV	2
BL8-1	RPVIWIAQEIARVGDAENAYYARRV	5
BL8-3	RPVIWLSQKLGRTGDGYNAYYARRV	3
BL8-7	RPIIWIAQQIARVGDEFNAYYARRV	1
BL8-8	RPIIWIAQQISRVGDEFNAYYARRV	14
BL8-9	RPIIWIGQQISRVGDEFNAYYARRV	2
BL8-11[b]	RPIIWIAQQIGRVGDRFNAYYARRV	38

[a]Number of copies for BHRF1 library clones is the sum of all copies found in any of the four conventionally sequenced BHRF1 library pools.
[b]BL8-11 was the clone with the growth advantage that took over the first library sorting attempt.

Table 3.5. Number of peptide sequences from Illumina sequencing of library pools

pool	Sequences passing quality score filtering	Sequences without stop codons	Unique sequences	Unique sequences contained in previous 2 pools[a]
KL0	12,342,360	10,568,961	2,616,343	
KL1	6,606,374	4,211,436	935,875	546,746
KL3	8,635,179	8,398,685	467,212	138,959
KL5	8,140,308	8,097,593	117,603	60,879
KL6	4,381,973	4,372,437	32,482	24,942
KL7	4,793,587	4,789,428	12,772	8,632
KL8	3,487,178	3,483,067	9,119	4,262
BL0	10,984,564	10,083,190	3,694,041	
BL1	6,771,022	6,013,775	714,775	456,405
BL3	8,041,691	7,971,975	342,085	107,703
BL5	8,965,239	8,945,462	76,932	33,827
BL6	3,774,705	3,765,149	13,496	8,891
BL7	4,346,314	4,342,741	11,575	5,808
BL8	4,497,084	4,494,624	2,900	1,577

[a] For KL1 and BL1 pools, the unique sequences were filtered to contain sequences also present in the naïve pools, KL0 and BL0.

To evaluate the overall agreement between predictions made by the models used in our library design protocol and the library enrichment results, we compared the model scores for all point mutations with the enrichment of individual mutations during screening (Figure 3.9). The SPOT arrays were converted to position-specific scoring matrices (PSSMs) by taking the logarithm of the mutant intensity divided by the average wild-type intensity (giving $PSSM_{SPOT}$ models) (DeBartolo et al., 2012; Dutta et al., 2010b). Enrichment was measured as $\log_2$(frequency of mutation in unique sequences in final pools KL8 or BL8'/frequency in the naïve pools KL0 or BL0). Enrichment is indicated by color in Figure 3.9, with enriched mutations in shades of orange to red. Scores for the wild-type Bim residues are indicated by gray lines on the plots in Figure 3.9, such that points that fall in the lower right quadrant of each plot correspond to point mutations predicted by the models to provide enhanced affinity for the target viral Bcl-2 and specificity against Mcl-1. Many of the most enriched mutations in each library fall into this quadrant for the $PSSM_{SPOT}$ models (Figure 3.9a and 3.9c). For example, 83% of the enriched mutations for KSBcl-2 ($\log_2$(frequency in final pool/frequency in naïve pool) > 0) fall in this quadrant, as do 58% of the enriched mutations for BHRF1. The STATIUM model predictions did not agree as well with the screening results, but enriched residues had better viral protein binding scores than non-enriched residues: 86% of the mutations with $\log_2$(frequency KL8/KL0) > 0 had $STATIUM_{KSBcl\text{-}2}$ scores > -3, whereas only 59% of un-enriched residues had scores this high. For BHRF1, 73% of mutations with $\log_2$(frequency BL8'/BL0) > 0 had $STATIUM_{BHRF1}$ scores > -3, compared to 56% of un-enriched residues. The high correlation of STATIUM scores for KSBcl-2 or BHRF1 binding with scores for Mcl-1 binding reflects the predicted difficulty of distinguishing interactions between the viral proteins and Mcl-1. This is

also seen, to a lesser degree, in the SPOT results and is consistent with our experimental observations.

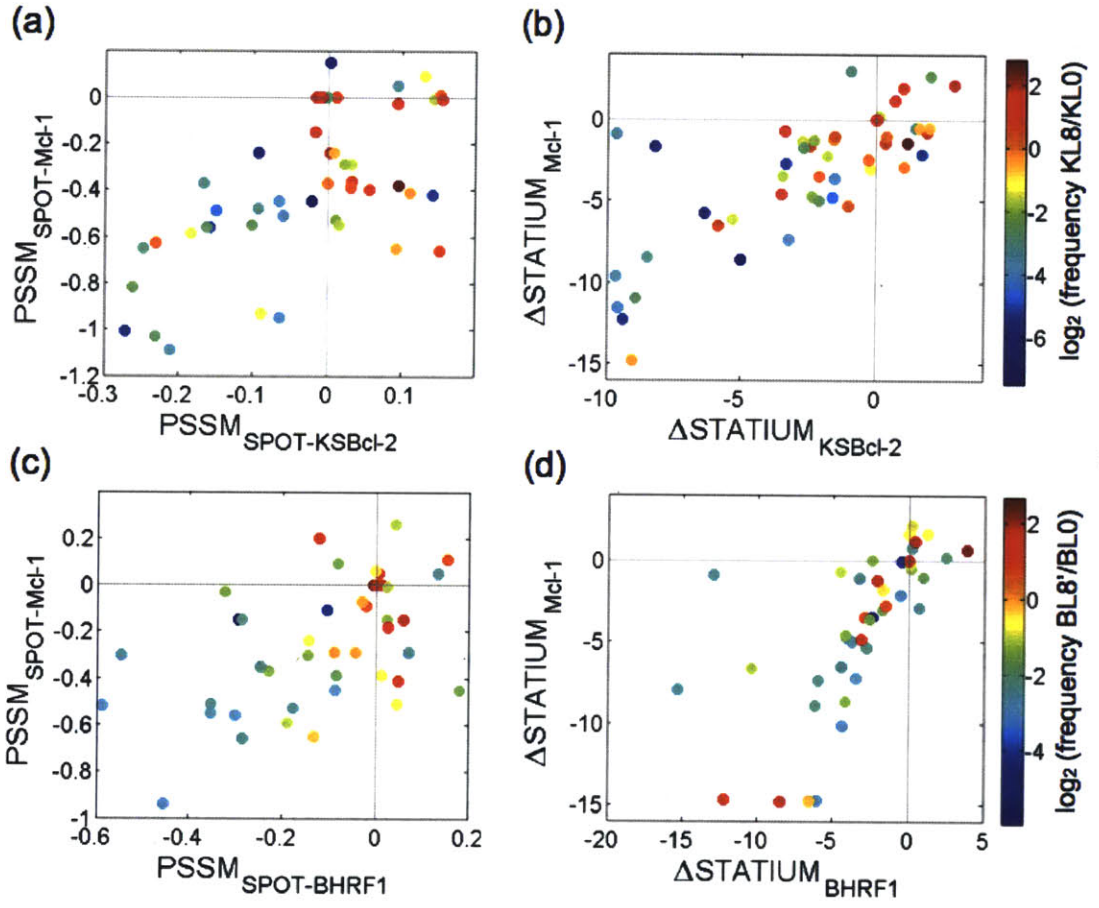


Figure 3.9. Enrichment of residues in library sequences versus viral Bcl-2 and Mcl-1 model scores. Each point represents a residue and is colored according to enrichment ($\log_2$ (frequency in final pool/frequency in naïve pool)). The frequencies were calculated from unique sequences filtered as described in the Materials and Methods. Residues in the KSBcl-2 library (a, b) and BHRF1 library (c, d) are plotted by (a, c) $PSSM_{SPOT}$ scores and (b, d) ΔSTATIUM scores. For all models, a score of zero is given for the wild-type residue and scores greater than zero indicate tighter binding than wild type.

Binding of library-derived peptides to 8 Bcl-2 homologs

We selected 22 peptides that sample a variety of sequence features to test in direct fluorescence anisotropy binding experiments. Peptide names designate the library ('KL' or 'BL') followed by the sorting pool from which they were sequenced, followed by an arbitrary

numerical designation. References to KL and BL peptides refer specifically to the peptides tested in solution binding experiments. The peptide names with alphabetical designations were chosen from the deep sequencing results, either because they represented a consensus sequence or because they contained enriched substitutions not sampled in the conventionally sequenced clones. The 22 peptides were tested for binding to all 5 human and 3 viral receptors.

A heat map of the K_D values is shown in Figure 3.10, clustered by receptor binding profile similarity (K_D values with 95% confidence intervals are given in Table 3.6). The binding patterns of peptides identified from the two libraries have a clear distinction. Most of the peptides identified by screening for binding to BHRF1 also bound to KSBcl-2, M11 and Mcl-1. In contrast, peptides identified on the basis of binding to KSBcl-2 only bound strongly to KSBcl-2 and Mcl-1. Despite our use of competitive screening against Mcl-1, it is notable that all of the peptides that bound to KSBcl-2 or BHRF1 also bound tightly to Mcl-1. Binding to Bcl-2, Bcl-x_L, and Bcl-w was substantially weaker for all of the peptides, consistent with the distinct sequence and structural properties of these receptors. Binding to Bfl-1 was universally very weak for peptides from both libraries, although Bfl-1 bound tightly to positive control peptides. This trend is especially interesting given the similarity between the native BH3 binding profiles of BHRF1 and Bfl-1.

Table 3.6. Dissociation constants (nM) and 95% confidence intervals for the library peptides and mutants.

	BHRF1	KSBcl-2	Mcl-1	Bfl-1	Bcl-w	Bcl-x_L	Bcl-2	M11
BL8-9	3.3 ± 1.6	4.2 ± 0.84	29 ± 4.6	>1000[a]	400 ± 170	180 ± 91	390 ± 110	7.6 ± 2.0
BL6-3	~4.8	5.2 ± 1.3	11 ± 1.3	600 ± 280	260 ± 53	510 ± 300	470 ± 180	1.1 ± 0.50
BL6-4	>1000	>1000	>1000	>1000	>1000	>1000	>1000	>1000
BL6-11	<1	2.0 ± 0.37	2.4 ± 0.62	610 ± 200	110 ± 22	55 ± 8.5	290 ± 99	2.3 ± 0.84

BL6*-16	~1.1	3.4 ± 0.44	5.2 ± 1.0	710 ± 590	11 ± 1.3	17 ± 3.0	14 ± 2.2	6.6 ± 1.3
BL6-19	<1	2.4 ± 0.46	5.3 ± 0.91	380 ± 140	72 ± 15	120 ± 26	390 ± 100	0.83 ± 0.53
BL6-22	2.2 ± 1.3	1100 ± 860	45 ± 6.1	>1000	>1000	>1000	>1000	>1000
BL6-23	<1	1.6 ± 0.41	2.8 ± 0.56	390 ± 180	110 ± 22	120 ± 17	55 ± 8.3	0.99 ± 0.36
BL_a	<1	1.6 ± 0.47	1.9 ± 0.29	690 ± 450	79 ± 11	120 ± 40	360 ± 150	1.5 ± 0.38
BL_b	<1	180 ± 31	10 ± 1.2	>1000	>1000	>1000	>1000	76 ± 17
BL_c	1.9 ± 1.2	27 ± 2.2	3.4 ± 0.99	>1000	>1000	>1000	>1000	130 ± 28
BL_d	1.6 ± 1.2	8.5 ± 1.7	15 ± 2.7	420 ± 250	620 ± 170	~2000	~2000	<1
KL6-1	>1000	1.1 ± 0.29	3.1 ± 1.5	~1200	10 ± 3.5	19 ± 5.5	19 ± 5.8	>1000
KL6-5	600 ± 340	<1	0.77 ± 0.72	~1000	110 ± 27	350 ± 130	~4400	>1000
KL6-7	>1000	1.6 ± 0.45	22 ± 2.8	>1000	>1000	>1000	>1000	>1000
KL6-14	>1000	0.62 ± 0.20	1.5 ± 0.44	>1000	~3800	>1000	>1000	>1000
KL6-18	>1000	0.72 ± 0.40	7.1 ± 1.7	>1000	500 ± 170	~720	>1000	>1000
KL8-7	~650	<1	21 ± 3.7	>1000	40 ± 9.2	100 ± 38	46 ± 16	77 ± 30
KL8-16	>1000	0.48 ± 0.3	3.9 ± 1.5	>1000	28 ± 8.0	83 ± 23	120 ± 36	200 ± 88
KL8-18	~440	0.52 ± 0.20	6.7 ± 1.2	>1000	110 ± 30	43 ± 8.5	160 ± 73	~3600
KL_a	~375	1.7 ± 0.54	4.7 ± 1.2	600 ± 340	910 ± 660	~3000	>1000	>1000
KL_b	>1000	1.1 ± 0.39	4.3 ± 0.83	>1000	>1000	>1000	>1000	>1000
KL6-7 _G3eC[b]	>1000	1300 ± 1100	~16000	900 ± 780	>3000	>3000	>3000	>3000
KL6-7 _Y4eK	>1000	2.9 ± 1.3	~7800	>3000	>3000	>3000	>3000	>3000
KL6-7 _D3fA	>1000	180 ± 80	>5000	>3000	>3000	>3000	>3000	>3000
KL6-7 _I2dR[d]	>1000	400 ± 110	3100 ± 2000	>3000	>3000	>3000	>3000	>3000

KScomp	>1000	13 ± 7.3	510 ± 190	>3000	1200 ± 1000	~8900	~10000	~560
BL6-22_G3eC[b]	~1800	>1000	>5000	>3000	>3000	>3000	>3000	>3000
BL6-22_Y4eK	30 ± 17	750 ± 300	~2200	>3000	>3000	>3000	>3000	>3000

[a] For values designated >1000 (or 3000 or 5000) there was no binding observed, and the number given is the highest concentration of receptor used for that curve, in nM. Values designated as <1 were too tight to be fit. Values designated as approximate had large 95% confidence intervals either because they were very weak, or for the BHRF1 curves, because the dynamic range of anisotropy values was low.

[b] The G3eC mutation was predicted to provide specificity for KSBcl-2 and BHRF1 over Mcl-1 based on sequence similarity of helix 5 between the viral Bcl-2 proteins and Bcl-2/x_L/w, which can bind peptides with larger side chains at 3e. This mutation proved to have a similarly negative effect on viral Bcl-2 and Mcl-1 binding.

[c] The I2dR mutation was predicted to provide specificity for KSBcl-2 over Mcl-1 based on SPOT array binding and sequence differences between the receptors near this position. This mutation proved to have a similarly negative effect on KSBcl-2 and Mcl-1 binding.

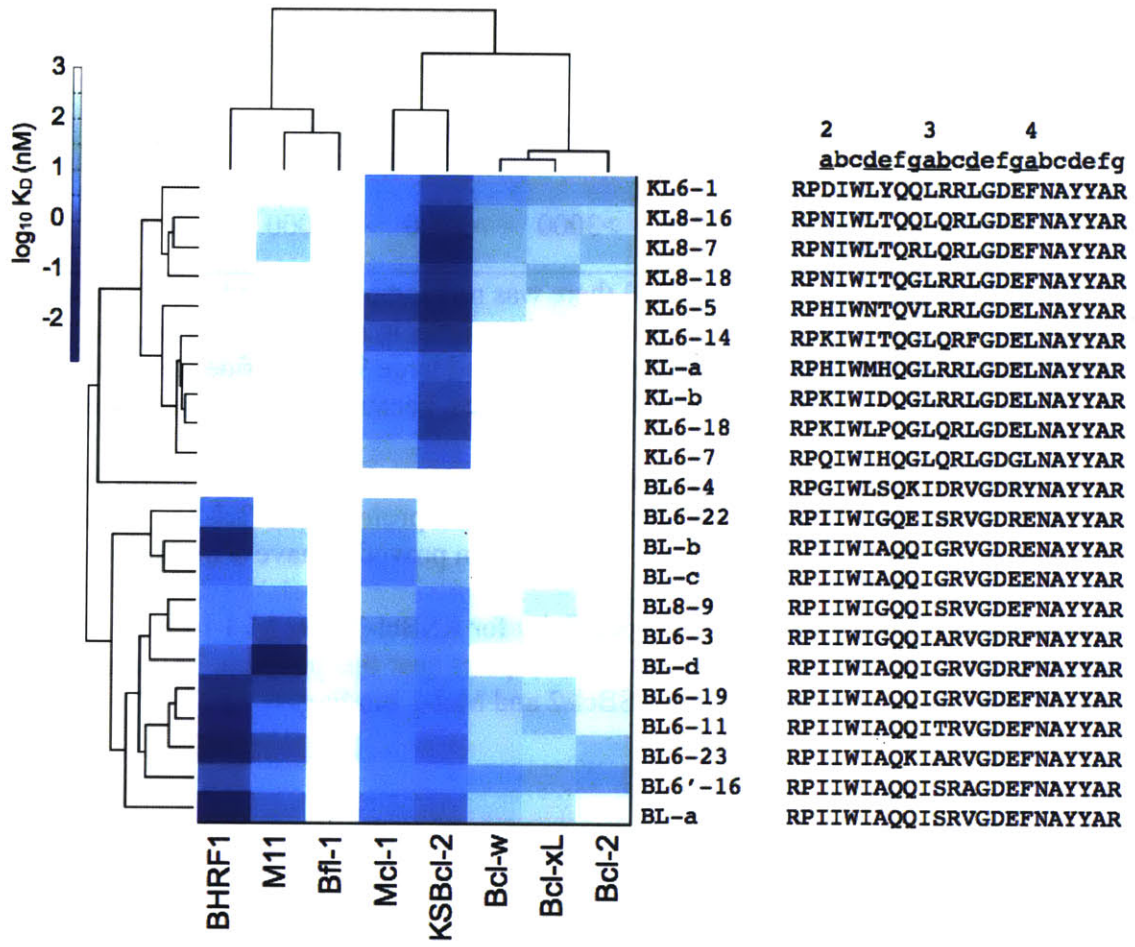


Figure 3.10. Binding of human and viral Bcl-2 homologs to peptides identified from library screening. K_D values, measured by fluorescence anisotropy, here plotted as $\log_{10}$ (K_D in nM), are given with 95% confidence intervals in Table 3.6. White indicates no binding up to 1000 nM. Library peptide sequences are shown to the right, with the heptad register indicated above and with varied positions underlined. Bcl-2 protein receptors (in columns) were clustered by the correlation of their binding profiles and peptides (in rows) were clustered by Euclidean distance.

Specificity mechanisms underlying receptor binding similarity patterns

We analyzed the binding patterns observed for different Bcl-2 proteins by integrating data from the SPOT arrays, the SiteMAP structural analysis and solution binding studies of peptides and selected point mutants. Our analyses suggested mechanisms by which the library peptides achieve specificity. These mechanisms will be discussed on three levels. First, we discuss peptide sequence features that distinguish binding to the three viral homologs KSBcl-2,

BHRF1, and M11. Second, we address sequence and structural features that differentiate peptides that bind to the KSBcl-2/Mcl-1/BHRF1 vs. Bcl-x_L/2/w/Bfl-1 groups. Finally, we discuss mutations that contribute to the modest level of specificity achieved by some peptides for binding to the viral Bcl-2 proteins over Mcl-1.

To investigate the origins of binding differences between the viral Bcl-2 homologs, we tested residue substitutions in Bim BH3 that were common in the peptides identified from the libraries. Threonine was enriched at position 2e in the KSBcl-2 library (Figure 3.6c), and the specificity of Bim_A2eT for KSBcl-2 and Mcl-1 has been addressed in prior work (Foight et al., 2014). Our nomenclature for mutants lists the parent peptide (Bim) followed by the wild-type residue that was mutated (A), the heptad position (2e) and the new residue identity at that site (T). The alanine-to-threonine mutation in Bim_A2eT weakens binding to Bfl-1, Bcl-x_L, Bcl-2 and Bcl-w (Foight et al., 2014). Here, we show that this mutation also weakens BHRF1 and M11 binding (Table 3.7). Another peptide position that can favor KSBcl-2 binding is 2g. In contrast to the BHRF1-binding peptides in Figure 3.10, which have glutamine, lysine or glutamate at position 2g, many of the KSBcl-2-binding peptides that we analyzed have a glycine at this site. BHRF1 binding is weakened dramatically by the two mutations in Bim_A2eT_E2gG, and 60% of the KL peptides tested combine the E2gG mutation with a larger residue (proline, threonine, aspartate, or histidine) at 2e. This likely explains why BHRF1 shows limited binding to many of the KSBcl-2 library peptides. As for BHRF1 specificity determinants, serine at position 3b is common in the BHRF1 library peptides. Bim_R3bS binds tightly to BHRF1 but the serine mutation weakens binding to Bfl-1, Bcl-w, Bcl-2, and M11 by ~4-9-fold when introduced into Bim BH3 (Table 3.7), providing a partial explanation of the BHRF1 library peptide specificity. Notably, many of the BHRF1 library peptides bind more tightly to M11 than does wild-type Bim

BH3, suggesting that even though R3bS weakens M11 binding, other mutations present in the BL peptides must compensate for this effect.

Table 3.7. Dissociation constants for Bcl-2 homologs binding to mutants of Bim BH3.

	K_D (nM) [a]			
	Bim	Bim_R3bS	Bim_A2eT	Bim_A2eT_E2gG
BHRF1	1.2 ± 0.74	~1.2[b]	64 ± 37	>1000[c]
KSBcl-2	1.2 ± 0.79	1.5 ± 0.57	1.8 ± 0.81	1.1 ± 0.49
Mcl-1	<1 [c]	<1 [c]	0.62 ± 0.55	1.6 ± 0.57
Bfl-1	<1 [c]	5.2 ± 1.7	32 ± 6.9	260 ± 64
Bcl-w	2.6 ± 2.1	18 ± 5.3	40 ± 9.3	98 ± 35
Bcl-x_L	2.6 ± 1.9	3.8 ± 1.4	18 ± 4.9	42 ± 7.1
Bcl-2	1.9 ± 1.3	7.2 ± 3.1	44 ± 3.1	150 ± 27
M11	23 ± 10	210 ± 160	78 ± 19	15 ± 5.3

[a] Values are given ± 95% confidence intervals.
[b] An approximate value is given where the confidence interval was very large, which was a problem for some BHRF1 curves that had a low dynamic range for the anisotropy signal.
[c] Values designated >1000 or <1 nM were too weak or tight to quantify using this assay.

The four BH3 positions that are conserved as hydrophobic (2d, 3a, 3d and 4a) have been demonstrated to make strong contributions to human Bcl-2 specificity patterns (Boersma et al., 2008; Dutta et al., 2010b; Fire et al., 2010; Lee et al., 2008). Here, we observed that these positions are important for differentiating binding to KSBcl-2/Mcl-1/BHRF1 vs. Bcl-x_L/2/w/Bfl-1. The SPOT arrays indicated that viral proteins BHRF1 and KSBcl-2, and human proteins Mcl-1 and Bcl-w, are broadly tolerant of substitutions at position 4a (Figure 3.11a). We observed the same trend in library screening. Several of the tested KL peptides had an F4aL substitution, and an F4aE mutation was found in a minority of the BL peptides. Notably, Bcl-x_L and Bcl-2 showed no binding to these peptides (Figure 3.10), in keeping with the observation made in several past studies that Bcl-x_L and Bcl-2 have a more enclosed 4a pocket that favors aromatics of complementary size and shape (Dutta et al., 2010b; Fire et al., 2010; Lee et al., 2007). Human

Mcl-1 and viral KSBcl-2, M11, and BHRF1 appear to have similar preferences at hydrophobic position 3a. The BHRF1 library peptides showed a strong preference for isoleucine at position 3a, where most natural BH3 peptides have a leucine. The Bim substitution SPOT arrays indicate that BHRF1, KSBcl-2 and Mcl-1 exhibit no loss in affinity for the L3aI mutation, whereas Bfl-1, Bcl-x_L, Bcl-2, and Bcl-w binding is weakened (Figure 3.11a). The shape of the 3a pocket is influenced by two positions in helices 4 and 5 that are underlined in Fig. 1c. KSBcl-2, Mcl-1 and BHRF1 have either a β-branched isoleucine or valine, or a long but flexible methionine on helix 4, where the other receptors have glutamate or leucine, which leave less space for a branched Cβ at peptide position 3a. The extra space for a branched Cβ can be seen in KSBcl-2, BHRF1, and Mcl-1 hydrophobic SiteMAPs (Figure 3.11b). M11 also has a leucine on helix 4, but has a unique glycine on helix 5, which may allow more space for a branched Cβ than is provided by the alanine or threonine found in the other receptors. Thus, preferences at two of the conserved hydrophobic positions, 4a and 3a, distinguish the binding properties of the library peptides, with KSBcl-2, BHRF1, Mcl-1, and M11 exhibiting broader tolerance for residue substitution than Bfl-1, Bcl-x_L, and Bcl-2.

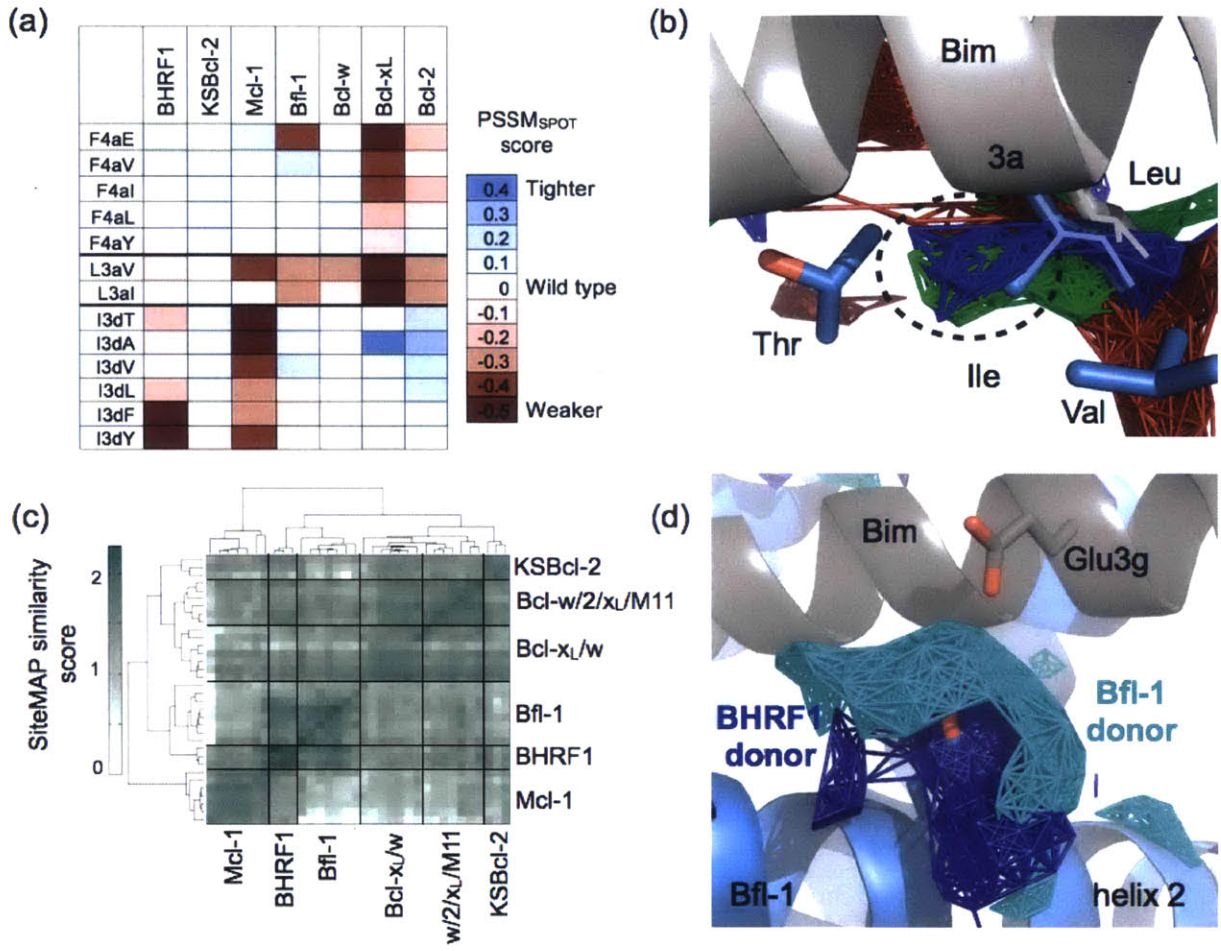


Figure 3.11. Specificity mechanisms employed by library peptides. (a) $PSSM_{SPOT}$ scores for substitutions at Bim peptide positions L3a, I3d, and F4a. (b) SiteMAP hydrophobic density near position 3a shows extra density (circled) for KSBcl-2, BHRF1, and Mcl-1 (red, green, and blue, respectively). Bim with leucine at position 3a (2PQK)(Fire et al., 2010) is in gray, and isoleucine at 3a from an Mcl-1-specific peptide (3KZ0)(Dutta et al., 2010b) is shown in blue. The threonine on helix 5 and valine on helix 4 from 3KZ0 are also shown in blue. (c) SiteMAP similarity score for the region around position 3d, clustered by structure. Black lines divide clusters, and Mcl-1 structures form a cluster distinct from the other receptors. (d) SiteMAPs for BHRF1 and Bfl-1 have significant donor density near the peptide 3g position.

Our libraries were designed and screened for peptides that would bind to the viral proteins in preference to all human proteins including Mcl-1. Specificity over Bcl-x_L, Bcl-2, Bcl-w, and Bfl-1 was readily achieved, but the library peptides showed only modest specificity (2- to 20-fold) over Mcl-1. We traced much of the viral protein vs. Mcl-1 binding specificity to

hydrophobic interactions at positions 3d and electrostatic interactions at positions 2g and 3g. Most KSBcl-2 binders that we characterized had leucine at 3d, and most BHRF1 binders had valine. SPOT arrays indicate that Bcl-x_L, Bcl-2, Bcl-w, Bfl-1, and KSBcl-2 are highly tolerant of mutations at 3d, whereas BHRF1 prefers valine and tolerates alanine (Figure 3.11a). However, Mcl-1 is highly intolerant of any mutation of I3d according to both the SPOT array and a previous BH3 peptide library sorted for Mcl-1 specificity (Figure 3.11a) (Dutta et al., 2010b). Comparing the region surrounding position 3d using SiteMAP, we found that Mcl-1 structures form a cluster separate from the other homologs (Figure 3.11c). The structural mechanism behind the position 3d tolerance exhibited by Bcl-x_L may be related to the multiple different conformations accessible at the helix 2-3 bend (Figure 3.12). These conformations are stabilized by different arrangements of a cluster of three aromatic residues highlighted in red in Figure 3.1c, and they result in different environments for the peptide side chain at position 3d in different structures. These three aromatic residues form a motif that is also present in Bcl-2, Bcl-w, and M11. In contrast to Bcl-x_L, all solved Mcl-1 structures have a similar conformation in the helix 2-3 bend. Until more structures are solved of KSBcl-2 and BHRF1 bound to diverse peptides it will remain unclear how they accommodate residues other than isoleucine at 3d. However, the different degrees of tolerance at 3d between KSBcl-2 (very tolerant), BHRF1 (moderately tolerant), and Mcl-1 (very intolerant) may arise from the different residues at the structural motif between helices 2 and 3 (L/K/F for KSBcl-2, I/N/F for BHRF1 and V/H/F for Mcl-1, in the same positions as the aromatic residues in the aforementioned Bcl-x_L motif; see Figure 3.1).

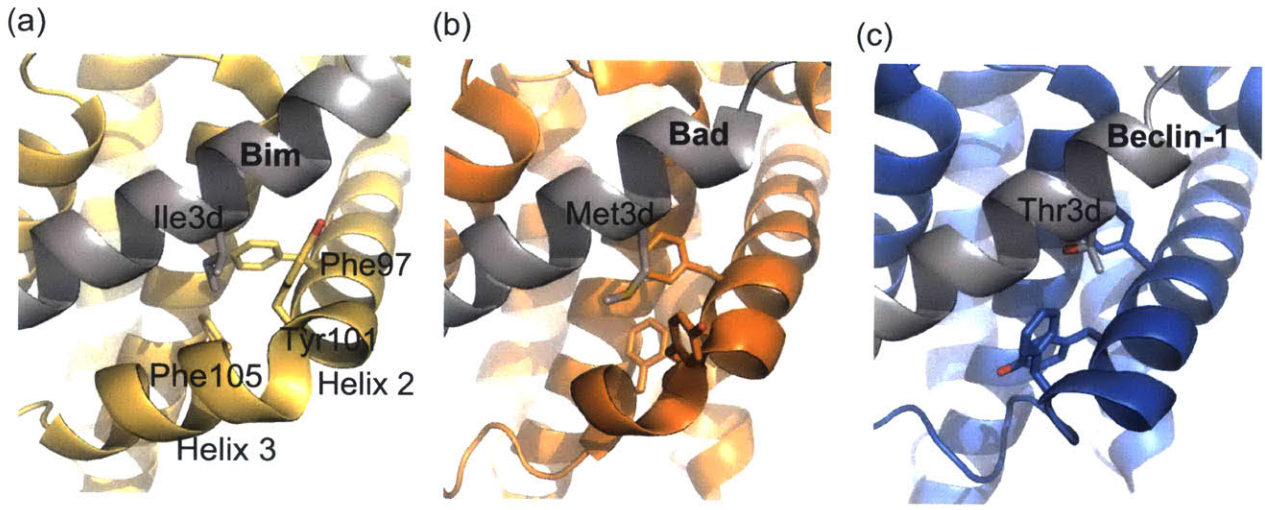


Figure 3.12. Conformational flexibility at Bcl-x_L helix 2 - helix 3 bend correlates with tolerance at peptide position 3d. Structures of Bcl-x_L bound to three different BH3 peptides show different conformations of helices 2 and 3 stabilized by different arrangements of three aromatic residues. (a) Bcl-x_L structures bound to Bim (3FDL (Lee et al., 2009c), yellow, I3d) and (b) Bad (1G5J (Petros et al., 2000), orange, M3d) have a short helix 2 and similar arrangements of the aromatic residues, with Phe105 buried in the groove, and Tyr101 extended out toward the solvent, but have different conformations of helix 3. (c) Bcl-x_L bound to Beclin-1 (2PON (Feng et al., 2007), blue, T3d) has a longer helix 2, with Tyr101 flipped into the groove and Phe105 oriented out of the groove. All peptides are in gray with the side chain of position 3d shown.

Charge patterns at positions 2g and 3g also contribute to selectivity for viral proteins over Mcl-1. For example, KL8-7 and KL8-16 both bind with K_D < 1 nM to KSBcl-2, but KL8-7 has a K_D of 21 nM for Mcl-1, whereas KL8-16 has a K_D of 3.9 nM. The only difference between these peptides is an arginine at position 2g in KL8-7 versus a glutamine in KL8-16 (Table 3.6). Likewise, mutation of E3g (the wild-type residue in Bim) to arginine in BL peptides reduced affinity to Mcl-1 by approximately 3-fold. By comparing BL6-19 (E3g) to BL_d (R3g), or BL_c (E3g) to BL_b (R3g), it can be seen that only BHRF1, Bfl-1, and M11 tolerate R3g without a significant loss of affinity (Table 3.6; the glutamate to arginine mutation is the only difference between these peptide pairs). Consistent with this, BHRF1 and Bfl-1 are the only receptors that show large patches of donor density proximal to position 3g in their SiteMAPs (Figure 3.11d).

Designed peptides with improved specificity for KSBcl-2 and BHRF1 vs. Mcl-1

To increase the specificity of our designed peptides against Mcl-1, we considered positions beyond the 9 that were varied in our libraries. We searched for differences between the viral homologs and Mcl-1 in sequence and structure, and mined our binding data and the literature for observations that could guide further optimization. Based on this analysis, we made additional mutations at C-terminal positions 3f and 4e that improved specificity, as described below.

(a)

Peptide	Sequence	Fold specificity (Mcl-1/ KSBcl-2 K_D)
TRPM7	FERVEQMCIQIKEVGDRVNYIKRSLQ	10
DDX4	FSKREKLVEILRNIGDERTMVFVETK	10
TXNDC11	TRELQELARKLQELADASENLLTENT	15
POFUT2	TRRSMVFARHLREVGDEFRSRHLNST	~16
MCF2L2	ADAIRPRCVELRHLCDDFINGNKKKW	~17
c6orf222	DAIIQMIVELLKRVGDQWEEEQSLAS	~24
TRIM58	KSRLVQQSKALKELADELQERCQRPA	~25
MINA	TVATRRLSGFLRTLADRLEGTKELLS	63
NBEAL2	AELRLFLAQRLRWLCDSCPASRATCV	84
	abcdefgabcdefgabcdefg	
	2 3 4	

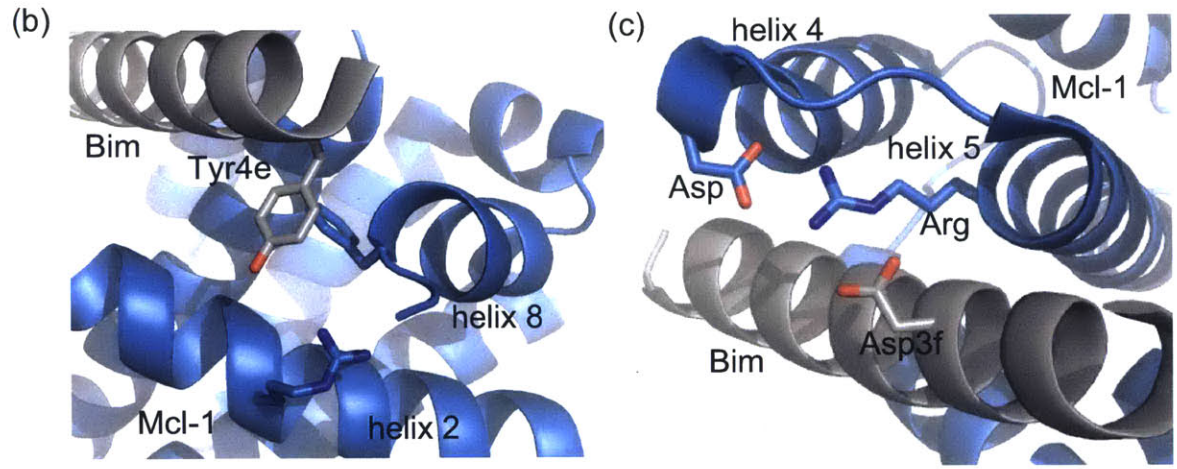


Figure 3.13. Specificity mechanisms that disfavor Mcl-1 binding. (a) Natural BH3-like sequences showing specificity for binding KSBcl-2 over Mcl-1 have diverse residues at position 4e (highlighted), including positively charged residues. (b) Tyrosine at position 4e in an Mcl-1:Bim BH3 complex (2PQK, blue:gray) (Fire et al., 2010). (c) Aspartate at position 3f in the same Mcl-1:Bim BH3 complex.

Among the 36 BH3-like peptides in Figure 3.4, 9 peptides showed ≥10-fold tighter binding to KSBcl-2 than to Mcl-1 (fold specificity and sequences are in Figure 3.13a). Three out of the nine KSBcl-2-specific peptides had arginine or lysine at position 4e. Boersma et al. showed that Mcl-1 binding to Bim BH3 is weakened by a Y4eK mutation, which may arise in part from repulsion from an arginine present on helix 2 and/or from the loss of a favorable aromatic interaction with a phenylalanine on helix 8 (both highlighted in green in Figure 3.1c, with the interactions in the Mcl-1:Bim BH3 structure shown in Figure 3.13b) (Boersma et al., 2008). KSBcl-2 and BHRF1 have negatively charged residues along the peptide interface with helix 2 (green in Figure 3.1c), and helix 8 in these proteins lacks the aromatic residue present in the human homologs. Finally, STATIUM predicts that substitution of tyrosine with lysine imparts a preference for KSBcl-2 over Mcl-1 binding, but is also moderately disruptive for KSBcl-2 binding. To test the role of lysine at position 4e, we made a Y4eK substitution in the background of the most specific KL and BL peptides (KL6-7 and BL6-22). KL6-7_Y4eK maintained tight binding to KSBcl-2 (K_D of 2.9 nM) but Mcl-1 binding was weakened by 350-fold; this peptide was 195-fold more specific for KSBcl-2 over Mcl-1 than KL6-7 (Table 3.8). The other Bcl-2 homologs displayed no binding to KL6-7_Y4eK (Table 3.6). Y4eK also improved the specificity of BL6-22 for BHRF1 by 3.6-fold, though it reduced BHRF1 binding affinity from a K_D of 2.2 nM to 30 nM. BL6-22_Y4eK had 70-fold specificity for BHRF1 over Mcl-1 and >100-fold specificity over the other human Bcl-2 homologs (no binding detected, Table 3.6). Thus, Y4eK is the single strongest contributor to specificity between KSBcl-2/BHRF1 and Mcl-1 that has been identified.

Table 3.8. Dissociation constants for Bcl-2 homologs binding to peptides designed for increased specificity against Mcl-1.

		K_D in nM with 95% confidence interval		
Peptide	Sequence	BHRF1	KSBcl-2	Mcl-1
KL6-7	RPQIWIHQGLQRLGDGLNAYYAR	>1000[a]	1.6 ± 0.45	22 ± 2.8
KL6-7_Y4eK	RPQIWIHQGLQRLGDGLNAYKAR	>1000[a]	2.9 ± 1.3	~7800[b]
KL6-7_D3fA	RPQIWIHQGLQRLGAGLNAYYAR	>1000[a]	180 ± 80	>5000[a]
KScomp	RPNIWLTQRLQRTGDGLNAYYAR	>1000[a]	13 ± 7.3	510 ± 190
BL6-22	RPIIWIGQEISRVGDRENAYYAR	2.2 ± 1.3	1100 ± 860	45 ± 6.1
BL6-22_Y4eK	RPIIWIGQEISRVGDRENAYKAR	30 ± 17	750 ± 300	~2200

[a] No binding was observed and the value given is the highest concentration of receptor used for that curve.
[b] Approximate values had large 95% confidence intervals due to the absence of an upper baseline.

Our library design strategy favored residues predicted to be non-disruptive for viral Bcl-2 binding. While appropriate for obtaining high affinity binders, this strategy may exclude mutations that provide specificity at the cost of affinity. We identified an example of such a mutation at peptide position 3f. Huang et al. showed that mutation of aspartate at 3f to alanine in a Bak BH3 peptide weakened Bcl-x_L binding more than KSBcl-2 binding (Huang et al., 2002). SPOT arrays also showed that mutation of aspartate at 3f in Bim to any other residue abolished binding to Mcl-1, Bfl-1, Bcl-2, and Bcl-x_L, but only weakened KSBcl-2 binding. To test the idea that substitutions at position 3f might provide specificity by decreasing Mcl-1 binding more than KSBcl-2 binding, we introduced D3fA into KL6-7. This mutation reduced binding to KSBcl-2 110-fold (K_D 180 nM), but weakened binding to Mcl-1 to the extent that it was undetectable up to 5000 nM (Table 3.8). Structural analysis suggests a plausible mechanism for this differential effect. Mcl-1 forms a salt bridge between an aspartate at the end of helix 4 and the BH1 motif arginine at the N-terminus of helix 5. This interaction appears to position the arginine optimally for interaction with the BH3 aspartate at 3f (Figure 3.13c). Although KSBcl-2 also has an

aspartate in the helix 4-helix 5 loop, it has an insertion of 3 residues preceding the aspartate, relative to Mcl-1 (Figure 3.1c). In homology models of KSBcl-2 complexes built on three different alignments with Mcl-1, the interactions observed in Mcl-1 were not able to form in an optimal configuration. Thus, the interaction between the 3f aspartate and the BH1 arginine may be weaker in KSBcl-2:BH3 complexes than in Mcl-1 complexes. We tested two other mutations with the potential to improve specificity for KSBcl-2 over Mcl-1 (G3eC and I2dR, see notes in Table 3.6), but these had equally detrimental effects on KSBcl-2 and Mcl-1 binding.

We also synthesized and tested a composite of the most KSBcl-2 specific library peptides, KL6-7 and KL8-7. KScomp has the N-terminal 5 substitutions present in the KL8-7 peptide, including the 2gR specificity mutation, and the 2 C-terminal substitutions of KL6-7 (3gG and 4aL). A 3dT substitution was also introduced, as the SPOT arrays suggested that threonine would provide more specificity against Mcl-1 than leucine. According to deep sequencing data, KScomp was preserved in our library from the naïve pool to pool KL5, but at a very low frequency (<0.0007%). We found that KScomp had reduced affinity for KSBcl-2 (13 nM K_D) when compared to library peptides that were tested from pools KL6 and KL8 (all had K_D values of ~ 1 nM), but KScomp had 39-fold specificity over Mcl-1, which is greater than the specificity of the other tested KL peptides. Given its low frequency, KScomp may have been lost from the library by chance. However, its 10-fold lower affinity than the peptides that survived the screen suggests that our sorting conditions were too stringent for KSBcl-2 affinity to enrich this peptide.

Discussion

As evolutionarily distant homologs of the mammalian anti-apoptotic Bcl-2 proteins, viral Bcl-2 proteins offer an interesting opportunity to study structure-function relationships in binding. In this paper, we characterized KSBcl-2, BHRF1 and M11 in terms of their sequence, structural, and binding-profile similarity to the human Bcl-2 homologs. The binding specificities of the different proteins provided insights into possible functional analogy, whereas analysis of sequence and structural similarity was useful for testing whether or not these measures are a good proxy for biochemical function similarity. We found that the low sequence identity of the human and viral proteins primarily emphasizes the high degree of divergence and is not particularly useful in identifying similarities between these groups. The SiteMAP results capture the clear similarities between Bcl-x_L, Bcl-2, and Bcl-w that are apparent from sequence analysis, but present a more nuanced view of the more dissimilar homologs Mcl-1, Bfl-1, KSBcl-2, and BHRF1. Though by sequence identity Bfl-1 is closest to Mcl-1, the SiteMAP results suggests that Bfl-1 has structural features closer to those of the x_L/2/w/M11 group. Conversely, BHRF1 has higher sequence identity to Bcl-x_L and Bcl-2, but appears structurally more similar to Mcl-1. KSBcl-2 is similar to Mcl-1 by structure, but to a lesser degree than the highly similar x_L/2/w/M11 group. The structural relatedness of both KSBcl-2 and BHRF1 to Mcl-1 was reflected in the library screening experiments, in which it was difficult to achieve large margins of binding specificity over Mcl-1.

By looking at parallels in binding similarity and structure we can make hypotheses about functional similarity between the viral and human Bcl-2 homologs. For example, BHRF1 shows structural similarity to Mcl-1, and appears to have BH3 binding preferences representative of a narrower subset of the binding preferences of Mcl-1. The restrictive binding of BHRF1 to only

Bim, Bid, and Puma and the pro-apoptotic effectors Bak and Bax is reminiscent of human Bcl-b and viral Bcl-2 homolog F1L from vaccinia virus. Bcl-b only binds Bim, Bik and Bax, and F1L is reported only to bind Bim, Bak, and Bax, out of a set of pro-apoptotic BH3-only proteins and Bax and Bak (Kvansakul et al., 2008; Rautureau et al., 2012). BHRF1 binds Bak in cells, and its interaction with Bim appears to be especially critical to its anti-apoptotic activity in lymphocytes (Desbien et al., 2009) .The restrictive binding profiles of BHRF1 and F1L may be a strategy employed by viruses either to focus on only the most critical activators and effectors of apoptosis, or to prevent off-target effects by avoiding the "moonlighting" functions of human Bcl-2 homologs in autophagy, mitochondrial homeostasis, or other pathways (Hardwick et al., 2012).

M11 and KSBcl-2 have binding preferences similar in many respects to those of Mcl-1 (Fig. 3). However, structural similarity suggests that M11 also has features similar to the Bcl-x_L/2/w class, whereas KSBcl-2 is closest to Mcl-1 by both structure and binding. As previously noted, both KSBcl-2 and M11 bind Beclin-1 and prevent autophagy as an important part of their cellular function (Ku et al., 2008; Pattingre et al., 2005). Bcl-2 and Bcl-x_L also bind Beclin-1 to prevent autophagy; Mcl-1 binds Beclin-1 more weakly (Erlich et al., 2007). However, KSBcl-2 and M11 do not bind tightly to the classical Bcl-2/x_L/w binder, Bad, but do bind the Mcl-1 binder, Noxa (Flanagan and Letai, 2008; Huang et al., 2002; Sinha et al., 2008). It will be interesting to investigate functionally why KSBcl-2 and M11 have binding characteristics of both the Bcl-2/x_L/w and Mcl-1 classes. Much of the viral Bcl-2 literature only focuses on comparisons between viral Bcl-2 and human Bcl-2 or Bcl-x_L function. Our results suggest that an investigation of the relationship between viral Bcl-2 and Mcl-1 function would be informative in discerning the relationship between the roles played by these viral and human Bcl-2 homologs.

Several lessons can be learned from our attempts to design KSBcl-2 and BHRF1 specific BH3 peptides. First, trade-offs between affinity and specificity must be considered when dealing with homologs with very similar binding preferences, such as the viral Bcl-2s and Mcl-1. The importance of such trade-offs has been previously discussed in the context of protein-protein interaction design (Chen and Keating, 2012; Grigoryan et al., 2009). Our library screen imposed criteria on both affinity and specificity, but with an emphasis on tight binding. As a consequence, we did not identify viral Bcl-2 selective peptides that had weakened binding to KSBcl-2 and BHRF1, such as KScomp (Table 3.8). Only very tight binders (K_d ~1 nM) survived to round 8. Screening conditions could be adjusted to maintain destabilizing but specific mutations by relaxing target affinity requirements in the design stage and reducing the stringency of screening. Peptides identified in this way could then be further optimized for tight binding. Another consideration in achieving specific binders is library diversity. Discriminating homologs with very similar binding preferences such as KSBcl-2 and Mcl-1 may require broader exploration of sequence space than can be achieved from screening just one library of ~10^7 molecules, or may require more precise selection of the sequences to be screened. For example, we did not vary position 4e in our library, the site of our most specific mutation, because we were limited by library size. Yet modeling using STATIUM, as well as hints from the literature, indicated this as a promising substitution. A recent paper by Dutta et al. also found that the Y4eK mutation imparted specificity for Bcl-x_L over Bcl-2, suggesting that positions outside the core BH3 motif can strongly influence specificity (Dutta et al., 2014). Searching sequence space more extensively, with larger libraries or multiple rounds of library design and sorting, could accelerate discovery.

In this work we ultimately identified selective peptide inhibitors of viral Bcl-2 family proteins KSBcl-2, BHRF1 and M11 by various means, including library screening, rational mutagenesis and profiling of BH3-like sequences previously identified in the human proteome. Although we did not sort a library for M11 specificity, one of the 36 BH3-like peptides identified from the proteome, AGBL2 showed moderate binding to M11 (210 nM), and at least 20-fold weaker binding to all of the other Bcl-2 homologs (Table 3.1). For use as a research reagent to specifically inhibit M11, this offers a 30-fold improvement in affinity over the existing M11-specific Beclin-1 mutant (Su et al., 2014). Several of our BL peptides also have higher affinity for M11 (~1 nM) than is offered by natural BH3 sequences, although these peptides show only moderate specificity over Mcl-1. The Y4eK mutants of the most specific KL and BL peptides (KL6-7_Y4eK and BL6-22_Y4eK) achieved large margins of specificity over all of the human Bcl-2 homologs. KL6-7_Y4eK is an especially promising peptide for future studies of KSBcl-2, as it binds very tightly and selectively to this protein.

As recently demonstrated by the use of a designed protein inhibitor of BHRF1 to suppress tumor growth in a mouse xenograft model of Epstein-Barr-positive human lymphoma, anti-viral Bcl-2 agents have potential for use in disease therapy (Procko et al., 2014). Large proteins face obstacles in terms of intracellular delivery, and in this respect the much smaller designed peptides presented here may have advantages. More immediately, designed potent and selective inhibitors of viral Bcl-2 proteins can provide useful reagents for studying the function of viral Bcl-2 homologs in herpesvirus-associated pathologies, and for testing the extent to which these proteins contribute to cancer onset or progression.

Materials and Methods

Sequence identity, homology modeling and SiteMAP

Multiple-sequence alignments of Bcl-2 family proteins were manually constructed based on preliminary Clustal alignments and structural analysis. For calculation of sequence identity in the BH3 binding groove, Pymol was used to select residues of the Bcl-2 homologs within 7 Å of the BH3 peptide. Structures used for this were: 2PQK(Mcl-1:Bim) (Fire et al., 2010), 2WH6 (BHRF1:Bim) (Kvansakul et al., 2010), 2VM6 (Bfl-1:Bim) (Herman et al., 2008), 3FDL (Bcl-x_L:Bim) (Lee et al., 2009a), 2XA0 (Bcl-2:Bax) (Ku et al., 2011), 3BL2 (M11:Beclin-1) (Ku et al., 2008), and the homology models of KSBcl-2 and Bcl-w bound to Bim. An inclusive set of groove residues was defined to include any residue found within 7 Å of any peptide atom in any structure. See Figure 3.1c for alignments.

Homology models of KSBcl-2 and Bcl-w were built using MODELLER (version 9.1) (Eswar et al., 2007; Sali and Blundell, 1993). The KSBcl-2:Bim complex was modeled on 2PQK (Fire et al., 2010), 2WH6 (Kvansakul et al., 2010), 1O0L (Hinds et al., 2003), 2VM6 (Herman et al., 2008), 3FDL (Lee et al., 2009a), and 1G5M (Petros et al., 2001a). The Bcl-w:Bim complex was modeled on 3FDL, 1G5M, 2PQK, and 2VM6. The KSBcl-2:Bak complex was modeled on 3PK1 (Czabotar et al., 2011) and 2XPX (Kvansakul et al., 2010). The Bcl-w:Bak complex was modeled on 1BXL (Sattler et al., 1997) and 2XA0 (Ku et al., 2011).

For SiteMAP (SiteMap, version 3.0, Schrödinger, LLC, New York, NY, 2014) structure analysis, the following PDB IDs were aligned to 2PQK (Mcl-1:Bim): 2VM6 (Bfl-1:Bim) (Herman et al., 2008), 3I1H (Bfl-1:Bak) and 3MQP (Bfl-1:Noxa) (NESG, unpublished), 2VOG (murine Bfl-1: Bmf) (Smits et al., 2008), 2VOF (murine Bfl-1:Puma) (Smits et al., 2008), 2VOH (murine Bfl-1:Bak) (Smits et al., 2008), 2VOI (murine Bfl-1:Bid) (Smits et al., 2008), 2WH6

(BHRF1:Bim) (Kvansakul et al., 2010), 2XPX (BHRF1:Bak) (Kvansakul et al., 2010), 3FDL (Bcl-x_L:Bim) (Lee et al., 2009a), 1BXL (Bcl-x_L:Bak) (Sattler et al., 1997), 1G5J (Bcl-x_L:Bad) (Petros et al., 2001b), 3IO8 (Bcl-x_L:Bim L12F) (Lee et al., 2009a), 2PON (Bcl-x_L:Beclin-1) (Feng et al., 2007), 3PK1 (Mcl-1:Bax) (Czabotar et al., 2011), 2KBW (Mcl-1:Bid) (Liu et al., 2010), 3KJ0 (Mcl-1:Bim I2dY) (Fire et al., 2010), 3KJ1 (Mcl-1:BimI2dA) (Fire et al., 2010), 3KJ2 (Mcl-1:Bim F4aE) (Fire et al., 2010), 3BL2 (M11:Beclin-1) (Ku et al., 2008), 2XA0 (Bcl-2:Bax) (Ku et al., 2011), and homology models of Bcl-w and KSBcl-2 bound to Bak and Bim BH3s. The following structures were relaxed using minimization in Maestro in the absence of the peptide (minimization uses the OPLS_2005 force field and proceeds until no movements are >0.05 Å) and were aligned to 2PQK using just the equivalent receptor positions: 2PQK, 2WH6, 2VM6, 3FDL, 3BL2, 2XA0, and the KSBcl-2 and Bcl-w:Bim models. Cealign in PyMOL (version 1.3, Schrödinger, LLC) was used to do the alignments based on the Cα atoms of helix 2, the helix4/5 region, and peptide regions of the structures. The following residues were used for the alignments: 208-220, 243-280, and 6-17 for 2PQK; 32-44, 68-105, and 148-159 for 2VM6; 45-57, 80-117, and 58-69 for 2WH6; 32-103, 127-163, and 59-70 for 2XA0; 36-48, 68-104, and 110-121 for 3BL2; 85-97, 120-156, and 90-101 for 3FDL; 31-43, 63-103, and 153-164 for KSBcl-2; 41-53, 76-112, and 164-175 for Bcl-w. The numbering for KSBcl-2 and Bcl-w refers to the residue number in the natural gene, with the third set of numbers referring to BH3 positions 2d-4a. Equivalent residue numbers to those listed above were used to align all structures of each homolog. Maestro (version 9.7) Prepwizard was then run on the aligned coordinates (complex structures) with the following settings: find disulfides, fill sidechains, propKa_pH 7.0, OPLS force field version 2005, RMSD minimization cutoff of 0.3, cap termini. SiteMAP (version 3.0, analysis, Schrödinger, LLC) was then run on only the receptor

coordinates, restricting map finding to a zero Å box around the 2PQK Bim peptide, with the enclosure setting at 0.4 for finding shallow grooves, and with verbose file output (Halgren, 2009; 2007). SiteMAP identified the BH3 binding groove as the largest site for all structures. Scripts were written in Python (version 2.6.1) to process the ".smpot" SiteMAP potential output files (lists of coordinates and potentials). All points that met the SiteMAP default potential values for hydrophilic (-8 kcal/mol) and hydrophobic (-0.5 kcal/mol) maps were extracted. The environment around each peptide position was analyzed by selecting all map points (donor, acceptor, and hydrophobic maps) that were within 6 Å of the Cβ of each position in the Bim peptide in structure 2PQK. The intersection between receptors for each type of map at each position was computed by counting all points within 1 Å of a point in another receptor's map. A similarity score was computed by summing the intersection counts from the acceptor, donor, and hydrophilic maps at all positions.

Clustering

Clustering of proteins by sequence identity, binding profile, or SiteMAP similarity was performed in Matlab (version R2012b) with the clustergram function in the Bioinformatics toolbox using hierarchical clustering based on correlation. The exception was clustering of the peptides (rows in Figure 3.4 and 3.10), which were clustered by Euclidean distance to give better ordering according to affinity.

Expression and purification of Bcl-2 proteins

Biotin-acceptor peptide (BAP-tagged) and His_6-tagged variants of all eight Bcl-2 homologs were made in vector pDW363 (used for analysis of direct binding on the bacterial cell surface, and, in the case of M11, for the fluorescence anisotropy experiments) (Tsao et al., 1996). Sequence encoding MBP in the parent vector was replaced by the Bcl-2 homolog, resulting in a

product with the BAP-tag and His_6-tag at the N-terminus of the Bcl-2. C-myc-tagged versions of all Bcl-2 homologs were in the pSVM vector. Constructs are given in Table 3.9.

pDW363 constructs were expressed in BL21 (DE3) pLysS Rosetta cells. 50 mL overnights were grown at 37 °C with shaking, and 10 mL of the overnight culture was used to inoculate 1 L of LB including 100 μg/mL ampicillin, 25 μg/mL chloramphenicol, and 12-15 mg of D-(+)-biotin. Cells were grown at 37 °C with shaking to an optical density at 600 nm (O.D. 600) of 0.6 and then induced with 1 mM IPTG for ~5 hours before harvesting. pSVM constructs were grown similarly, with 50 μg/mL kanamycin, 25 μg/mL chloramphenicol, and no biotin.

Purification of pSVM constructs was performed as follows. 1 L cell pellets were resuspended in 25 mL of 5 mM imidazole, 500 mM NaCl, 1 mM DTT, 20 mM Tris pH 8.0, and 0.2 mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor. Cells were sonicated ten times for 30 seconds followed by 30 seconds of rest. The supernatant from the centrifuged lysis product was filtered through 0.2 μm filters before application to 3 mL of Ni-nitrilotriacetic acid agarose resin equilibrated in 20 mM Tris pH 8.0, 500 mM NaCl. After the supernatant was applied, the resin was washed 3 times with 8 mL of 20 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0. The His_6-MBP-c-myc-Bcl-2 construct was eluted with 8 mL 20 mM Tris pH 8.0, 500 mM NaCl, 300 mM imidazole. At this point, the protein in the eluate was quantitated by absorbance at 280 nm and diluted to 1 mg/mL with 50 mM NaCl, 50 mM Tris pH 8.0, 1 mM DTT, 0.5 mM EDTA, before the addition of TEV protease at a ratio of 50 mg Bcl-2 protein:1 mg TEV. The Bcl-2 protein and TEV were dialyzed overnight at 4 °C against 1 L of 50 mM NaCl, 50 mM Tris pH 8.0. The dialyzed protein was purified by nickel affinity chromatography as before, with the exception that the wash buffer did not contain imidazole, to minimize dissociation of His_6-MBP and His_6-TEV from the resin. The flow-through (and if not overly

contaminated, the wash) was applied to a S75 26/60 size exclusion column equilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 1% glycerol, 1 mM DTT. Purity was verified by SDS-PAGE, and proteins were frozen at -80 °C in the final buffer. pDW363 constructs were purified similarly, but with no cleavage step, and with just one nickel affinity step followed by the size exclusion column.

Table 3.9. Sequences of Bcl-2 constructs and primers

Construct	Sequence
pDW363 tags[a]	MAGGLNDIFEAQKIEWHEDTGGSSHHHHHH
pSVM tags (before TEV cleavage)	MGSSHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIK VTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDK AFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIP ALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDV GVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGP WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKE FLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQK GEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNN NNNNNNLGIEENLYFQGSEQKLISEEDLGGSGGTS
BHRF1	MAYSTREILLALCIRDSRVHGNGTLHPVLELAARETPLRLSPEDTVVLRYH VLLEEIIERNSETFTETWNRFITHTEHVDLDFNSVFLEIFHRGDPSLGRALA WMAWCMHACRTLCCNQSTPYYVVDLSVRGMLEASEGLDGWIHQQGGW STLIEDNIPGSRRF
KSBcl-2	MDEDVLPGEVLAIEGIFMACGLNEPEYLYHPLLSPIKLYITGLMRDKESLFE AMLANVRFHSTTGINQLGLSMLQVSGDGNMNWGRALAILTFGSFVAQKL SNEPHLRDFALAVLPVYAYEAIGPQWFRARGGWRGLKAYCGRIVTD
M11	MSHKKSGTYWATLITAFLKTVSKVEELDCVDSAVLVDVSKIITLTQEFRRH YDSVYRADYGPALKNWKRDLSKLFTSLFVDVINSGRIVGFFDVGRYVCEE VLCPGSWTEDHELLNDCMTHFFIENNLMNHFPLED
Mcl-1	DELYRQSLEIISRYLREQATGAKDTKPMGRSGATSRKALETLRRVGDGVQ RNHETAFQGMLRKLDIKNEDDVKSLSRVMIHVFSDGVTNWGRIVTLISFG AFVAKHLKTINQESCIEPLAESITDVLVRTKRDWLVKQRGWDGFVEFFHV EDLEGG
Bcl-xL	MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSA INGNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQALREAGDEFELR YRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALC VESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNA AAESRKGQER

Bcl-2	MAHAGRTGYDNREIVMKYIHYKLSQRGYEWDAGDVGAAPPGAAPAPGIF SSQPGHTPHPAASRDPVARTSPLQTPAAPGAAAGPALSPVPPVVHLTLRQA GDDFSRRYRRDFAEMSSQLHLTPFTARGRFATVVEELFRDGVNWGRIVAF FEFGGVMCVESVNREMSPLVDNIALWMTEYLNRHLHTWIQDNGGWDAF VELYGPSMRPLFDFSWLSL
Bcl-w	MATPASAPDTRALVADFVGYKLRQKGYVCGAGPGEGPAADPLHQAMRAA GDEFETRFRRTFSDLAAQLHVTPGSAQQRFTQVSDELFQGGPNWGRLVA FFVFGAALCAESVNKEMEPLVGQVQEWMVAYLETRLADWIHSSGGWAE FTALYGDGALEEARRLRE
Bfl-1	MTDCEFGYIYRLAQDYLQCVLQIPQPGSGPSKTSRVLQNVAFSVQKEVEK NLKSCLDNVNVVSVDTARTLFNQVMEKEFEDGIINWGRIVTIFAFEGILIKK LLRQQIAPDVDTYKEISYFVAEFIMNNTGEWIRQNGGWENGFVKKFEPK
pBAD33-eCPX-FLAG-Bim BH3	MKKIACLSALAAVLAFTAGTSVADYKDDDDKAGGSGGSGGQSGRPEIWI AQELRRIGDEFNAYYARRVGGQSGQSGDYNKNQYYGITAGPAYRINDWA SIYGVVGVGYGKFQTTEYPTYKHDTSDYGFSYGAGLQFNPMENVALDFS YEQSRIRSVDVGTWILSVGYRFGSKSRRATSTVTGGYAQSDAQGQMNKM GGFNLKYRYEEDNSPLGVIGSFTYTEKSRTAS
Library assembly:	
ecpx_rev_ library	CGCCAAAACAGCCAAGCTTGG
BimC_fwd_ library	AACGCGTACTACGCCCGTCG
sfiI_fwd	CGATCGGCCAGTCTGGCCGTC
KSBcl2_ library	GATCGGCCAGTCTGGCCGTCCGVAMATCTGGWWSNMCCAGSDGSTGC RKCGTNHCGGTGACRDABKT AACGCGTACTACGCCCGTCG
BHRF1_ library	GATCGGCCAGTCTGGCCGTCCGRNAATCTGGHDCRVCCAGVAMMTTV VCCGTRYCGGTGACRVGKWK AACGCGTACTACGCCCGTCG
Illumina sequencing preparation:	
MmeI_fwd	GGGACCACCACCTCCGACCGGCGGCCAGTCTGGCCGTC
3prime_rev	CAAGCAGAAGACGGCATACGAGATCTGCGTCCAGGACCAGACTGCCC AGACTGCCCTCC

finalPCR_fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
finalPCR_rev	CAAGCAGAAGACGGCATACGAGATCTGCGTCCAGGACCAGACTGC
rev_seq	CTGCGTCCAGGACCAGACTGCCCAGACTGCCCTCC
example adapter and barcode ('B')	ACACTCTTTCCCTACACGACGCTCTTCCGATCTBBBBBTC

[a] pDW363 and pSVM tags were on N-terminus of all Bcl-2 constructs. Bcl-2 constructs were the same in both vectors.

Binding affinity measurements by fluorescence anisotropy

Library peptides, the Bim BH3 peptide, and Bim BH3 peptide mutants were 23 residues long with N-terminal 5/6-fluorescein amidite and C-terminal amidation and were synthesized by the MIT Biopolymers Laboratory. The crude synthesis product was verified by mass spectrometry to contain predominantly the desired species and then purified by HPLC on a C18 column with a linear gradient of acetonitrile in water. Proteome-derived BH3 peptides (26-mers, N-fluoresceinated, C-amidated) in Figure 3.4 were the same as those used by DeBartolo et al. (DeBartolo et al., 2014). Direct fluorescence anisotropy experiments were performed as in Foight et al., with a titration of twelve receptor concentrations, with a maximum concentration of 1 μM receptor for the library peptide curves and 3 μM for the proteome peptide curves (Foight et al., 2014). C-myc-tagged receptors were used for all Bcl-2 homologs, with the exception of M11, for which a BAP and His_6-tagged construct was used. All K_D values are from averaged data from three replicates done over three days. Data were fit as described for direct fluorescence anisotropy experiments in Foight et al., with the upper baseline and K_D fit, while the lower

baseline and concentration of the fluoresceinated peptide were fixed as described (Foight et al., 2014).

SPOT arrays

SPOT arrays were synthesized on activated nitrocellulose support using Fmoc protection/deprotection chemistry by an Intavis AutoSpot robot in the MIT Koch Biopolymers Laboratory. The peptides were synthesized with PEG3 (three ethylene glycol units) linkers at the carboxy terminus. All peptides were 26 residues long (wild-type Bim sequence MRPEIWIAQELLRIGEDEFNAYYARV). Native BH3 peptides in column "X" in Figure 3.5, top to bottom: Bad, Bid, Bmf, Hrk, Noxa, Bik, Bak, Bax, Bcl-x_L, Bcl-w, Bcl-2, Mcl-1, Bfl-1, BHRF1, KSBcl-2, with the BH3 residues equivalent to those described for Bim BH3. Arrays were processed by first hydrating the membranes in 100% methanol, followed by water. Arrays were blocked in blocking buffer (50 mM Tris pH 8.0, 100 mM NaCl, 0.01% Triton X-100, 1% BSA) for 20 min. 100 nM c-myc-BHRF1 or KSBcl-2 in blocking buffer was applied to the arrays and incubated 1 hr at ~ 23 °C with rocking. Arrays were washed 3 times with 10 mL blocking buffer before addition of 5 mL blocking buffer + anti-c-myc-Cy3 (Sigma Aldrich) at 1:100 dilution. Arrays were incubated in the dark for 30 minutes at ~23 °C with rocking. Finally, the array was washed 3 times with 10 mL blocking buffer, imaged on a Typhoon 9400, and analyzed with ImageQuant.

Models used for library design

Illumina sequencing data of a yeast display library sorted for selective binding to KSBcl-2 but not Mcl-1 and Bcl-x_L was used to inform selection of substitutions that are tolerated by KSBcl-2 binding (Foight et al., 2014). Enrichment of residues was computed as the percentage of unique sequences with a residue after five rounds of screening minus the percentage of unique

sequences with that residue in the naïve library. Sequences surviving the selection were likely to be moderate to high-affinity KSBcl-2 binders, but not all were selective for KSBcl-2. Peptides tested individually in solution bound KSBcl-2 with dissociation constants of ~1 nM but also bound to Mcl-1 with similar affinities (Foight et al., 2014).

Position-specific scoring matrices based on SPOT array intensities for KSBcl-2 and BHRF1 ($PSSM_{SPOT}$ models) were built by computing $\log_{10}$ of each mutant peptide fluorescence intensity divided by the average wild-type intensity, as described previously for the human Bcl-2 proteins (DeBartolo et al., 2012; Dutta et al., 2010b; London et al., 2012).

STATIUM is a statistical potential built to evaluate the fit of a sequence on a particular structural template. STATIUM models were built using the STATIUM_sidechain method reported by DeBartolo et al., with the exception of Bcl-w, for which the original STATIUM method was used (DeBartolo et al., 2012; 2014). The template structures used were as follows: a KSBcl-2:Bim BH3 homology model described above, 2PQK (Mcl-1:Bim BH3) (Fire et al., 2010), 3IO8 (Bcl-x_L:Bim BH3_L12F) (Lee et al., 2009a), 2VM6 (Bfl-1:Bim BH3) (Herman et al., 2008), 2WH6 (BHRF1:Bim BH3) (Kvansakul et al., 2010), and 2XA0 (Bcl-2:Bax BH3) (Ku et al., 2011). The Bcl-w model was built using template 1ZY3 (Denisov et al., 2006), an NMR-based docking model with Bid BH3.

Library design

KSBcl-2 library To guide library design, all possible point mutants of Bim at positions 2a-4e were scored with the three models described above, with some positions not having data from all models. Substitutions were rated as **non-disruptive** for KSBcl-2 binding if they met at least one of the following criteria:

i) Frequency increase of 1% in the library screening experiment targeting KSBcl-2, or present in at least 10% of unique sequences after 5 rounds of screening

ii) $PSSM_{SPOT\text{-}KSBcl\text{-}2}$ score greater than the median score for all mutations across all positions

iii) $\Delta STATIUM_{KSBcl\text{-}2}$ (raw Bim score – raw mutant score) greater than the median for all mutations across all positions

Substitutions were counted as **specific** for KSBcl-2 over an alternative receptor if $\Delta PSSM = PSSM_{KSBcl\text{-}2} - PSSM_{alternative\ receptor}$ or $\Delta\Delta STATIUM = \Delta STATIUM_{KSBcl\text{-}2} - \Delta STATIUM_{alternative\ receptor}$ was greater than 0.2. As for $\Delta STATIUM_{KSBcl\text{-}2}$, $\Delta STATIUM_{alternative\ receptor}$ was defined as the score for Bim on that receptor minus the mutant score on that receptor. The **specificity score** for each substitution was defined as the number of scores that met the aforementioned cutoff (e.g., a specificity score of 10 would be given if all ΔPSSM and all ΔΔSTATIUM scores for all 5 human Bcl-2 proteins were greater that 0.2). The 0.2 cutoff was chosen because it was close to the average over all receptor comparisons of the median ΔPSSM and ΔΔSTATIUM values for all mutations, and seemed to represent a reasonable minimal margin of specificity when compared to model scores for known specificity mutations.

Libraries were constructed using degenerate codons chosen by a computational optimization protocol (Chen et al., 2013). To guide the selection of a set of degenerate codons to consider at each position, residue substitutions were divided into two categories, "preferred" or "required". Substitutions were included in the **preferred** category if they were non-disruptive by at least one out of the three categories previously mentioned and had a specificity score of at least 6 (or at least 3 for positions with information available from only $PSSM_{SPOT}$ or STATIUM). Additionally, some substitutions were included that did not meet this criteria but had large

ΔPSSM or ΔΔSTATIUM scores for Mcl-1. **Required** residues included the wild-type residues and a subset of the preferred residues that were hand-selected for a high degree of specificity, with emphasis placed on specificity against Mcl-1. Degenerate codons selected for consideration encoded all of the required residues, and the choices were further narrowed by the elimination of any codon that included more trinucleotides than another but encoded fewer preferred residues.

Optimization of degenerate codon combinations was performed by integer linear programming, as previously described, with a limit on the DNA size of the library set at 1×10^7 (Chen et al., 2013). At positions where more than one degenerate codon of equal DNA size encoded the same number of "preferred" residues, one codon was selected to maximize specificity against Mcl-1. Codon options at certain positions were narrowed to encourage the ILP code to compose a library in which the majority of positions had a modest number of variants (more than 3). This strategy was employed with the aim of decreasing the likelihood that the library sequences would be overly reliant on one or two positions to achieve specificity. Several possible library designs arose from different choices of positions to mutate and codon choices to include. The final library design was chosen because it contained a large number of protein sequences (5.23×10^6), and it had the highest number of sequences that scored in a range predicted to have high affinity for KSBcl-2 and weaker binding than Bim to Mcl-1 on the $PSSM_{SPOT}$ and STATIUM models.

BHRF1 library Design of the BHRF1-targeted library used the same $PSSM_{SPOT}$ and STATIUM models as described above for the cellular receptors. BHRF1 models used to categorize mutations included a $PSSM_{SPOT}$ model and STATIUM models built on two templates, 2WH6 (BHRF1:Bim BH3) and 2XPX (BHRF1:Bak BH3) (Kvansakul et al., 2010). Preferred residues were defined as all residues that were non-disruptive by at least one out of three models

(i.e., a $PSSM_{BHRF1}$ or $\Delta STATIUM_{BHRF1}$ score above the median across all mutations at all positions). Required residues were selected to be non-disruptive and highly specific for BHRF1, with preference given to residues that disfavor binding to Mcl-1. A few residues were also included as required because $PSSM_{SPOT}$ indicated high affinity binding to BHRF1. Codon selection was performed as for the KSBcl-2 library. In selecting the final library, a similar emphasis was placed on including 4-8 mutations at as many positions as possible; manual selection of codons and adjustment of required residues achieved this goal. The final BHRF1 library had a large number of protein sequences (6.72×10^6), as well as large numbers of sequences that scored as high for BHRF1 affinity but low for Mcl-1 affinity according to the $PSSM_{SPOT}$ and STATIUM models. The final library designs are presented in Tables 3.2 and 3.3.

Library construction

The pBAD33-eCPX-FLAG-Bim BH3 construct used for bacterial surface display of the Bim variant libraries and the primers used for library construction are shown in Table 3.9. The protocol for library assembly was similar to that presented in Getz et al. (2012). The vector into which the library was cloned was pBAD333-eCPX with a FLAG tag included for use as an expression control and linkers optimized for Bim BH3 expression and binding. A vector containing the non-binding mutant BimL3aD was used as a cloning template to prevent background from undigested vector from influencing the screening results. 200 mL of this vector expressed in DH5α *E. coli* was minipreped over 8 Qiagen miniprep columns and digested with SfiI according to the NEB protocol for 8 hours. Digested vector was PCR purified over an appropriate number of Qiagen PCR purification columns and eluted with water. The digested vector was then dephosphorylated with Antarctic phosphatase (1 μL phosphatase: 1 μg DNA) at 37 °C for 2 hours, followed by 10 min at 65 °C for deactivation of the enzyme. This prevented

re-ligation of the original insert back into the vector. This protocol yielded 8-10 μg of DNA, which was the amount of vector needed per library.

To prepare the library inserts, a DNA fragment corresponding to the eCPX gene was generated by performing PCR with the primers eCPX_rev_library and BimC_fwd_library (Table 3.9) on the BimL3aD-eCPX vector. Platinum Taq HiFi polymerase was used according to manufacturers instructions. The product was run on an agarose gel using GelGreen DNA dye and blue light for visualization to minimize DNA damage. The appropriate band was extracted with a Zymo DNA gel recovery kit. Overlap PCR was then performed on this eCPX fragment to add the varied Bim region. For each library, the appropriate library primer (1.5 μL at 10 μM) was combined with 1 μL of the eCPX product in standard Platinum Taq HiFi conditions. Seven PCR cycles were run with an annealing temperature of 54 °C, and then 1 μL of a 12.5 μM mix of the eCPX_rev_library and SfiI_fwd primers were added and a further 28 PCR cycles were completed (Table 3.9). 15, 50 μL reactions were done per library to obtain sufficient insert DNA. The PCR products were purified over Zymo Clean & Concentrate columns and digested with SfiI for 4 hours. Digestion reactions were then purified again. This protocol yielded ~7 μg of insert DNA for each library.

Digested vector and library insert were ligated at a 5:1 molar ratio of insert:vector, using ~8 μg vector. An 800 μL reaction volume with standard T4 DNA ligase conditions was used, and ligations were performed at 14 °C overnight. A 20 μL control ligation of digested vector alone was also performed to allow estimation of vector background. Following ligation, the ligase was deactivated at 70 °C for 10 minutes, and the ligation was concentrated over four Zymo Clean & Concentrate columns, eluting each column in 12.5 μL DNA elution buffer (from the kit). Each ligation aliquot was then desalted for 20 minutes on a Millipore 0.025 μm filter suspended on an

Eppendorf tube of sterile MilliQ water. The ligations were then pooled on ice and split between 4, 250 µL aliquots of thawed MC1061 competent *E. coli* cells. MC1061 competent cells were prepared as in Getz et al. (2012). Each aliquot was electroporated in a cold 2 mm BioRad cuvette at 2.5 kV, 50 µF, 100 Ω on a BioRad Gene Pulser electroporator, and then immediately rinsed three times with 1 mL warm SOC and combined with 7 mL warm SOC and then allowed to recover at 37 °C for 1 hr on a rotator wheel. The library was then added to 500 mL LB + 25 µg/mL chloramphenicol + 0.2% w/v sterile-filtered glucose in a 2 L flask and grown at 37 °C until an O.D.600 of ~1.5 was reached (~7 hours). 400 mL of the library culture was pelleted at 3000 relative centrifugal force (rcf) for 15 minutes and re-suspended in 10 mL SOB + 15% v/v sterile glycerol and aliquoted and frozen at -80 °C. The number of transformants was 3.9×10^9 for the BHRF1 library and 1.8×10^9 for the KSBcl-2 library, with vector background estimated at < 0.001%.

Library Sorting

The general protocol for preparation of library samples for sorting was as follows. A quantity of glycerol stock sufficient to oversample the estimated library diversity by at least 10-fold was used to inoculate 5 mL LB plus 0.2% glucose and 25 µg/mL chloramphenicol and grown overnight at 37 °C on a rotator wheel. ~100 µL of overnight culture (or enough to maintain library diversity) was pelleted and used to inoculate 5 mL LB plus 25 µg/mL chloramphenicol. The culture was then grown to an O.D. 600 of 0.5-0.6 and induced with 0.04% w/v arabinose for 1 hour. Sufficient cells to oversample the library diversity were pelleted at 3000 rcf for 5 minutes and mixed with 100 µL PBS + 0.1% BSA (PBSA) and 100 µL of an appropriate 2x biotinylated viral Bcl-2 stock. Cells were incubated at room temperature with shaking for 1 hour. Cells were pelleted, washed with 200 µL PBSA, pelleted again, and then 210

μL of a mix of streptavidin-PE and anti-FLAG-APC (both at a 1:100 dilution in PBSA) was added. For three-color competition sorts, a two-step antibody labeling process was used. The first incubation was with streptavidin-PE (Molecular Probes), anti-FLAG-APC (Perkin Elmer), and rabbit-anti-c-myc (Sigma), and the second incubation (after a wash step) was with anti-rabbit-FITC (Sigma). Gates for the three-color competition sorts were set to exclude FITC-positive cells and include PE-positive cells. Cells were incubated with the labeling reagents for 15 minutes on ice in the dark. The pelleting and washing steps were repeated, and cells were resuspended in 1.5 mL PBSA for sorting on a BD FACSAria. Cells were collected in PBSA, which was diluted in SOC after sorting and cells were allowed to recover at 37 °C overnight, at which point some cells were harvested for glycerol stocks and some were used to inoculate cultures for the next day's sort.

The sorting scheme for the KSBcl-2 library was as follows (summarized in Figure 3.7): sort for binding to 1 μM biotinylated-KSBcl-2 (positive sort; 2.7% of cells collected, cells were frozen as a glycerol stock and used to inoculate overnight cultures at a later date), sort for cells that do not bind to 500 nM biotinylated-Mcl-1 (negative sort; 7.1% of cells collected, the next four sorts after this point were done on consecutive days), positive sort at 100 nM biotinylated-KSBcl-2 (2.1% collected), negative sort at 300 nM Bfl-1 and Bcl-w (21.5% collected), sort for binding to 100 nM biotinylated-KSBcl-2 and in the presence of 500 nM c-myc-Mcl-1 (competition sort; 3.4% collected), competition sort at 100 nM biotinylated-KSBcl-2 and 500 nM c-myc-Mcl-1 (9.9% collected, glycerol stocks were made at this point and used to inoculate overnight cultures), competition sort at 100 nM biotinylated-KSBcl-2 and 500 nM c-myc-Bcl-w and Bfl-1 (23.7% collected), and finally, competition sort at 20 nM biotinylated-KSBcl-2 and 1 μM c-myc-Mcl-1 (0.6% collected). Competition sorts were all three-color. As often as possible,

sorts were done on consecutive days to minimize growth time for the libraries, thereby minimizing potential for the occurrence of secondary mutations that imparted growth advantages.

The sorting scheme for the BHRF1 library was as follows: positive sort at 1 μM biotinylated-BHRF1 (0.4% of cells collected, cells were frozen as a glycerol stock and used to inoculate overnight cultures at a later date), negative sort at 500 nM biotinylated-Mcl-1 (21.1% of cells collected), positive sort at 100 nM biotinylated-BHRF1 (3.3% collected), negative sort at 500 nM biotinylated-Mcl-1 (39.5% collected), competition sort at 100 nM biotinylated-BHRF1 and 500 nM c-myc-Mcl-1 (19% collected), competition sort at 100 nM biotinylated-BHRF1 and 500 nM c-myc-Mcl-1 (11.3% collected), negative sort at 500 nM biotinyalted-Bfl-1 (33.3% collected), and finally, competition sort at 100 nM biotinylated-BHRF1 and 500 nM c-myc-Mcl-1 and Bfl-1 (5.5% collected). These competition sorts were done with no labeling of the c-myc labeled competitor. Clones were conventionally sequenced from the sixth and eighth pools of this sorting process. The sorts were largely not done on consecutive days, and we noticed that one clone with a growth advantage (clone BL8-11) had taken over the library by the final sort. Therefore, to get better diversity for Illumina sequencing, we went back to the fourth pool and re-cloned the library into fresh vector and fresh MC1061 cells using Gibson assembly, at which point the growth advantage disappeared. The BHRF1 library was re-sorted from this point on consecutive days with the following scheme (in which competition sorts were three-color): positive sort at 100 nM BHRF1 (1.5% collected), competition sort at 100 nM biotinylated-BHRF1 and 400 nM c-myc-Mcl-1 and Bfl-1 (3.4% collected), negative sort at 500 nM biotinylated-Mcl-1 and Bfl-1 (34.4% collected), competition sort at 100 nM biotinylated-BHRF1 and 500 nM c-myc-Mcl-1 and Bfl-1 (7.9% collected). We sequenced clones from the sixth and

eight pools from the first sorting attempt (pools BL6 and BL8). From the second sorting attempt, we sequenced 25 clones each from pools BL6' and BL7'. The accelerated growth clone from the first sorting attempt, BL8-11, was not found in the BL6' and BL7' pools.

FACS analysis samples were prepared as were samples for sorting, and were analyzed on a BD FACSCalibur.

Illumina sequencing and data processing

The KSBcl-2 Bim variant library sorted by yeast surface display was described by Foight et al. (2014). To decrease contamination by peptides that bound the antibody detection reagents, we performed two rounds of negative sorting against antibody binding on the fifth library pool. The resulting pool and the naïve, unsorted library were prepared for Illumina sampling on a Genome Analyzer II, and the data were processed as in DeBartolo et al (DeBartolo et al., 2012). The information from this screen was used in the design of the KSBcl-2 bacterial display library as described above.

Bacterial display pools were prepared for Illumina sequencing with a scheme similar to that reported by Hietpas et al. for yeast libraries (2012) First, 10 mL LB, 0.2% w/v glucose, 25 μg/mL chloramphenicol cultures of each library pool were started from glycerol stocks and grown overnight at 37 °C. The entire culture was mini prepped with a Qiagen mini prep kit and eluted in sterile water. The DNA was diluted to ~50-100 ng/μL, and 1 μL was used for the first PCR. The first PCR added an MmeI restriction enzyme site to the 5' end and a universal Illumina sequencing region on the 3' end. Primers MmeI_fwd and 3prime_rev were used in standard Phusion polymerase conditions with no annealing step and 25 cycles with a 30 second extension period in each cycle. All primer and adapter sequences are given in Table 3.9. PCR products were purified with the Qiagen PCR purification kit and eluted in 30 μL sterile water. MmeI

digestion was performed with 3.45 pmol DNA:2 μL MmeI (NEB) for 1 hour at 37 °C, followed by 20 minutes at 80 °C for enzyme deactivation. Double stranded DNA fragments containing Illumina adapters and 5-mer barcodes were then ligated onto the 5' end of the digested DNA. The adapter was double stranded through the end of the barcode, leaving a single strand "TC" overhang to anneal with the digestion product.

Barcodes were all different by at least 2 bases. 3 barcodes were used for each naïve library, 2 barcodes each for pools 1, 3, and 5, and 1 barcode each for pools 6, 7, and 8, for a total of 24 barcodes. Ligations were performed on 30 μL MmeI-digested DNA, with 4 μL 6 μM adapter, 4 μL 10x T4 ligase buffer, and 2 μL T4 DNA ligase for 30 minutes at room temperature, followed by deactivation of the ligase at 65 °C for 10 minutes. Ligation products were run on an agarose gel and ~200 bp bands were extracted with a Zymoclean Gel DNA recovery kit and eluted in 15 μL water. The second PCR amplified the ligation product and extended the 5' region to encompass the universal Illumina forward read sequencing primer. Standard Phusion polymerase conditions described above were used on 15 μL of gel-purified ligation product with 0.5 μM finalPCR_fwd and rev primers, with 25 cycles. PCR products were run an on agarose gel and extracted with the Zymoclean Gel DNA recovery kit and eluted in 30 μL water. Samples were then multiplexed and run in one lane on an Illumina Hiseq2000 with paired-end reads of 80 bp using the universal Illumina forward sequencing primer and rev_seq primer. A PhiX lane was also run for control.

Illumina sequencing data from the bacterial display library pools were filtered using in-house scripts written in Matlab and Python. First, the constant positions in the library between I2d and R4g were required to match the wild-type DNA sequence exactly, though no quality score filtering was imposed upon the constant positions. Second, the variable position bases were

required to have Illumina quality scores greater than 20 (99% base calling accuracy). Finally, sequences were sorted into pools according to their barcodes, which had to exactly match a theoretical barcode, though no quality score filtering was imposed. Sequence logos were made using Weblogo (Crooks et al., 2004).

Acknowledgements

The authors would like to thank C. Kougentakis for making the pDW363-Mcl-1, Bcl-w, Bcl-2, and Bcl-x_L constructs, O. Levsh for assistance with the fluorescence anisotropy assays and S.V. Gullá for performing the viral Bcl-2 SPOT arrays. We thank the MIT Koch Flow Cytometry core staff for assistance with FACS sorting and MIT Koch Biopolymers for peptide and SPOT array synthesis. This work was funded by the National Institute of General Medical Sciences through award R01GM110048. Computational resources to support this work were provided by the National Science Foundation under Grant No. 0821391. G.W.F. was supported by a National Science Foundation Graduate Research Fellowship.

References

Boersma, M., Sadowsky, J., and Tomita, Y. (2008). Hydrophile scanning as a complement to alanine scanning for exploring and manipulating protein-protein recognition: Application to the Bim BH3 domain. Protein Sci *17*, 1232-1240.

Certo, M., Del Gaizo Moore, V., Nishino, M., Wei, G., Korsmeyer, S., Armstrong, S.A., and Letai, A. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. Cancer Cell *9*, 351–365.

Chen, L., Willis, S.N., Wei, A., Smith, B.J., Fletcher, J.I., Hinds, M.G., Colman, P.M., Day, C.L., Adams, J.M., and Huang, D.C.S. (2005). Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. Mol Cell *17*, 393–403.

Chen, T.S., and Keating, A.E. (2012). Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods. Protein Sci *21*, 949–963.

Chen, T.S., Palacios, H., and Keating, A.E. (2013). Structure-based redesign of the binding specificity of anti-apoptotic Bcl-x(L). J Mol Biol *425*, 171–185.

Cheng, E.H.-Y., Nicholas, J., Bellows, D.S., Hayward, G.S., Guo, H.-G., Reitz, M.S., and Hardwick, J.M. (1997). A Bcl-2 homolog encoded by Kaposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. Proc Natl Acad Sci USA *94*, 690-694.

Chipuk, J.E., Moldoveanu, T., Llambi, F., Parsons, M.J., and Green, D.R. (2010). The BCL-2 family reunion. Mol Cell *37*, 299–310.

Cojohari, O., Burrer, C.M., Peppenelli, M.A., Abulwerdi, F.A., Nikolovska-Coleska, Z., and Chan, G.C. (2015). BH3 Profiling Reveals Selectivity by Herpesviruses for Specific Bcl-2 Proteins to Mediate Survival of Latently Infected Cells. J Virol. doi:10.1128/JVI.00236-15.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188–1190.

Czabotar, P.E., Lee, E.F., Thompson, G.V., Wardak, A.Z., Fairlie, W.D., and Colman, P.M. (2011). Mutation to Bax beyond the BH3 domain disrupts interactions with pro-survival proteins and promotes apoptosis. J Biol Chem *286*, 7123–7131.

DeBartolo, J., Dutta, S., Reich, L., and Keating, A.E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. J Mol Biol *422*, 124–144.

DeBartolo, J., Taipale, M., and Keating, A.E. (2014). Genome-wide prediction and validation of peptides that bind human prosurvival Bcl-2 proteins. PLoS Comput Biol *10*, e1003693.

Denisov, A.Y., Chen, G., Sprules, T., Moldoveanu, T., Beauparlant, P., and Gehring, K. (2006). Structural model of the BCL-w-BID peptide complex and its interactions with phospholipid

micelles. Biochemistry *45*, 2250–2256.

Desbien, A.L., Kappler, J.W., and Marrack, P. (2009). The Epstein-Barr virus Bcl-2 homolog, BHRF1, blocks apoptosis by binding to a limited amount of Bim. Proc Natl Acad Sci USA *106*, 5663–5668.

Dutta, S., Gullá, S., Chen, T., Fire, E., and Grant, R. (2010a). Determinants of BH3 binding specificity for Mcl-1 vs. Bcl-xL. J Mol Biol *398*, 747–762.

Dutta, S., Chen, T.S., and Keating, A.E. (2013). Peptide ligands for pro-survival protein Bfl-1 from computationally guided library screening. ACS Chem Biol *8*, 778–788.

Dutta, S., Gullá, S., Chen, T.S., Fire, E., Grant, R.A., and Keating, A.E. (2010b). Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. J Mol Biol *398*, 747–762.

Dutta, S., Ryan, J., Scott Chen, T., Kougentakis, C., Letai, A., and Keating, A.E. (2014). Potent and specific peptide inhibitors of human pro-survival protein Bcl-xL. J Mol Biol 427, 1241-1253.

Xiaofei, E., Hwang, S., Oh, S., Lee, J.-S., Jeong, J.H., Gwack, Y., Kowalik, T.F., Sun, R., Jung, J.U., and Liang, C. (2009). Viral Bcl-2-mediated evasion of autophagy aids chronic infection of gammaherpesvirus 68. PLoS Pathog *5*, e1000609.

Erlich, S., Mizrachy, L., Segev, O., Lindenboim, L., Zmira, O., Adi-Harel, S., Hirsch, J.A., Stein, R., and Pinkas-Kramarski, R. (2007). Differential interactions between Beclin 1 and Bcl-2 family members. Autophagy *3*, 561–568.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci *Chapter 2*, Unit2.9.

Feng, W., Huang, S., Wu, H., and Zhang, M. (2007). Molecular basis of Bcl-xL's target recognition versatility revealed by the structure of Bcl-xL in complex with the BH3 domain of Beclin-1. J Mol Biol *372*, 223–235.

Fire, E., Gullá, S., Grant, R., and Keating, A. (2010). Mcl-1-Bim complexes accommodate surprising point mutations via minor structural changes. Protein Sci *19*, 507-519.

Flanagan, A.M., and Letai, A. (2008). BH3 domains define selective inhibitory interactions with BHRF-1 and KSHV BCL-2. Cell Death Differ *15*, 580–588.

Foight, G.W., Ryan, J.A., Gullá, S.V., Letai, A., and Keating, A.E. (2014). Designed BH3 Peptides with High Affinity and Specificity for Targeting Mcl-1 in Cells. ACS Chem Biol. *9*, 1962-1968.

Galluzzi, L., Brenner, C., Morselli, E., and Touat, Z. (2008). Viral control of mitochondrial apoptosis. PLoS Pathog, e1000018.

Gangappa, S., van Dyk, L.F., Jewett, T.J., Speck, S.H., and Virgin, H.W. (2002). Identification of the in vivo role of a viral bcl-2. J Exp Med *195*, 931–940.

Gelgor, A., Kalt, I., Bergson, S., Brulois, K.F., Jung, J.U., and Sarid, R. (2015). KS-Bcl-2 Encoded by the Kaposi's Sarcoma-Associated Herpesvirus is Vital for Virus Reactivation. J Virol, doi:10.1128/JVI.00098-15.

Getz, J.A., Schoep, T.D., and Daugherty, P.S. (2012). Peptide discovery using bacterial display and flow cytometry. Meth Enzymol *503*, 75–97.

Grigoryan, G., Reinke, A.W., and Keating, A.E. (2009). Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature *458*, 859–864.

Halgren, T.A. (2009). Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model *49*, 377–389.

Halgren, T. (2007). New method for fast and accurate binding-site identification and analysis. Chem Biol Drug Des *69*, 146–148.

Hardwick, J.M. (2000). Cyclin' on the viral path to destruction. Nat Cell Biol *2*, E203–E204.

Hardwick, J.M., and Bellows, D.S. (2003). Viral versus cellular BCL-2 proteins. Cell Death Differ *10 Suppl 1*, S68–S76.

Hardwick, J.M., Chen, Y.-B., and Jonas, E.A. (2012). Multipolar functions of BCL-2 proteins link energetics to apoptosis. Trends Cell Biol *22*, 318–328.

Herman, M.D., Nyman, T., Welin, M., Lehtiö, L., Flodin, S., Trésaugues, L., Kotenyova, T., Flores, A., and Nordlund, P. (2008). Completing the family portrait of the anti-apoptotic Bcl-2 proteins: crystal structure of human Bfl-1 in complex with Bim. FEBS Lett *582*, 3590–3594.

Hietpas, R., Roscoe, B., Jiang, L., and Bolon, D.N.A. (2012). Fitness analyses of all possible point mutations for regions of genes in yeast. Nat Protoc *7*, 1382–1396.

Hinds, M., Lackmann, M., Skea, G., and Harrison, P. (2003). The structure of Bcl-w reveals a role for the C-terminal residues in modulating biological activity. *EMBO J 22*(7), 1497-1507.

Huang, Q., Petros, A., Virgin, H., and Fesik, S. (2003). Solution structure of the BHRF1 protein from Epstein-Barr virus, a homolog of human Bcl-2. *J Mol Biol 332*, 1123-1130.

Huang, Q., Petros, A.M., Virgin, H.W., Fesik, S.W., and Olejniczak, E.T. (2002). Solution structure of a Bcl-2 homolog from Kaposi sarcoma virus. Proc Natl Acad Sci USA *99*, 3428–3433.

Ku, B., Liang, C., Jung, J.U., and Oh, B.-H. (2011). Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. Cell Res *21*, 627–641.

Ku, B., Woo, J.-S., Liang, C., Lee, K.-H., Hong, H.-S., E, X., Kim, K.-S., Jung, J.U., and Oh, B.-H. (2008). Structural and biochemical bases for the inhibition of autophagy and apoptosis by viral BCL-2 of murine gamma-herpesvirus 68. PLoS Pathog *4*, e25.

Kvansakul, M., Yang, H., Fairlie, W.D., Czabotar, P.E., Fischer, S.F., Perugini, M.A., Huang, D.C.S., and Colman, P.M. (2008). Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. Cell Death Differ *15*, 1564–1571.

Kvansakul, M., Wei, A.H., Fletcher, J.I., Willis, S.N., Chen, L., Roberts, A.W., Huang, D.C.S., and Colman, P.M. (2010). Structural basis for apoptosis inhibition by Epstein-Barr virus BHRF1. PLoS Pathog *6*, e1001236.

Lantner, F., Starlets, D., Gore, Y., Flaishon, L., Yamit-Hezi, A., Dikstein, R., Leng, L., Bucala, R., Machluf, Y., Oren, M., et al. (2007). CD74 induces TAp63 expression leading to B-cell survival. Blood *110*, 4303–4311.

Leber, B., Lin, J., and Andrews, D.W. (2007). Embedded together: the life and death consequences of interaction of the Bcl-2 family with membranes. Apoptosis *12*, 897–911.

Lee, E.F., Czabotar, P.E., Smith, B.J., Deshayes, K., Zobel, K., Colman, P.M., and Fairlie, W.D. (2007). Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. Cell Death Differ *14*, 1711–1713.

Lee, E., Czabotar, P., Yang, H., and Sleebs, B. (2009a). Conformational changes in BCL-2 pro-survival proteins determine their capacity to bind ligands. Journal of Biological ....

Lee, E., Fedorova, A., Zobel, K., and Boyle, M. (2009b). Novel Bcl-2 homology-3 domain-like sequences identified from screening randomized peptide libraries for inhibitors of the pro-survival Bcl-2 proteins. *J Biol Chem 284*(44), 30508-30517.

Lee, E.F., Czabotar, P.E., van Delft, M.F., Michalak, E.M., Boyle, M.J., Willis, S.N., Puthalakath, H., Bouillet, P., Colman, P.M., Huang, D.C.S., et al. (2008). A novel BH3 ligand that selectively targets Mcl-1 reveals that apoptosis can proceed without Mcl-1 degradation. J Cell Biol *180*, 341–355.

Lee, E.F., Dewson, G., Evangelista, M., Pettikiriarachchi, A., Gold, G.J., Zhu, H., Colman, P.M., and Fairlie, W.D. (2014). The functional differences between pro-survival and pro-apoptotic B cell lymphoma 2 (Bcl-2) proteins depend on structural differences in their Bcl-2 homology 3 (BH3) domains. J Biol Chem *289*, 36001–36017.

Lee, E.F., Sadowsky, J.D., Smith, B.J., Czabotar, P.E., Peterson-Kaufman, K.J., Colman, P.M., Gellman, S.H., and Fairlie, W.D. (2009c). High-resolution structural characterization of a helical alpha/beta-peptide foldamer bound to the anti-apoptotic protein Bcl-xL. Angew Chem Int Ed Engl *48*, 4318–4322.

Letai, A., Bassik, M.C., Walensky, L.D., Sorcinelli, M.D., Weiler, S., and Korsmeyer, S.J. (2002). Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as

prototype cancer therapeutics. Cancer Cell *2*, 183–192.

Liu, Q., Moldoveanu, T., Sprules, T., Matta-Camacho, E., Mansur-Azzam, N., and Gehring, K. (2010). Apoptotic regulation by MCL-1 through heterodimerization. J Biol Chem *285*, 19615–19624.

Llambi, F., Moldoveanu, T., Tait, S.W.G., Bouchier-Hayes, L., Temirov, J., McCormick, L.L., Dillon, C.P., and Green, D.R. (2011). A unified model of mammalian BCL-2 protein family interactions at the mitochondria. Mol Cell *44*, 517–531.

London, N., Gullá, S., Keating, A.E., and Schueler-Furman, O. (2012). In silico and in vitro elucidation of BH3 binding specificity toward Bcl-2. Biochemistry *51*, 5841–5850.

Ojala, P.M., Tiainen, M., Salven, P., Veikkola, T., Castaños-Vélez, E., Sarid, R., Biberfeld, P., and Mäkelä, T.P. (1999). Kaposi's sarcoma-associated herpesvirus-encoded v-cyclin triggers apoptosis in cells with high levels of cyclin-dependent kinase 6. Cancer Res *59*, 4984–4989.

Ojala, P.M., Yamamoto, K., Castaños-Vélez, E., Biberfeld, P., Korsmeyer, S.J., and Mäkelä, T.P. (2000). The apoptotic v-cyclin-CDK6 complex phosphorylates and inactivates Bcl-2. Nat Cell Biol *2*, 819–825.

Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., Bruncko, M., Deckwerth, T. L., Dinges, J., Hajduk, P. J., Joseph, M. K., Kitada, S., Korsmeyer, S. J., Kunzer, A. R., Letai, A., Li, C., Mitten, M. J., Nettesheim, D. G., Ng, S., Nimmer, P. M., O'Connor, J. M., Oleksijew, A., Petros, A. M., Reed, J. C., Shen, W., Tahir, S. K., Thompson, C. B., Tomaselli, K. J., Wang, B., Wendt, M. D., Zhang, H., Fesik, S. W., and Rosenberg, S. H. (2005). An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature *435*, 677–681.

Pattingre, S., Tassa, A., Qu, X., Garuti, R., and Liang, X. (2005). Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. Cell *122*, 927-939.

Petros, A.M., Nettesheim, D.G., Wang, Y., Olejniczak, E.T., Meadows, R.P., Mack, J., Swift, K., Matayoshi, E.D., Zhang, H., Thompson, C.B., et al. (2000). Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. Protein Sci *9*, 2528–2534.

Petros, A., Medek, A., and Nettesheim, D. (2001a). Solution structure of the antiapoptotic protein bcl-2. Proc Natl Acad Sci *USA 98*(6), 3012-3017.

Petros, A., Nettesheim, D., Wang, Y., and Olejniczak, E. (2001b). Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci 9*, 2528-2534.

Polster, B.M., Pevsner, J., and Hardwick, J.M. (2004). Viral Bcl-2 homologs and their role in virus replication and associated diseases. Biochim Biophys Acta *1644*, 211–227.

Procko, E., Berguig, G.Y., Shen, B.W., Song, Y., Frayo, S., Convertine, A.J., Margineantu, D.,

Booth, G., Correia, B.E., Cheng, Y., et al. (2014). A computationally designed inhibitor of an epstein-barr viral bcl-2 protein induces apoptosis in infected cells. Cell *157*, 1644–1656.

Rautureau, G.J.P., Yabal, M., Yang, H., Huang, D.C.S., Kvansakul, M., and Hinds, M.G. (2012). The restricted binding repertoire of Bcl-B leaves Bim as the universal BH3-only prosurvival Bcl-2 protein antagonist. Cell Death Dis *3*, e443.

Roberts, A.W., Seymour, J.F., Brown, J.R., Wierda, W.G., Kipps, T.J., Khaw, S.L., Carney, D.A., He, S.Z., Huang, D.C.S., Xiong, H., et al. (2012). Substantial susceptibility of chronic lymphocytic leukemia to BCL2 inhibition: results of a phase I study of navitoclax in patients with relapsed or refractory disease. J Clin Oncol *30*, 488–496.

Rudin, C.M., Hann, C.L., Garon, E.B., Ribeiro de Oliveira, M., Bonomi, P.D., Camidge, D.R., Chu, Q., Giaccone, G., Khaira, D., Ramalingam, S.S., et al. (2012). Phase II study of single-agent navitoclax (ABT-263) and biomarker correlates in patients with relapsed small cell lung cancer. Clin Cancer Res *18*, 3163–3169.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol *234*, 779–815.

Sattler, M., Liang, H., Nettesheim, D., Meadows, R.P., Harlan, J.E., Eberstadt, M., Yoon, H.S., Shuker, S.B., Chang, B.S., Minn, A.J., et al. (1997). Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. Science *275*, 983–986.

Sinha, S., Colbert, C.L., Becker, N., Wei, Y., and Levine, B. (2008). Molecular basis of the regulation of Beclin 1-dependent autophagy by the gamma-herpesvirus 68 Bcl-2 homolog M11. Autophagy *4*, 989–997.

Smits, C., Czabotar, P.E., Hinds, M.G., and Day, C.L. (2008). Structural plasticity underpins promiscuous binding of the prosurvival protein A1. Structure *16*, 818–829.

Souers, A.J., Leverson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature Medicine *19*, 202–208.

Su, M., Mei, Y., Sanishvili, R., Levine, B., Colbert, C.L., and Sinha, S. (2014). Targeting γ-herpesvirus 68 Bcl-2 mediated down-regulation of autophagy. *J Biol Chem 289*(12), 8029-8040.

Taylor, G.S., and Blackbourn, D.J. (2011). Infectious agents in human cancers: lessons in immunity and immunomodulation from gammaherpesviruses EBV and KSHV. Cancer Lett *305*, 263–278.

Tsao, K., Debarbieri, B., and Michel, H. (1996). A versatile plasmid expression vector for the production of biotinylated proteins by site-specific, enzymatic modification in Escherichia coli. Gene *169*, 59-64.

Vogler, M. (2012). BCL2A1: the underdog in the BCL2 family. Cell Death Differ *19*, 67–74.

Watanabe, A., Maruo, S., Ito, T., Ito, M., Katsumura, K.R., and Takada, K. (2010). Epstein-Barr virus-encoded Bcl-2 homologue functions as a survival factor in Wp-restricted Burkitt lymphoma cell line P3HR-1. J Virol *84*, 2893–2901.

Yang, E., Zha, J., Jockel, J., Boise, L.H., Thompson, C.B., and Korsmeyer, S.J. (1995). Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. Cell *80*, 285–291.

Zhang, S., and Link, A.J. (2011). Bcl-2 family interactome analysis using bacterial surface display. Integr Biol (Camb) *3*, 823–831.

Zhang, S., Long, A., and Link, A.J. (2012). A comparison of two strategies for affinity maturation of a BH3 peptide toward pro-survival Bcl-2 proteins. ACS Synth Biol *1*, 89–98.

Zuo, J., Thomas, W.A., Haigh, T.A., Fitzsimmons, L., Long, H.M., Hislop, A.D., Taylor, G.S., and Rowe, M. (2011). Epstein-Barr Virus Evades CD4 T Cell Responses in Lytic Cycle through BZLF1-mediated Downregulation of CD74 and the Cooperation of vBcl-2. PLoS Pathog *7*, e1002455.

Chapter 4

Peptide binding preferences of TRAFs 2, 3, and 5

Introduction

Tumor necrosis factor receptor-associated factors (TRAFs) were originally identified in the mid-1990's as proteins that interacted with the cytoplasmic tails of TNFR super family members in yeast two hybrid assays (Cheng et al., 1995; Ishida et al., 1996a; 1996b; Rothe et al., 1994). The TRAF family in humans has since been extended to include seven members, TRAFs 1-7 (Ha et al., 2009). TRAFs mediate pathways downstream of a diverse array of signaling receptors including TNFR super family members, Toll-like receptors, the T cell receptor, interleukin receptors, NOD-like receptors, RIG-I-like receptors, IFN receptors, and TGFβ receptors (Xie, 2013). These pathways control inflammation, adaptive and innate immunity, and apoptosis, and they are important in many human diseases.

The multiple domains present in TRAFs allow them to both connect and regulate components of signaling pathways. TRAFs 2-6 contain an N-terminal RING domain, followed by 5-7 zinc-finger domains, a coiled-coil domain, and a MATH domain. Collectively, the coiled-coil domain and the MATH make up the TRAF domain, which mediates the homo- and hetero-trimerization of TRAFs (Pullen et al., 1998). TRAF1 does not have the RING domain and only contains one zinc finger, while TRAF7 does have the RING and zinc-finger domains, but contains WD40 repeats in place of the TRAF domain. The MATH domain (also known as the TRAF-C domain) binds peptides present in the cytoplasmic tails of receptors or their downstream adapter proteins. These interactions are weak (dissociation constants in the tens-to-hundreds of micromolar) for monomeric MATH-peptide interactions and therefore, generally do not take place without the aid of avidity supplied by receptor and TRAF oligomerization. This dependence on oligomerization allows for control of pathway activation by binding of extracellular ligands, which oligomerize the receptors.

The zinc-finger and RING domains of TRAFs mediate interactions with downstream effectors and are important for activation of kinase cascades. The RING finger domains possess E3 ligase activities, and TRAFs are known to mediate K63-linked poly-ubiquitination of themselves and other proteins to recruit and activate effectors. Degradative ubiquitination is also important in TRAF function and regulation, as TRAFs 1 and 2 recruit the E3 ubiquitin ligases cIAP1 and 2, which mediate K48-linked ubiquitination of TRAFs themselves and other associated proteins (Häcker et al., 2011). The cIAP proteins interact with the coiled-coil domains of TRAFs 1 and 2 (Zheng et al., 2010). Thus, TRAFs posses multiple protein-protein interaction interfaces with which to scaffold signaling complexes and utilize their E3 ubiquitin ligase activities to further modulate signaling.

TRAFs can have overlapping and distinct functions related to their differential expression and binding preferences. TRAFs 2, 3, and 6 have widespread expression, and their deletion in mice leads to death within several weeks of birth due to a variety of effects including systemic inflammation, and, in the case of TRAF6, severe osteopetrosis and developmental defects (Ha et al., 2009). TRAFs 1 and 5 are less critical, as their deletion yields viable adults, which contain defects in their innate and adaptive immune systems (Nakano et al., 1999; Tsitsikov et al., 2001). The non-critical nature of these TRAFs reflects their limited tissue distribution; TRAF1 is expressed primarily in lymphoid tissue, and TRAF5 is expressed in lung, spleen, and testis (Ishida et al., 1996a; Nakano et al., 1996; Rothe et al., 1994). TRAF4 expression is greatest during embryogenesis, particularly in neural tissues (Régnier et al., 1995). Mice with TRAF4 deletion often die as embryos, and those that survive to adulthood have severe developmental defects including impaired neural tube closure, axial skeletal malformations and tracheal ring defects (Régnier et al., 2002; Shiels et al., 2000).

These differential effects of TRAF deletion arise from functions in different pathways, as well as redundant or opposing roles in some of the same pathways. An example of the latter is the perhaps best-studied pathway of TRAF function, that downstream of the TNFR super family member CD40. CD40 contains two TRAF binding sites (TBSs) in its cytoplasmic tail, one shared by TRAFs 1, 2, 3, and 5, and another bound by TRAF6 (Pullen et al., 1998). Downstream of CD40 binding, TRAF6 activates the canonical NF-κB pathway, while TRAFs 2 and 5 can activate both the canonical and non-canonical NF-κB pathways (Hauer et al., 2005). TRAF3, by competing for binding to the same site on CD40, can block activation mediated by TRAFs 2 and 5, but not by TRAF6 (Hauer et al., 2005). Activation of p38 kinase and JNK downstream of TRAFs 2, 5, and 6 involves assembly of multi-protein complexes including cIAP, IKKγ, E2 ligases, and MAP3Ks, including TAK1 and MEKK1 (Häcker et al., 2011). TRAF3 inhibits dissociation of kinases TAK1 and MEKK1 from complexes with TRAFs 2/5 and 6, thus preventing downstream activation of JNK and p38 MAPK (Matsuzawa et al., 2008). TRAF3 inhibition is released by the K48-linked poly-ubiquitination and subsequent degradation of TRAF3 mediated by cIAP. In non-canonical NF-κB activation, TRAF2 and cIAP-mediated degradation of TRAF3 releases NIK, leading to activation of NF- κB2 (Vallabhapurapu et al., 2008). Therefore, due to differential interactions with downstream proteins, TRAFs 2/5 and 3 have opposing effects on CD40 signaling, despite binding the same TBS.

A significant component of TRAF functional specificity, however, does arise from the interaction specificity of their MATH domains. Evolutionary analysis of TRAFs suggests that TRAFs 4 and 6 are the more ancient homologs, with examples of TRAF6 present in insects and TRAF4 in *cnidaria* and early chordates (Zapata et al., 2007b). TRAFs 1, 2, 3, and 5 appear to be more recent homologs largely found in vertebrates (Zapata et al., 2007b). The pairs of TRAFs 1

and 2 and TRAFs 3 and 5 have highest sequence similarity to each other in their MATH domains, suggesting that they may have evolved through gene duplication, and this idea is supported by the fact that these pairs have also been found to hetero-oligomerize (Pullen et al., 1998). Different core binding motifs have been defined for TRAFs 1/2/3/5 and TRAF6. TRAFs 1, 2, 3, and 5 share many of the same binding partners, recognizing a major ((P/S/A/T)x(Q/E)E) and minor (PxQxxD) motif (Ye et al., 1999). A number of TRAF 1/2/3/5 binding partners also fit the motif PxQxT, suggesting that these TRAFs can accommodate a variety of binding strategies C-terminal to the critical central Q/E position (Devergne et al., 1996). TRAF6 recognizes a different motif (PxExx(Aromatic/Acidic)), and binding partners of TRAF4 have not been directly characterized, though TRAF4 may bind some of the same sites as TRAF6 (Marinis et al., 2011; Ye et al., 2002; Zepp et al., 2012). However, at least one example of a TBS exists that binds both TRAFS 1/2/3/5 and TRAF6. TRAF6 association with the Epstein-Barr virus protein LMP1 was shown to occur at the same PQQATD site used by TRAFs 1/2/3/5 (Arcipowski et al., 2011). This site fits the minor TBS motif of PxQxxD identified for TRAF2, and is close to the TRAF6 motif PxExx(Ar/Ac), suggesting that peptides that fit the more general motif Px(Q/E)xx(Ac) might be accessible to both groups of TRAFs. These results suggest that there is considerable plasticity in the established motifs, and a more detailed examination of binding capabilities is needed to create more accurate binding models.

The classification of TRAF binding specificity has focused on the core motifs described above, but the literature contains numerous hints that peptide regions outside of the core motif can influence affinity and specificity. A number of these examples support the idea that TRAF3, especially, makes interactions outside of the peptide core. An early report by Devergne et al. showed that mutation of either the core motif proline or glutamine to alanine in LMP1 abrogated

TRAF1 and TRAF2 binding, but both mutations were necessary to lose TRAF3 binding (Devergne et al., 1996). This suggested that TRAF3 might make use of additional interactions to bind LMP1. Indeed, this was shown to be the case for CD40, the human receptor that LMP1 mimics. Substitution SPOT arrays used to test binding of short (core binding motif) and C-terminally extended CD40 peptides showed more relaxed binding preferences in the core for TRAF3 binding when the C-terminal extension was present, relative to the shorter peptides (Pullen et al., 1999a). Structures of TRAF3 bound to long peptides from TANK and BAFF-R also show extended interactions C-terminal to the peptide core, with the peptide wrapping around the MATH domain (Li et al., 2002; Ni et al., 2004). However, the relative contributions of core versus extended peptide regions to TRAF3 binding affinity are unknown. Little is known about the importance of peptide sequence N-terminal to the core motif, but one recent study found that a histidine to tyrosine mutation three residues N-terminal to the core proline in BAFF-R increases binding affinity to TRAFs 2, 3, and 6 (Hildebrand et al., 2010). Signaling through BAFF-R prevents apoptosis, and this mutation was found in a subset of patients with non-Hodgkin lymphoma. As more interactions are characterized, further examples of interactions mediated by peptide regions outside the core are likely to be discovered. Interaction preferences outside the core could vary widely due to lower sequence identity between the TRAFs outside of the core binding groove. Therefore, these interactions may be an important source of specificity.

A better understanding of TRAF binding preferences could be applied to interactome prediction and the design of specific peptide inhibitors. Given the diversity in TRAF binding preferences, it is not straightforward to identify the TBS in an interaction partner identified by pull-down or other methods. Because TRAFs are often present in multi-protein assemblies, and themselves have more than one interaction interface, it is not always clear whether interactions

are with the MATH peptide binding groove. Better models of TRAF binding preferences would allow identification of TBSs on known partners, as well as prediction of new partners. Knowledge of the interaction preferences of each TRAF protein would allow prediction of the relative affinity of TRAFs for a given TBS, providing hypotheses about signaling mechanisms. Specific inhibitors of TRAFs would provide the means to test such hypotheses. It is a common practice to swap cytoplasmic domains of TNFR super family members in order to put the downstream effects under control of a different extracellular domain/ligand pair (Arch and Thompson, 1998; Hildebrand et al., 2010). It is possible that specific TBSs could be swapped into an interaction partner of interest to examine pathway requirements for individual TRAFs.

Due to the importance of TRAFs in many disease states, specific peptide inhibitors would be useful therapeutic leads. Several groups have demonstrated this idea for TRAF6. TRAF6 signaling downstream of RANK is responsible for osteoclast differentiation, which can lead to osteoporosis and cancer-induced bone lesions when it occurs aberrantly (Arron and Choi, 2000). Ye et al. fused peptides corresponding to the TRAF6 binding sites on RANK to a cell penetrating peptide and showed that treating osteoclast precursor cells with these peptides reduced NF-κB activation and associated osteoclast differentiation (Ye et al., 2002). Later variations included fusion of palmitate to the RANK peptide for cellular delivery, and PEGylation of the peptide for enhanced bioavailability (Akhtar et al., 2012; Poblenz et al., 2007). Inhibiting TRAF1 is also of interest, as its overexpression is associated with several B-cell malignancies. Its role in these leukemias and lymphomas may be to prevent apoptosis by recruiting anti-apoptotic proteins like the cIAPs to activated TNFRs (Zapata et al., 2007a). The roles of TRAF3 and TRAF6 in autoimmunity and inflammation make them attractive targets for the treatment of related diseases.

This chapter will present an initial characterization of the peptide binding preferences of TRAFs 2, 3, and 5. Because these TRAFs are close homologs, the possibility of differences in binding preferences was of interest. I compared the sequence identity and physicochemical characteristics of the TRAF binding sites. I also developed a bacterial surface display protocol for characterizing the peptide binding preferences of TRAFs. Screening single and double point mutant libraries of peptides from CD40 and TANK yielded a variety of interesting insights into differences between TRAFs 2, 3, and 5, as well as leads for the development of specific peptide inhibitors.

Results and Discussion

Comparison of TRAFs by sequence identity of the MATH domain and peptide-binding site

Sequence identity in the TRAF MATH domains and peptide binding sites illustrate which TRAF family members are most similar to each other (Figure 4.1). The peptide-binding site considered includes the 18 TRAF residues within 7 Å of the 5-mer peptide core ('PVQET') in a structure of CD40 bound to TRAF2. TRAF6 has the lowest sequence identity to the other TRAFs, with ~30% identity over the full MATH domain, and only ~20% identity in the core peptide binding site. TRAF4 has ~10% higher sequence identity to TRAFs 1, 2, 3, and 5 than to TRAF6 over both the MATH domain and the peptide binding site. TRAFs 1, 2, 3, and 5 have high identity (>52%) over the MATH domain, and very high identity in the core-binding site (>78%). Therefore, the TRAF 1/2/3/5-binding motif contacts a nearly identical set of residues when binding these four homologs. The lower sequence identities for the full MATH domain suggest that differences between these four homologs outside the core-binding site could influence binding specificity.

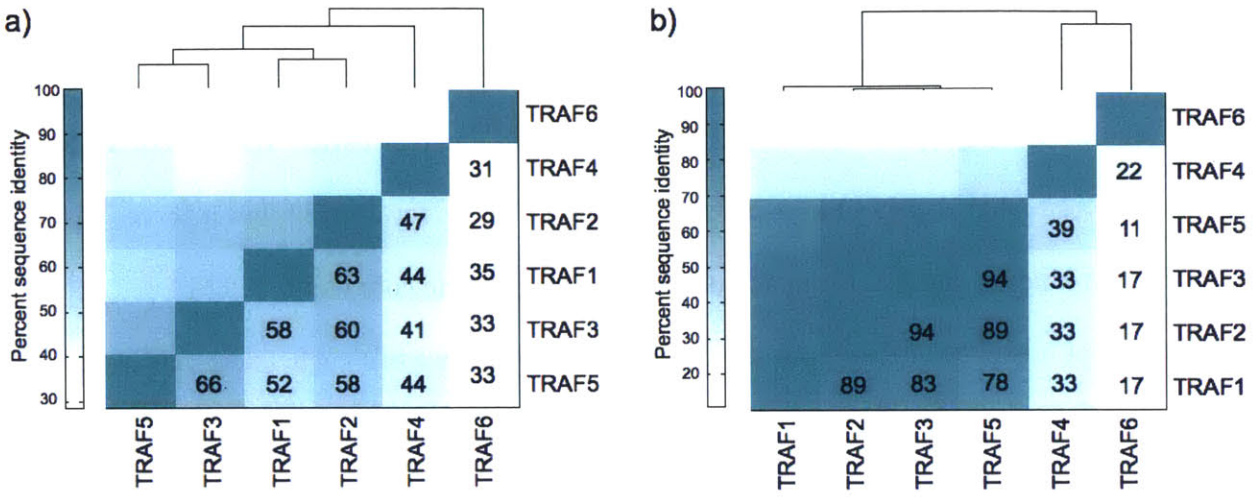


Figure 4.1 Sequence identities show similarities between TRAFs 1, 2, 3, and 5. a) Sequence identities over the entire TRAF MATH domain. b) Sequence identities in the core peptide binding site. Residues included are within 7 Å of the core 'PVQET' peptide sequence in a structure of CD40 bound to TRAF2.

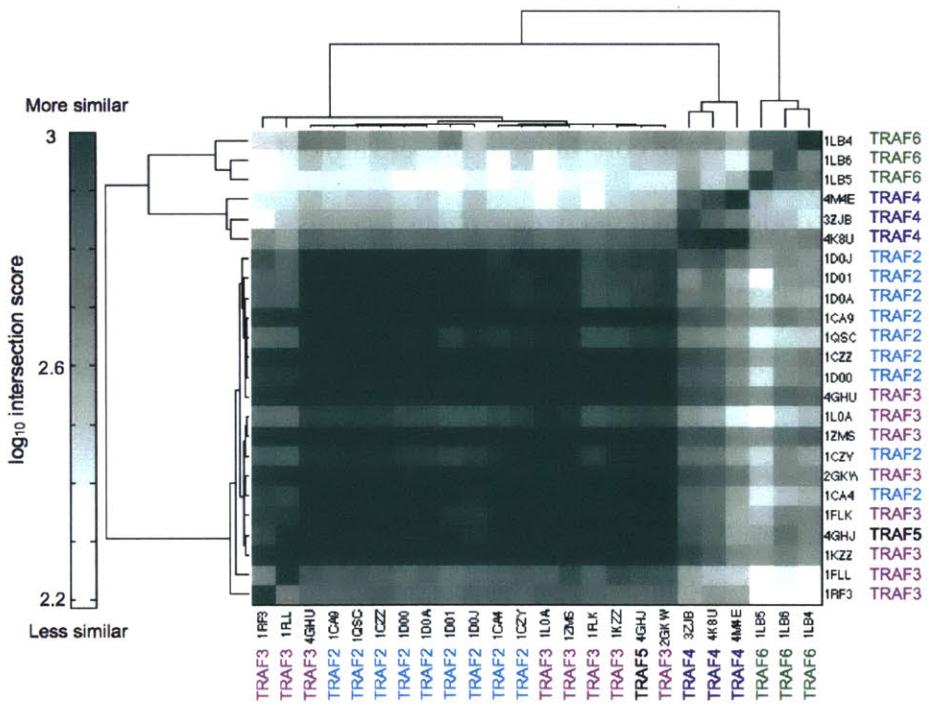
Comparison of TRAFs by physicochemical properties of the peptide-binding site

In order to compare the binding environment on the surface of the TRAFs, I looked at the physicochemical properties around the core-binding site and around an exosite. An exosite is a binding site outside of the region bound by the core peptide motifs. I aligned all available TRAF structures to a structure of TRAF3 bound to a long peptide from TANK and used SiteMAP to characterize binding potential on the TRAF surface for hydrophobic, hydrogen bond donor and hydrogen bond acceptor groups as described in the Methods (Halgren, 2009; 2007). The core-binding site was defined as the region within 6 Å of the Cβ residues of the core 'VPIQCTD' sequence of TANK. Though TRAF6 binds peptides in a different orientation than TRAF3 does, its core-binding motif residues fall within this same region. The core-binding site comparison (Figure 4.2a) reflects the sequence identity results in that TRAFs 2, 3, and 5 have highly similar binding sites, and TRAF4 and TRAF6 are each unique and not very similar to any other TRAFs. Two TRAF3 structures, 1FLL and 1RF3 (the bottom rows), have lower similarity, but these

structures are of low quality and have long peptides bound in unique hairpin conformations that may be influenced by crystal contacts.

The exosite region was defined as the area surrounding the 'EALF' residues of TANK. This is an exosite that is only of known functional significance for TRAF3, but it serves as an example of a binding site that is outside of the site bound by the core motif. SiteMAP comparison of this exosite region for the different TRAFs shows that each TRAF has unique surface features in this region (Figure 4.2b). TRAF 2, 3, and 5 structures segregate into their own clusters, in contrast to the comparison of the core-binding sites, where they were intermingled. This demonstrates that outside of the core binding site, TRAFs 2, 3, and 5 have differences that may be of relevance for peptide binding specificity.

a)



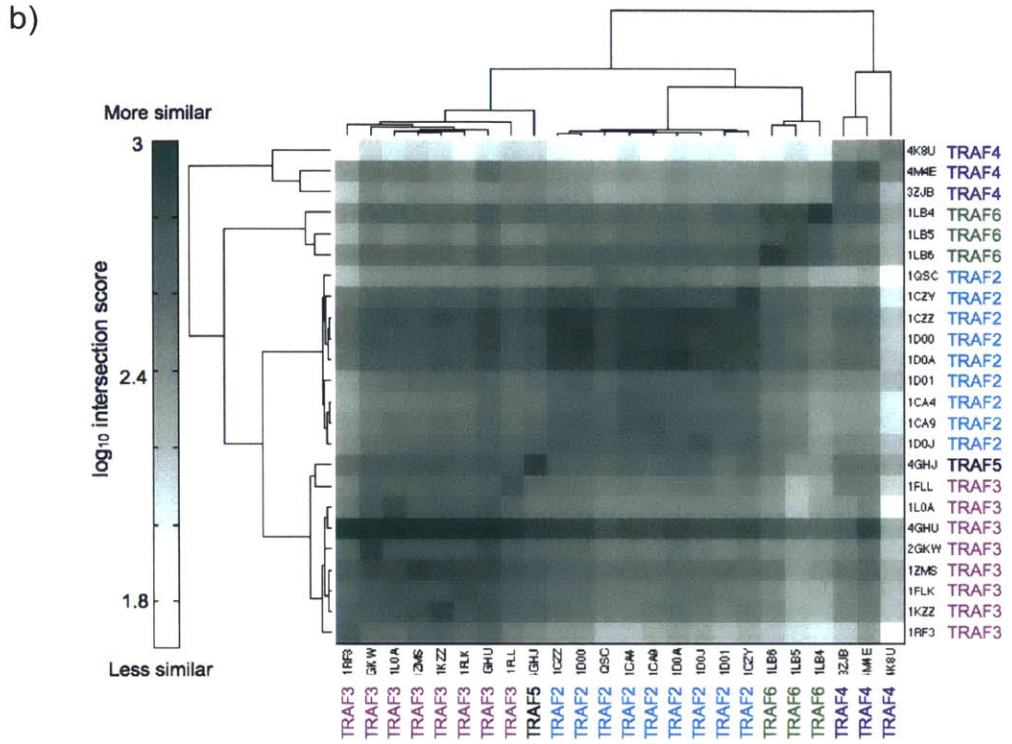


Figure 4.2. Comparison of the physicochemical characteristics of core and exosite regions of TRAF binding grooves by SiteMAP. The intersection score is a measure of the similarity of binding potential for hydrophobic, hydrogen bond acceptor and hydrogen bond donor groups between structures. a) Comparison of the core peptide-binding groove. b) Comparison of an exosite that is bound by a C-terminal extension of TANK in a structure of TRAF3 (1L0A) (Li et al., 2002).

Development of a surface display system for TRAF binding

To compare binding preferences of TRAFs in a large-scale manner, I created a display method for peptide libraries in the eCPX *E.coli* surface display system used in Chapter 3. TRAF-peptide binding is weak, and there are several wash steps and a significant time (on the order of 1-2 hrs) between protein incubation and analysis by FACS. For these reasons, it was expected that it would be challenging to observe binding on cell surface display systems. Therefore, I first tried to use TRAF domains oligomerized via their natural coiled-coil domains to provide avidity and increase binding signal. These TRAF constructs included both the TRAF-N coiled-coil domains and the TRAF-C MATH domains, and TRAF2 and TRAF3 were soluble and trimeric

when analyzed by size exclusion chromatography. The TRAF5 coiled-coil construct did not express well. However, the trimeric TRAF constructs displayed weak binding to peptides displayed as fusions to eCPX on the surface of *E. coli*. In contrast, MATH-domain only constructs of TRAFs 2, 3, and 5 demonstrated much better binding. This was contrary to my expectations, but the lower binding exhibited by the trimeric TRAF constructs could arise from steric issues in binding the peptide, or in being bound by the bulky labeling reagent, streptavidin-phycoerythrin. A MATH-domain only construct of TRAF1 was insoluble, but could be rescued by fusion to a GB1 domain. However, GB1 showed high non-specific binding to the *E. coli* surface. The TRAF domain of TRAF1 has been reported to be insoluble even when expressed in insect cells, and no structures of it have yet been solved (Pullen et al., 1999b). TRAFs trimerize via their coiled-coil domains and through interfaces between their MATH domains. The MATH domains of TRAFs 2, 3, and 5 appeared monomeric by size exclusion at an injection concentration of 10 μM and have also been reported to be monomeric by analytical ultracentrifugation experiments (Pullen et al., 1999b). However, it remains possible that MATH-only constructs could trimerize at the higher local concentrations created by binding peptides on the *E. coli* surface.

The manner in which I fused peptides to eCPX also strongly influenced my ability to see TRAF binding. In our past work with BH3:Bcl-2 (Chapter 3) and polyproline:EVH1 interactions (unpublished), we displayed peptides on the N-terminus of eCPX with an epitope tag N-terminal to the peptide included for use as an expression control. This orientation did not give a strong binding signal for TRAFs. I saw much-improved TRAF binding when I displayed the peptide on the C-terminus of eCPX (Figure 4.3a). Initially, I included an epitope tag C-terminal to the peptide for expression control. However, screening libraries in this construct resulted in strong

enrichment for stop codons after the core peptide motif, and consequent loss of the epitope tag. Therefore, I proceeded with constructs that did not contain an expression control. In the future, it is possible that an epitope tag could be placed between the eCPX and the peptide, but this will require optimization of linkers to minimize competition between TRAF and labeling reagent binding. Without an expression control, it is possible that peptide libraries sorted for clones displaying high binding signal could instead enrich clones with high expression. However, I proceeded with the one-dimensional system with the idea that most single and double mutations would be unlikely to have large effects on peptide expression.

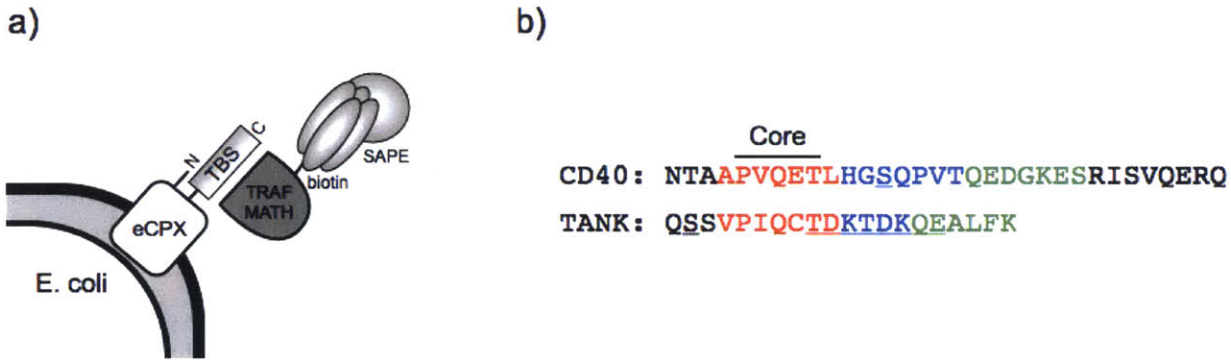


Figure 4.3. Cell surface display system and peptide constructs. a) Peptides ('TBS') were displayed on the C-terminus of eCPX without an expression control. Binding of biotinylated TRAF MATH domains was detected with streptavidin-phycoerythrin (SAPE). b) TBSs from CD40 and TANK were displayed and mutated in the libraries. The core site is labeled. The three, 7-mer segments that are colored (and underlined for the middle TANK segment) were mutated to generate all single and double point mutants. The underlined serines were mutated from the wild-type cysteine.

I displayed TRAF binding sites from CD40 or TANK on eCPX (Figure 4.3b). As controls for the ability to detect changes in binding affinity, I used threonine to alanine mutations at the conserved T in the PxQxT motif of CD40 and TANK. This mutation has been reported to substantially weaken binding of TRAFs 1, 2, and 3 to CD40 (Pullen et al., 1999a). However, I did not initially see a difference in binding between to the wild-type and mutant

peptides when displayed on *E. coli*. After mutating a cysteine residue to serine in both CD40 and TANK, I was able to see the expected difference in binding (serines underlined in Figure 4.3b). TRAFs 1, 2, 3, and 5 have cysteines in the binding groove positioned to form disulfides with cysteines occurring in the peptides both N-terminal (in TANK) and C-terminal (in CD40) to the core motif. Binding of the MATH domains of TRAFs 2, 3, and 5 at 15 μM to the cysteine-to-serine versions of the wild-type and threonine-to-alanine mutants of CD40 and TANK is shown in Figure 4.4. The cysteine-to-serine mutants shown in Figure 4.3b will henceforth be referred to simply as CD40 and TANK.

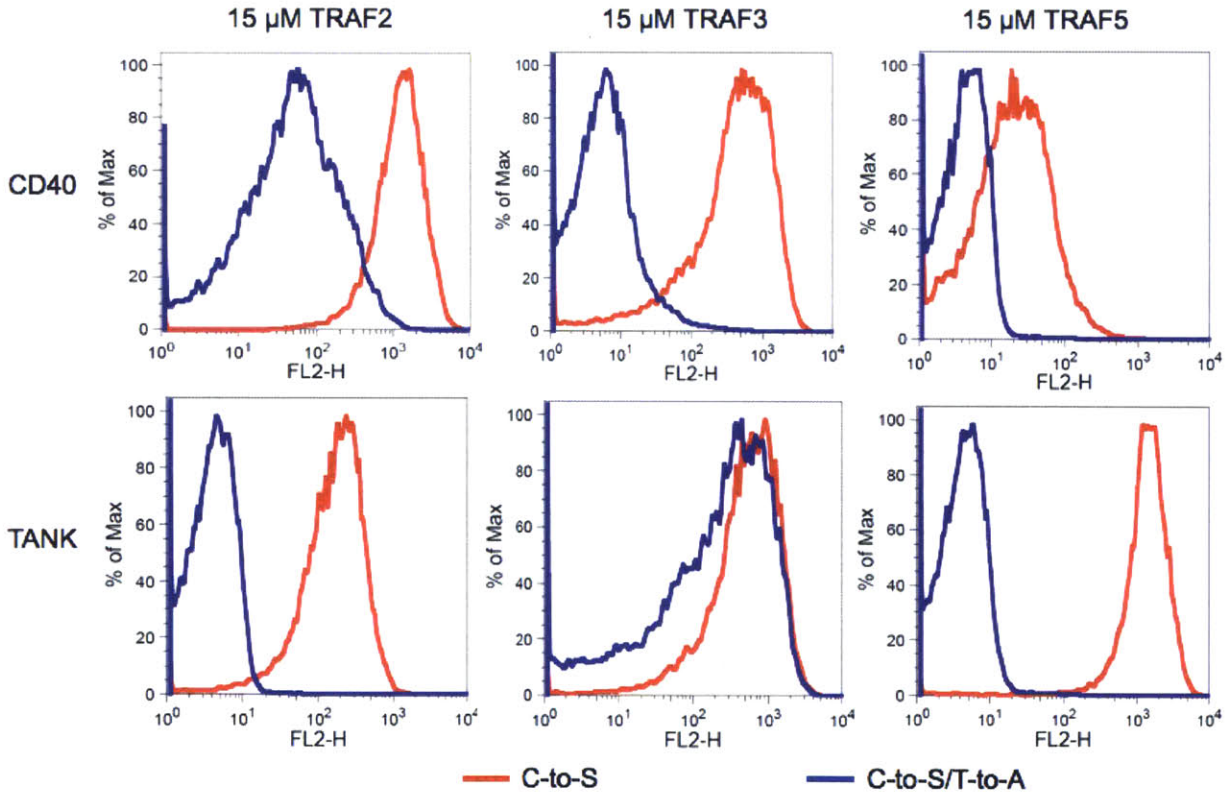


Figure 4.4. Binding of TRAFs 2, 3, and 5 to control peptides displayed on *E. coli*. Each column is labeled with the protein condition used for the analysis. The top row shows binding to the CD40 C-to-S (red) and C-to-S/T-to-A (blue) peptides. The bottom row shows binding to the TANK C-to-S (red) and C-to-S/T-to-A (blue) peptides. X-axis is the binding fluorescence, and y-axis is proportional to the number of events.

Screening of single and double mutant libraries of CD40 and TANK

To explore the binding preferences of TRAFs 2, 3, and 5, I created libraries of single and double point mutants of CD40 and TANK. The libraries included all single mutants in the 21-mer (CD40) or 18-mer (TANK) colored regions in Figure 4.3b. All double mutants in three, 7-mer segments of different colors (or underlined for the middle TANK section) were also included. I sorted the libraries by FACS at two different concentrations of TRAF (3 μM and 10 μM) to give different degrees of stringency. The library sorting schemes, including the number of rounds, are illustrated in Figure 4.5.

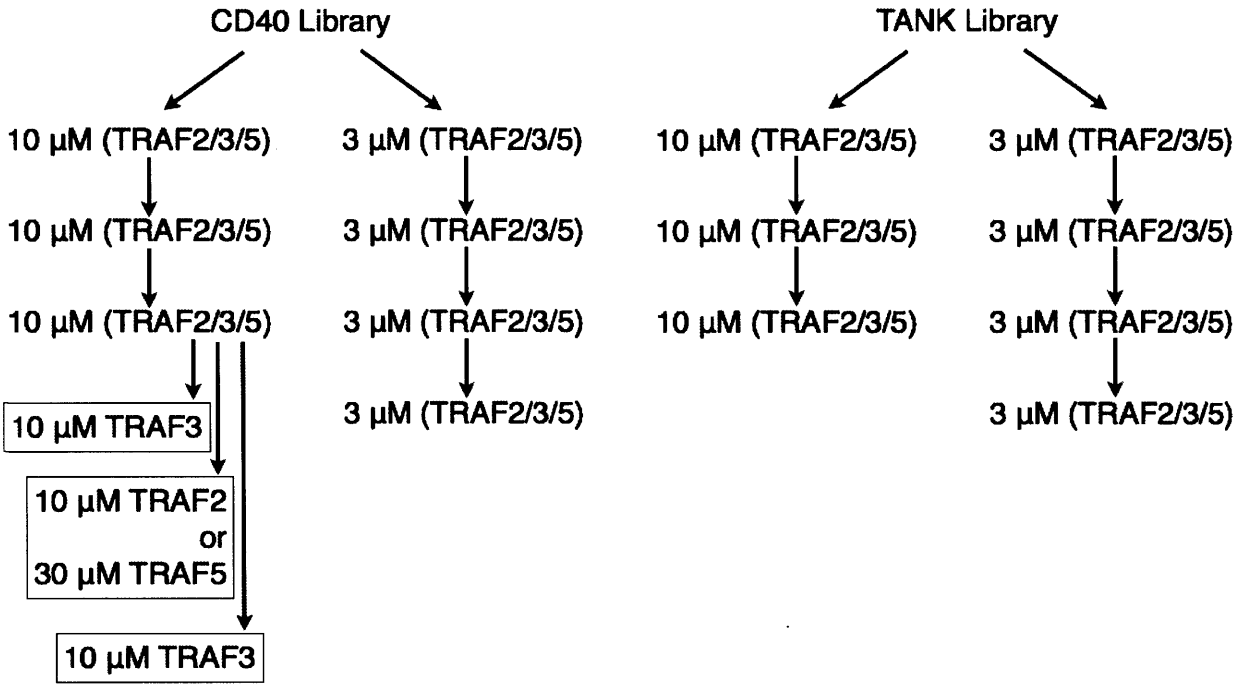


Figure 4.5. Sorting schemes for the CD40 and TANK libraries. All rounds not boxed were affinity sorts for the given concentration of TRAF2, TRAF3, or TRAF5. The four negative sorts are boxed. These used the last round of the CD40 10 μM sorts as input and screened for non-binders at the concentrations of TRAFs listed in the box.

Most sorting paths resulted in final pools showing ~60-95% of events above the noise level of the binding signal. Single wild-type clones at saturating TRAF concentrations have ~95% of events above this level. This indicates that the final pools in each sorting path predominantly contain binders. The final pools from the 10-μM sorts were analyzed for binding to 10 μM of each TRAF. The FACS distributions of events according to binding signal are shown in Figure 4.6. The enriched TANK library pools showed no specificity between TRAFs 2, 3, and 5, in that all three TRAFs bound equally to pools enriched for binding the other TRAFs. In contrast, the enriched CD40 library pools showed varying degrees of specificity. The CD40 pool enriched for TRAF3 binding showed greatly reduced binding to TRAF2 and TRAF5. The CD40 pools enriched for binding to TRAFs 2 and 5 showed modestly reduced binding to TRAF3. However, there was no apparent specificity between TRAFs 2 and 5, as both bound equally to CD40 pools enriched for binding either TRAF. This led me to perform one round of specificity sorting on the final 10-μM CD40 library pools at the concentrations shown in Figure 4.5. The goal of these specificity sorts was to further enrich specific clones from pools that had demonstrated specificity between TRAF3 and TRAFs 2/5.

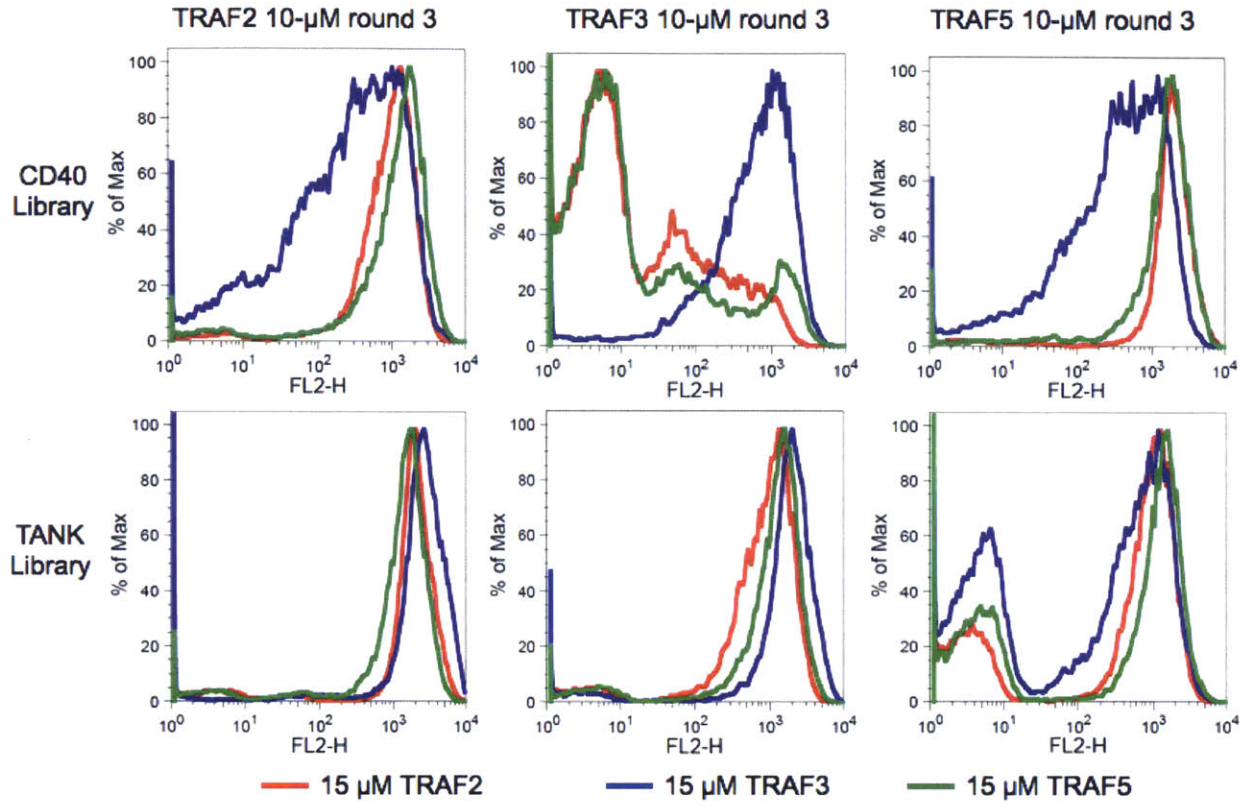


Figure 4.6. TRAFs 2, 3, and 5 binding to the final 10-μM CD40 and TANK library pools. Columns are labeled with the sorting pool analyzed (10-μM sorts round 3 for each TRAF), with CD40 library pools on top and TANK library pools on bottom. The conditions used for analysis are colored: 15 μM TRAF2 (red), 15 μM TRAF3 (blue), 15 μM TRAF5 (green). X-axis is the binding fluorescence, and y-axis is proportional to the number of events.

Enrichment analysis of CD40 and TANK library sequences

To analyze the enrichment of variants in each sort, I Illumina sequenced each library pool. I did not sequence all of the CD40 library 3-μM pools for TRAF3 and TRAF5, as I believed them to be contaminated with clones with a growth advantage (see Methods), and I abstain from any analysis of those screens here. I calculated a functional score for each peptide variant similarly to Starita et al. (Starita et al., 2015). See Methods for details. Briefly, an enrichment ratio was calculated from the frequency of each variant in each round relative to the naïve library. A line was fit to the enrichment ratios across rounds, and the slope of this line was converted to a functional score, after correction for non-specific carryover. The functional scores

were normalized by the wild-type sequence functional score (1), such that variants with functional scores >1 are enriched relative to wild type, and variants with scores <1 are reduced relative to wild type. This approach to measuring enrichment of variants from deep sequencing data has been used by other groups (Araya et al., 2012; Starita et al., 2015). I expected to be able to at least see the strongest trends from this analysis. In the future, a more rigorous analysis is needed of how to treat noise from sources such as under-sampling of the naïve library.

The functional scores for single mutants are plotted as heatmaps for the 10-μM sorts of the CD40 and TANK libraries in Figure 4.7. In the CD40 libraries (Figure 4.7a), the TRAF2 and TRAF5 preferences are notably similar, and different from the TRAF3 preferences. This agrees with the FACS analysis performed on the final pool of these libraries, which showed that TRAF2 and TRAF5 bound each other's pools equally, but bound weakly to the TRAF3 pool and *vice versa* (Figure 4.6). In the TANK libraries (Figure 4.7b), the overall patterns are similar for all three TRAFs, in agreement with all three TRAFs binding each other's pools equally in Figure 4.6.

The highest functional scores for single mutants in the CD40 libraries (Figure 4.7a) appear at the first position (alanine in wild type) of the peptide and the first glutamate in the core. At the alanine position, TRAF2 and TRAF5 have high scores for hydrophobic residues and cysteine, while those residues score similarly to wild type for TRAF3 binding. Conversely, TRAF3 has high functional scores for hydrophobic residues at the glutamate position, which score similarly to wild type for TRAF2 and TRAF5. Other CD40 mutants with high functional scores that show agreement of TRAF2 and TRAF5 binding preferences include leucine-to-aspartate and the last aspartate-to-isoleucine. It is possible that these single mutant preferences could contribute to the specificity seen for binding the CD40 library pools; however, there are

many more double mutants in the libraries than single mutants, so it is important to consider all of the data when looking for specificity mechanisms. In the TANK libraries, all three TRAFs are permissive at the cysteine (equivalent to the first glutamate in CD40). The strongest preference is for stop codons after the core binding sequence. TRAF3 also has high functional scores for cysteines after the core.

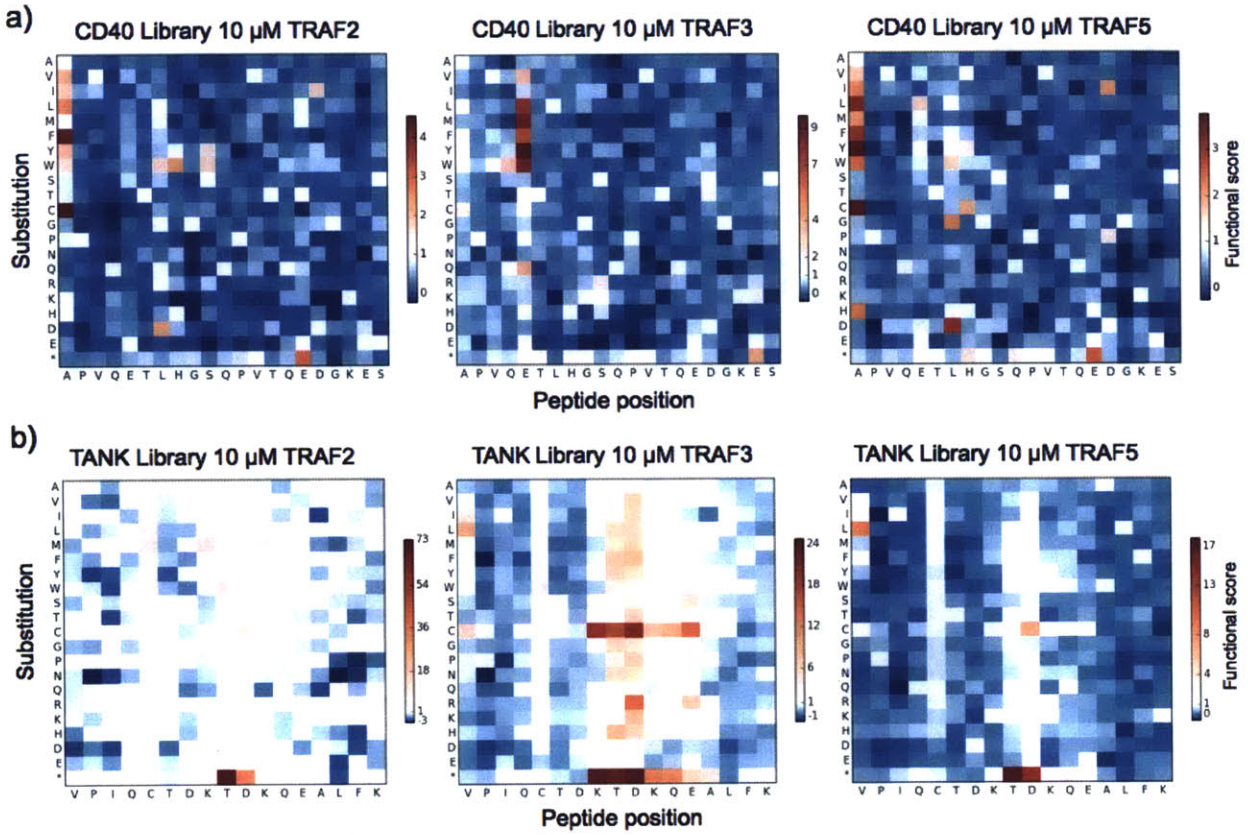


Figure 4.7. Functional scores for peptide single mutants. a) CD40 library sorts performed at 10 μM. b) TANK library sorts performed at 10 μM. The wild-type residue has a score of 1 (white) and is printed along the x-axis.

To look at trends seen in the most highly enriched sequences, I made sequence logos from variants (both single and double mutant) with functional scores greater than two standard deviations above the mean functional score (~500-1100 sequence in each sort). Figure 4.8 shows the logos for the 10-μM CD40 sorts and the TRAF2 3-μM CD40 sort. For comparison, I

included SPOT arrays from the literature that show TRAF2 and TRAF3 from insect cell lysates binding to single mutants of CD40 peptides (Pullen et al., 1999a). The sequence logo for TRAF3 emphasizes that it is very tolerant of mutation at the first glutamate position, which agrees with the SPOT arrays. The 10-μM TRAF2 logo emphasizes that this protein prefers to bind the wild-type core sequence, 'PVQET', in agreement with its SPOT array. However, the 3-μM TRAF2 logo shows some preference for hydrophobic residues at the first glutamate position, similar to the TRAF3 logo. The TRAF2 and TRAF5 logos show a low enrichment of stop codons ('X') in the positions after the core, and TRAF2 also shows a preference for tryptophans.

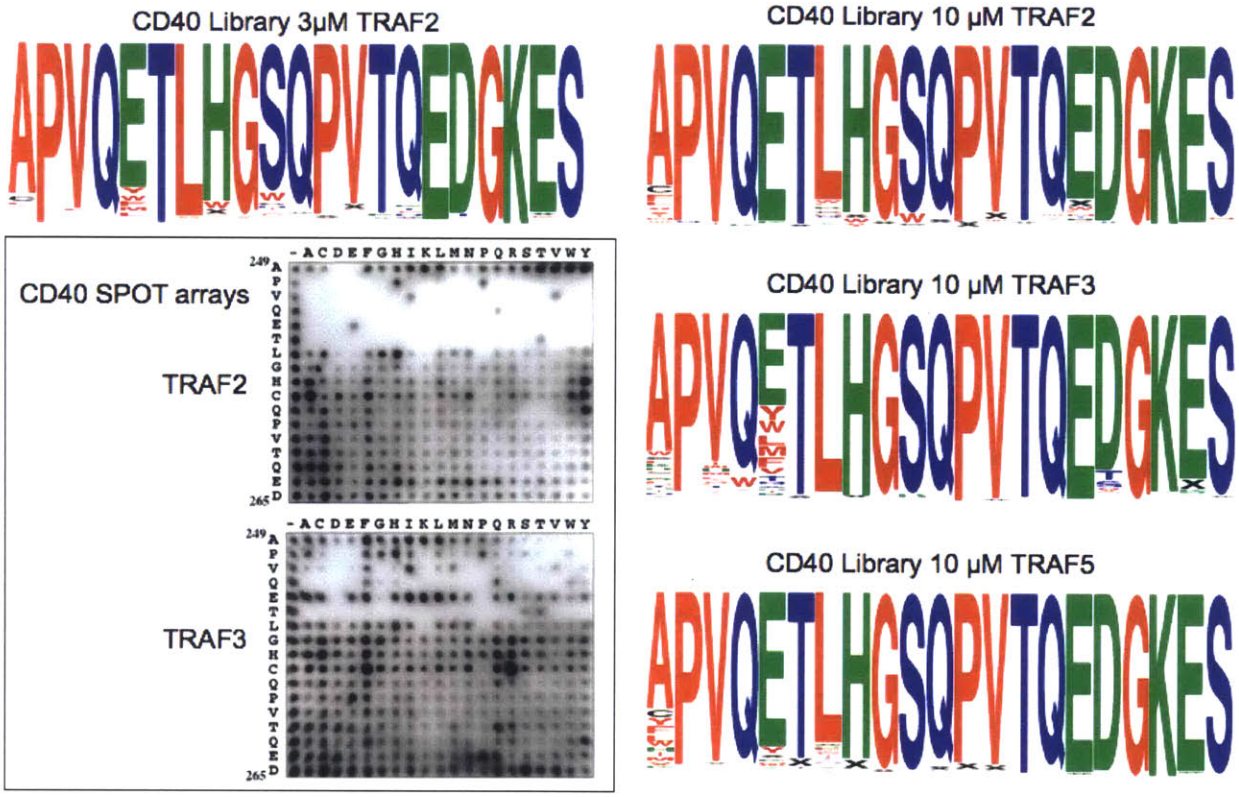


Figure 4.8. Sequence logos and SPOT arrays of CD40 peptides. The sequence logos show the frequency of residues at the peptide positions varied in the library, with the wild-type residue being the most frequent in all cases. The sequences used to build the logos had functional scores >2 standard deviations above the mean. 'X': stop codon. Inset: SPOT arrays for CD40 single mutants binding to lysates containing TRAF2 and TRAF3 reproduced from Pullen et al. 1999.

Sequence logos for TANK library sequences with functional scores greater than two standard deviations above the mean are shown in Figure 4.9. Both of the affinity sorts look very similar for each TRAF, suggesting that the mechanisms employed to give high binding signal at 10 μM were also sufficient to give high binding signal at 3 μM. TRAF3 and TRAF5 show a strict preference for the wild-type 'PIQCT' core sequence. TRAF2 allows some variation of the core cysteine to hydrophobic residues. The strongest trend for all three TRAFs is a preference for stop codons and cysteine starting two residues after the core 'T' and peaking at the second aspartate. Notably, this second aspartate position is at an equivalent position to the cysteine found in wild-type CD40, which we mutated to serine to reduce background binding signal. Other residue preferences distributed similarly to the stop codon and cysteines (black) include hydrophobic residues (red) and basic residues (arginine, lysine, or histidine, in green). This distribution of preferences suggests that the peptide after the core binding sequence may not be bound in one position. Instead, whatever TRAF surface features are attracting these groups (hydrophobic, basic, cysteine, and negative charge from the terminus), are accessible from a range of peptide positions spanning the wild-type 'KTDKQ' sequence. These preferences are discussed below in the context of TRAF structures. The agreement between the TANK library results for all three TRAFs suggests that these sequences could be general solutions for binding these TRAFs, in contrast to the CD40 library results, which showed more paralog-specific variation.

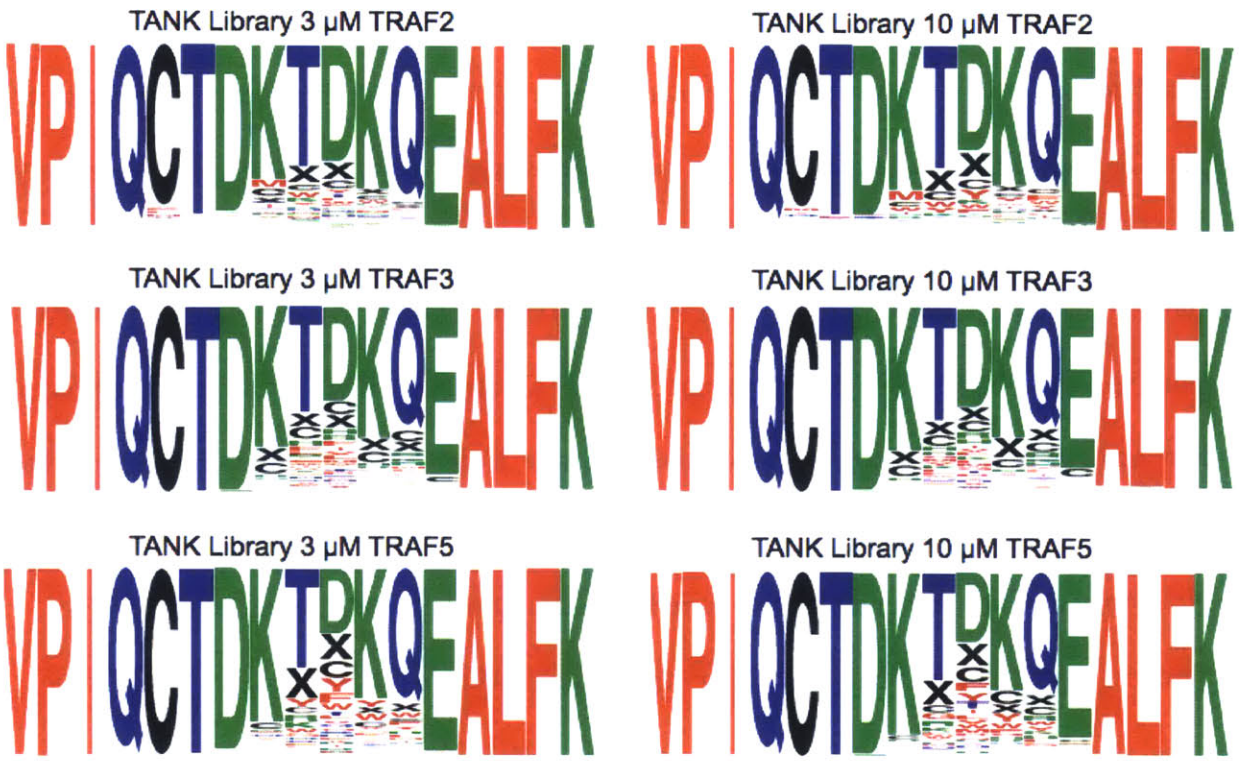


Figure 4.9. Sequence logos from TANK library pools. The sequence logos show residue frequency and were built from sequences with functional scores >2 standard deviations above the mean. Left, 3-μM sorts. Right, 10-μM sorts. 'X': stop codon.

Structural hypotheses for peptide binding preferences

Cysteines were enriched in both the CD40 and TANK libraries at positions both N-terminal and C-terminal to the core binding motif. TRAFs 1, 2, 3, and 5 all contain cysteines located in proximity to the N-terminus of the peptide. TRAFs 1, 2, 3, 5, and 6 all contain a conserved cysteine proximal to the C-terminus of the peptide core. These cysteines are highlighted in magenta on a structure of TRAF3 bound to a TANK peptide in Figure 4.10a. TRAF2 and TRAF5 demonstrated a preference for binding cysteines at the first varied position in the CD40 library (the residue before the conserved motif proline), but not in the TANK library. In analyzing the frequencies of mutations in the raw sequencing data, I observed that TRAF2 sorting also enriched a cysteine at the position two residues N-terminal to the conserved

proline in the TANK library. This position was not varied in the library, so these sequences were excluded in my analysis of the theoretical library sequences. The wild-type cysteine in TANK, which I mutated to serine, is three residues N-terminal to the core proline. The TRAF3 binding partner CARDIF also has a cysteine three residues N-terminal to the core proline. Figure 4.10b shows a crystal structure in which this cysteine forms a disulfide with the conserved cysteine in TRAF3 (Zhang et al., 2012). Therefore, these results show that TRAFs 2, 3, and 5 favor cysteines across a distribution of at least three peptide positions N-terminal to the motif proline, possibly due to their ability to form disulfides with a conserved cysteine on the TRAF surface.

The cysteine on the TRAF surface located in proximity to the peptide region C-terminal to the core motif could potentially form disulfides with cysteines located at a distribution of positions along the peptide. Cysteines were enriched at positions C-terminal to the core in the TANK library by all three TRAFs. The frequency of mutation to cysteine or a stop codon peaked at the second aspartate in TANK, which is highlighted in orange in Figure 4.10a. This aspartate position is a cysteine in wild-type CD40, which would likely be able to reach the conserved C-terminal cysteine on the TRAF surface. In the structure, this region of the TANK peptide is arched away from the surface of TRAF3 and does not make extensive contacts. As with the results at the N-terminus, the distribution of preferences for cysteine C-terminal to the core motif suggests a flexibility of the peptide termini.

I performed the TRAF and *E. coli* incubations in the presence of 2 mM DTT, but subsequent wash steps did not contain a reducing agent, so it is likely that disulfides were able to form and were favored because they would greatly reduce the off-rate. It is unclear if these cysteines in TRAFs and their binding partners would form disulfides in the reducing environment of the cytoplasm. Disulfide formation could be a mechanism for sustained

signaling. Whether or not disulfide formation is physiologically relevant, it offers a strategy for the future design of covalent inhibitors of TRAFs. For future libraries, the enrichment of cysteines could be avoided by mutating the TRAF cysteines or by using reducing agent in all screening steps.

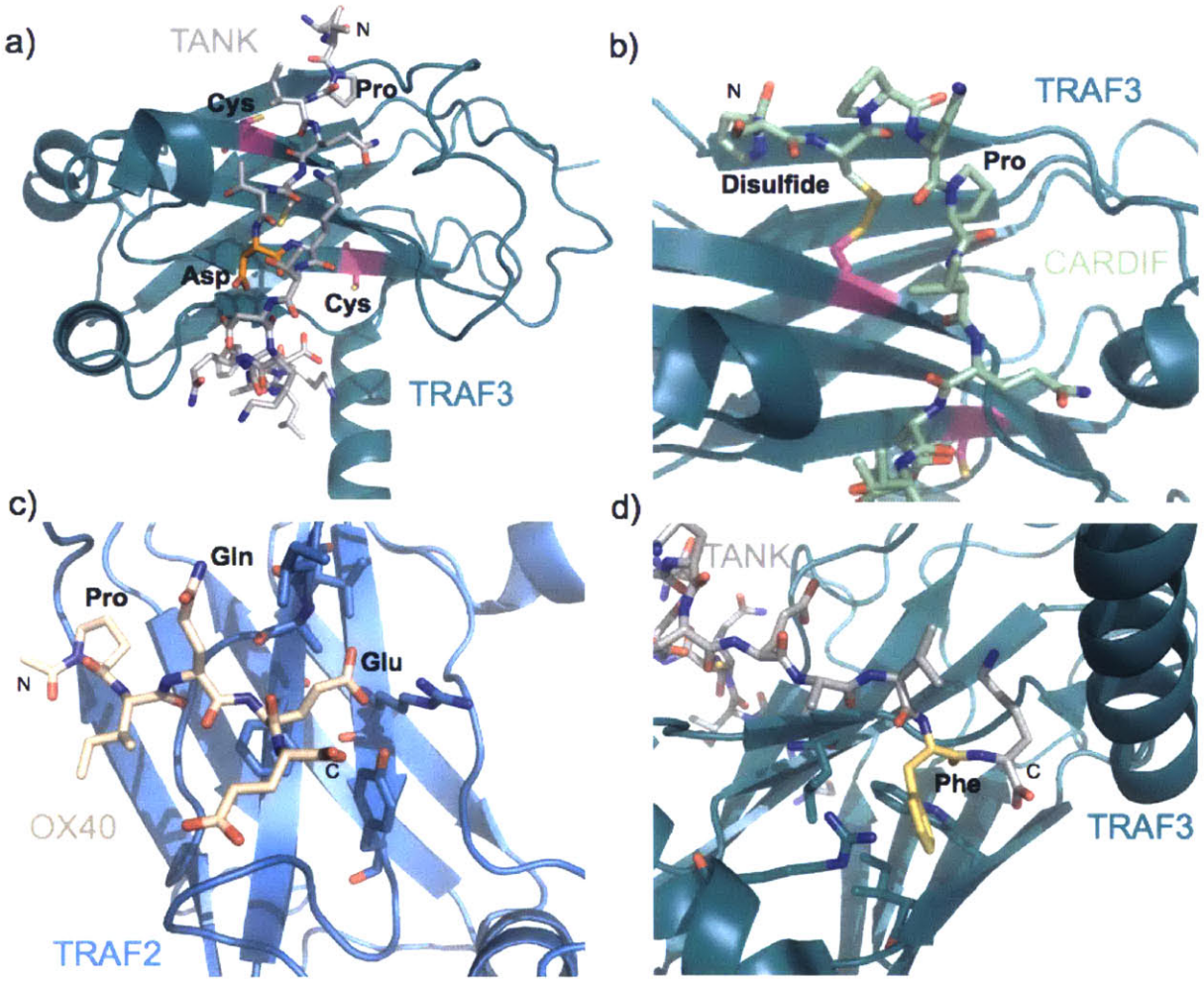


Figure 4.10. Structural environments of peptide positions of interest. a) Structure of TRAF3 (teal) bound to a peptide from TANK (gray) with the two cysteines of TRAF3 highlighted in magenta (Li et al., 2002). The aspartate in TANK that is at the equivalent position to the cysteine in wild-type CD40 is highlighted in orange. The core proline residue is labeled for reference in parts a-c. b) TRAF3 bound to a peptide from CARDIF showing a disulfide formed between TRAF3 and a cysteine three residues N-terminal to the core proline (Zhang et al., 2012). c) A peptide from OX40 bound to TRAF2 shows the conserved core glutamate with its pocket formed by hydrophobic residues and an arginine from TRAF2 shown as sticks. d) The pocket for the phenylalanine (yellow) near the C-terminus of the TANK peptide is formed by hydrophobic residues and an arginine from TRAF3, shown as sticks.

The TRAFs demonstrated a preference for hydrophobic residues and stop codons in the region C-terminal to the peptide core in TANK sequences. TRAF3 also showed a preference for hydrophobic residues at the core glutamate position in CD40 sequences. The interaction of the core glutamate with the surface of TRAF2 is shown in Figure 4.10c. All TRAFs have a conserved arginine positioned to interact with the negative charge from the glutamate. In other structures (e.g. TRAF3:TANK and TRAF2:LMP1), this arginine interacts with negatively charged side chains located at peptide positions of up to eight residues C-terminal to the core glutamate (Li et al., 2002; Ye et al., 1999). The preference for stop codons in the library peptides could arise from favorable electrostatic interactions between the peptide C-terminus and this arginine. However, if that was the case, one might also expect to see an enrichment of aspartates and glutamates, which was not observed. Therefore, the enrichment of stop codons may have occurred for another reason. One possibility is that the C-terminal region of the peptide could form inhibitory interactions with other parts of the peptide, or other constituents of the *E. coli* outer membrane, which could reduce TRAF access to the core motif.

The preference for hydrophobic residues at the core glutamate and positions C-terminal to the core may represent an alternate strategy for binding. Besides the conserved TRAF arginine, the environment around the core glutamate in Figure 4.10c shows a hydrophobic pocket formed by alanine, isoleucine, valine, phenylalanine, and tyrosine side chains on the TRAF surface. All of these residues are conserved in TRAFs 2, 3, and 5. Peptide positions C-terminal to the core motif may also be able to access this pocket, or perhaps access other hydrophobic pockets on the TRAF surface. Several TRAF binding partners have residues with large hydrophobic side chains located at positions between one and six residues C-terminal to the motif threonine including CD40, CD30, TNFR2, LMP1, RANK, BAFFR, and LTβR.

A final positional preference of note is the preference for isoleucine at the aspartate position in CD40 shown in the single mutant functional score heatmaps of TRAF2 and TRAF5 (Figure 4.7). This aspartate is at an equivalent position to the phenylalanine in TANK. The binding environment of this phenylalanine in the TRAF3:TANK structure is shown in Figure 4.10d. It occupies a pocket formed by hydrophobic residues and an arginine on the underside of TRAF3, near the coiled-coil stalk. The physicochemical properties of this pocket are conserved in TRAFs 2 and 5, though some of the residue identities change. It is possible that CD40 could bind in a similar orientation as TANK and the isoleucine enriched by TRAF2 and TRAF5 could bind this pocket. This is one of many hypotheses generated by the library data that need to be explored by structural modeling and mutational analysis.

Specificity features in the CD40 library results

After observing paralog-specific binding behavior to the enriched CD40 library pools (Figure 4.6), I performed a final round of sorting under conditions meant to further enrich specific binders. I did negative sorts to collect clones that did not show binding to 10 μM TRAF3 in the final CD40 10-μM TRAF2 and TRAF5 pools. I also did negative sorts on the final CD40 10-μM TRAF3 pool to collect clones that did not show binding to 10 μM TRAF2 or 30 μM TRAF5. As it is not possible to calculate functional scores from sorts performed under changing conditions, I analyzed the enrichment ratios of the sequences in the resulting specificity pools relative to the naïve libraries. To identify residues that were further enriched by the negative sorts, I compared enrichment ratios in the specificity pool to enrichment ratios in the last affinity pool by subtracting the affinity pool enrichment ratios from the specificity pool ratios. Sequence logos were built from sequences with the top 100 enrichment ratio differences (Figure 4.11).

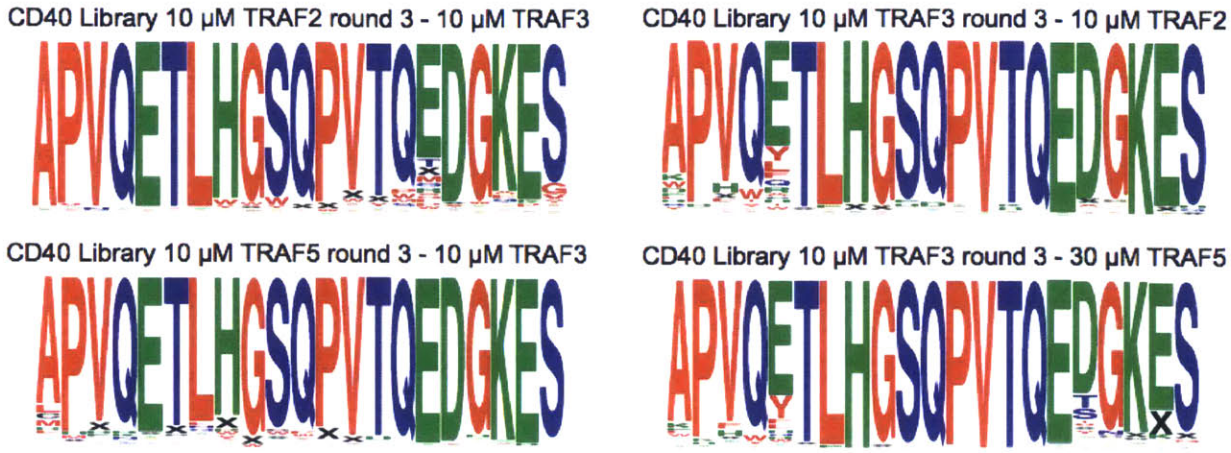


Figure 4.11. Sequence logos of enriched sequences in the specificity pools from the CD40 library. Logos with residue heights proportional to frequency were built from 100 sequences with the highest increase in enrichment ratio between the specificity pool and its source affinity pool.

The specificity logos contain residue preferences that provide hypotheses about possible specificity mechanisms between TRAFs 2/5 and TRAF3. The TRAF2-specific pool shows an increase in tryptophans at wild-type positions 'HGS', and enrichment for stop codons after these positions. The TRAF5-specific pool primarily shows a preference for stop codons, but also contains hydrophobic residues or cysteine at the first position. The TRAF3 logos emphasize the previously noted preference for hydrophobic residues at the core glutamate position. Another TRAF3-specific feature of note is the preference for basic residues at the first position, which show enrichment in both specificity pools relative to the final TRAF3 pool shown in Figure 4.8. The tryptophan at the highly conserved core glutamine position is also an unexpected feature. Sequences with threonine or serine in place of the aspartate were also enriched in the pool of TRAF3 sequences showing specificity against TRAF5. These data demonstrate that, despite the high sequence identity between TRAFs 2, 3 and 5, these paralogs do have distinct binding preferences. These library features will need to be tested in solution binding assays to determine the degree of specificity they impart.

Conclusions

Despite their essential role in many signaling pathways, the peptide binding preferences of TRAFs have been relatively uncharacterized to date. The prevailing model was that the high sequence identity in the core binding groove resulted in nearly identical binding preferences for TRAFs 1, 2, 3, and 5. My results demonstrate that there are difference between these close homologs, especially between TRAF3 and TRAFs 2 and 5. Given that the observed specificity was dependent on the peptide backbone, it is possible that other peptides could discriminate between TRAFs 2 and 5. Single and double mutant libraries only begin to sample the sequence space present in natural binding partners. Higher-order combinatorial libraries may be needed to access alternate binding modes that could impart specificity.

The library results emphasized that the TRAFs do have distinct binding preferences outside of the core motif. This was not apparent in previously published SPOT arrays, which showed little variation in binding to single mutants of the CD40 C-terminal region (Pullen et al., 1999a). An interesting feature of the most strongly enriched codons was that they were present in distributions across several peptide positions. Future models of TRAF binding preferences will need to represent these distributions.

The library datasets are rich with hypotheses that will be tested by structural modeling and mutagenesis of the TRAFs and peptides. They also will be useful for testing different approaches to analyzing sequencing enrichment data and its correlation with binding affinity. The data on both single and double mutants will allow us to begin to examine epistatic relationships between peptide positions. This library approach will be useful in the future for studies of the other TRAF homologs, and will reveal many new insights into this important signaling family.

Methods

SiteMAP and sequence identity analysis

SiteMAP analysis of the TRAF MATH domains and comparison of the TIM binding grooves was performed similarly as in Chapter 3 for the Bcl-2 homologs. Given the relatively high sequence identity and similar structures of TRAF MATH domains, alignment was successfully performed with the 'align' command in PyMol. All structures were aligned to the structure of TRAF3 bound to a long peptide from TANK (PDB ID 1L0A) (Li et al., 2002). For structures that included multiple MATH domain copies, one monomer with a peptide bound was chosen. All existing structures of TRAF MATH domains were analyzed: TRAF3:TANK (1L0A) (Li et al., 2002), TRAF3 apo (1FLK) (Ni et al., 2000), TRAF3:CD40 (1FLL) (Ni et al., 2000), TRAF3:TANK (1KZZ) (Li et al., 2002), TRAF3:LTβR (1RF3) (Li et al., 2003), TRAF3:LMP1 (1ZMS) (Wu et al., 2005), TRAF3:BAFFR (2GKW) (Ni et al., 2004), TRAF3:CARDIF (4GHU) (Zhang et al., 2012), TRAF2 apo (1CA4) (Park et al., 1999), TRAF2:TNFR2 (1CA9) (Park et al., 1999), TRAF2:LMP1 (1CZY) (Ye et al., 1999), TRAF2:CD40 (1CZZ) (Ye et al., 1999), TRAF2:CD40 (1D00) (Ye et al., 1999), TRAF2:OX40 (1D0A) (Ye et al., 1999), TRAF2:m4-1BB (1D0J) (Ye et al., 1999), TRAF2:CD30 (1D01) (Ye et al., 1999), TRAF2:CD40 mutant (1QSC) (McWhirter et al., 1999), TRAF6 apo (1LB4) (Ye et al., 2002), TRAF6:RANK (1LB5) (Ye et al., 2002), TRAF6:CD40 (1LB6) (Ye et al., 2002), TRAF4 apo (3ZJB, 4K8U, 4M4E) (Niu et al., 2013; Rousseau et al., 2013; Yoon et al., 2014), TRAF5 apo (4GHJ) (Zhang et al., 2012).

Briefly, aligned complex structures were prepared for SiteMAP analysis with Maestro (version 9.7) Prepwizard. SiteMAP (version 3.0, analysis, Schrödinger, LLC) was run on the prepared structures with the peptides removed using a 3 Å sitebox around the 1L0A peptide to

search for sites. Coordinates and potentials of sitemaps for individual sites on the surface of each TRAF were concatenated to create one inclusive sitemap for each structure. The binding site environments were compared by computing an intersection score between sitemaps for different structures as previously described (Chapter 3). The Cβ coordinates used as binding site locators were from TANK in 1L0A: VPIQCTD for the core binding site, and EALF for the TRAF3 exosite.

Sequence identity was calculated from alignments built using Clustal Omega (Sievers et al., 2011). Hierarchical clustering of TRAFs by sequence identity or SiteMAP similarity score was performed with the clustergram function in Matlab (version R2012b), based on correlation.

Protein constructs and purification

The TRAF constructs used contained only the MATH domain, which was inserted into the pDW363 biotinylation vector after the biotin acceptor peptide-tag, a His_6-tag and a linker (N-terminal tag sequence: MAGGLNDIFEAQKIEWHEDTGGSSHHHHHHGSGSGS). The TRAF sequences are shown in Table 4.1. TRAF proteins were expressed in BL21 (DE3) pLysS Rosetta cells. 10 mL overnights were started from a fresh colony and grown at 37 °C with shaking, and 5 mL of the overnight culture was used to inoculate 1 L of LB including 100 μg/mL ampicillin, 25 μg/mL chloramphenicol, and 12-15 mg of D-(+)-biotin. Cells were grown at 37 °C with shaking to an optical density at 600 nm (O.D. 600) of 0.5-0.6 and then moved to 18 °C and induced with 0.5 mM IPTG overnight before harvesting.

Purification of TRAFs was performed as follows. 1 L cell pellets were resuspended in 25 mL of 5 mM imidazole, 500 mM NaCl, 1 mM DTT, 20 mM Tris pH 8.0, 0.1% v/v Tween-20, and 0.2 mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor. Cells were sonicated ten times for 30 seconds followed by 30 seconds of rest. The supernatant from the centrifuged lysis

product was filtered through 0.2 μm filters before application to 3 mL of Ni-nitrilotriacetic acid agarose resin equilibrated in 20 mM Tris pH 8.0, 500 mM NaCl. After the supernatant was applied, the resin was washed 3 times with 8 mL of 20 mM imidazole, 500 mM NaCl, 20 mM Tris pH 8.0. The protein was eluted with 8 mL 20 mM Tris pH 8.0, 500 mM NaCl, 300 mM imidazole. The protein was applied to a S75 26/60 size exclusion column equilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM DTT. Purity was verified by SDS-PAGE, and proteins were frozen at -80 °C in the final buffer at a concentration of 100-700 μM.

Bacterial surface display constructs and library assembly

CD40 and TANK peptides were expressed on the C-terminus of eCPX. The constructs are shown in Table 4.1. A cysteine in each peptide was mutated to serine to reduce background binding of TRAFs. A threonine to alanine mutation (PVQEA or PIQCA) was used as a negative control in library sorting experiments, as this mutation is reported to reduce TRAF binding (Pullen et al., 1999a). The single and double point mutant libraries were originally constructed in an eCPX construct that contained a FLAG-tag C-terminal to the peptide. Sorting these libraries led to stop codons after the core binding sequence, which removed the FLAG-tag. Therefore, the FLAG-tag was removed by PCR on the assembled library plasmid DNA, leaving CD40 at the C-terminus and a 'GGSGGS' cloning artifact C-terminal to TANK.

The libraries were assembled in 3 pieces, with each 7-mer mutated segment covered by a different primer. All single and double point mutants in 3, 7-mers were encoded by NNK (N=A/C/G/T, K=G/T) in oligonucleotides made by machine mixing of nucleotides (Integrated DNA Technologies). The sections mutated in the CD40 library were: APVQETL, HGSQPVT, and QEDGKES. The sections mutated in the TANK library were: VPIQCTD, TDKTDKQ, KQEALFK. Underlined residues in the middle TANK section overlap with the first and third

section; the TANK library contained all double mutants within each of these 7-mers, but single mutants were not duplicated for the underlined residues. The libraries were assembled analogously to those in Chapter 3, with the exception that the piece of insert DNA included the peptide and ~150 bp of vector backbone 3' to the peptide. The 3' SfiI site was introduced further into the vector to allow for ~200 bp between the SfiI site immediately 5' to the peptide, which SfiI needs for efficient cleavage (according to New England Biolabs). The SfiI site located between the signal peptide and the N-terminus of eCPX in the original eCPX construct was removed. See Table 4.1 for vector sequences. Empty vectors (no peptide) were digested and treated with alkaline phosphatase to serve as the vector for library insertion (see Chapter 3). Inserts were assembled by overlap PCR using standard conditions for Phusion HiFi polymerase with the library primer on a 3' segment corresponding to the DNA from immediately 3' of the last codon in each 7-mer segment through the 3' SfiI site. Each single and double mutant insert PCR was done individually, and PCR products were pooled at equimolar ratios before insert digestion with SfiI. For the second and third 7-mer segment of each library, an additional primer was used to extend the insert DNA to the 5' SfiI site. All PCRs were performed with Phusion high-fidelity polymerase (New England Biolabs). Each 7-mer library was electroporated into MC1061 *E. coli* individually and grown for glycerol stocks as in Chapter 3.

To create the libraries without a FLAG tag, plasmid DNA from the original libraries (assembled as described above) was miniprepped from overnight cultures started from glycerol stocks. The FLAG tag was removed from the CD40 libraries (parts 1, 2, and 3) by amplification of the plasmid with primers that excluded the FLAG-tag and linker after the CD40 peptide, but had a 40 bp overlap covering the 3' end of the peptide sequence and the vector after the stop codon (primers CL_GA_fwd and CL_GA_rev). The PCR was performed with Phusion HiFi

polymerase (New England Biolabs) according to standard procedures, with an annealing temperature of 64 °C and 30 cycles. Three reactions were performed per library part, with 50 ng of original library template used in each reaction. The PCR products were digested with DpnI at 37 °C for 1 hr and then run on a 0.65% agarose gel with Gel Green dye. Vector-sized bands were extracted with Zymo Gel Extraction Kit and each library part was purified over two columns and eluted with 25 μL water. Gibson Assembly was performed to connect the overlap and ligate the plasmid. Three, 20 μL Gibson Assembly reactions were performed per library part, with ~150 ng library DNA in each reaction.

The FLAG-tag was removed from the TANK library by round-the-horn PCR and blunt end ligation. Part of the linker was retained for PCR due to the fact that the mutated region of the TANK peptide went to the end of the peptide. FLAG_cut_fwd and TL_rth_rev primers (below) were phosphorylated with PNK (New England Biolabs) using T4 ligase buffer. Phosphorylated primers were used for PCR with Phusion HiFi polymerase, with an annealing temperature of 69 °C and 30 cycles. 50 ng of TANK library template was used in each reaction, and two reactions were performed per library part. PCR products were digested and purified as for the CD40 library. Ligations were performed overnight at 4 °C with T4 DNA ligase. Two, 20 μL reactions were performed for each library part, with ~350 ng DNA/reaction.

DNA for the libraries without the FLAG-tag was desalted on sterile MilliQ water on a 0.025 μm Millipore filter for 20 minutes. The DNA for each library part was split equally between two, 250 μL MC1061 *E. coli* competent cell stocks. Each cell stock plus DNA was electroporated in a 2 mm cuvette at 2.5 kV, 100 Ω, 50 μF and recovered in 10 mL warm SOC for an hour at 37 °C with rotation. Each library part was then added to 150 mL LB with 0.2 % w/v

glucose and 25 μg/mL chloramphenicol and grown at 37 °C for 8-20 hours (to an O.D.600 of 1.0-1.8). Cells were then pelleted and resuspended in 20% glycerol and frozen at -80 °C.

Table 4.1. TRAF and peptide-eCPX constructs

Construct name	Sequence
TRAF2 MATH	DGVFIWKISDFARKRQEAVAGRIPAIFSPAFYTSRYGYKMC LRIYLNGDGTGRGTHLSLFFVVMKGPNDALLRWPFNQKVT LMLLDQNNREHVIDAFRPDVTSSSFQRPVNDMNIASGCPLF CPVSKMEAKNSYVRDDAIFIKAIVDLTGL
TRAF3 MATH	NGVLIWKIRDYKRRKQEAVMGKTLSLYSQPFYTGYFGYK MCARVYLNGDGMGKGTHLSLFFVIMRGEYDALLPWPFKQ KVTLMLMDQGSSRRHLGDAFKPDPNSSSFKKPTGEMNIAS GCPVFVAQTVLENGTYIKDDTIFIKVIVDTSDLPDP
TRAF5 MATH	NGKLIWKVTDYKMKKREAVDGHTVSIFSQSFYTSRCGYRL CARAYLNGDGSGRGSHLSLYFVVMRGEFDSLLQWPFRQR VTLMLLDQSGKKNIMETFKPDPNSSSFKRPDGEMNIASGCP RFVAHSVLENAKNAYIKDDTLFLKVAVDLTDLEDL
Constructs used for library screening:	
eCPX-CD40[a]	MKKIACLSALAAVLAFTAGTSVAGGQSGQSGDYNKNQYY GITAGPAYRINDWASIYGVVGVGYGKFQTTEYPTYKHDTS DYGFSYGAGLQFNPMENVALDFSYEQSRIRSVDVGTWILS VGYRFGSKSRRATSTVTGGYAQSDAQGQMNKMGGFNLK YRYEEDNSPLGVIGSFTYTEKSRTASGSGGQSGR**NTAAPVQ** **ETLHGSQPVTQEDGKESRISVQERQ**
eCPX-CD40 vector	GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT CGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC TCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCA ACTATGGATGAACGAAATAGACAGATCGCTGAGATAGG TGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT TTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAG CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTC CTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGG TTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA CTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCC TGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT TCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACT CAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATT TTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT

	TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACG TTTACAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGA TAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGG TAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCG GCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCG AATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATC ACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAA TAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCA TCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACAT GGAAGCCATCACAGACGGCATGATGAACCTGAATCGCC AGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTG CCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATT GGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATC TTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTG GTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTC ATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA TCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCG GATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAG GCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTA AAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT TACGATGCCATTGGGATATATCAACGGTGGTATATCCAG TGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAA TCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTAT TTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATC AACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCC GGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAA GTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGC GTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGAT GGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAG CTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAA CGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTG TATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAG TGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACC GGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGC TTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG

	CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTT CCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGA GAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCC CCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTG GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT TTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGC CTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGT TTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGC AGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT GAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGC AGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGA AGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGT TTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCA AAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCT GCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGC GCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATC AGATAAAATATTTGCTCATGAGCCCGAAGTGGCGAGCCC GATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAG CAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGC GTCCGGCGTAGAGGATCTGCTCATGTTTGACAGCTTATC ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGG GATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAAT TGTCTGATTCGTTACCAATTATGACAACTTGACGGCTAC ATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTC GGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAA TAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTG GCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCC TGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTA ATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGA CGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATAT CAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGAC AAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCG ACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCT CAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTT CCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGC TGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACC CCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCAT GCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGA TACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGA ATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGG CAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGAC CGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCA GCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCT CAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACG AGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCA

	TACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATT GTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTA CTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGAC AAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAA AAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTA CCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATA CCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTACCTTT GAGGTGGTTATGAAAAAAATTGCATGTCTTTCAGCACTG GCCGCAGTTCTGGCTTTCACCGCAGGTACTTCCGTAGCT GGAGGGCAGTCTGGGCAGTCTGGTGACTACAACAAAAA CCAGTACTACGGCATCACTGCTGGTCCGGCTTACCGCAT TAACGACTGGGCAAGCATCTACGGTGTAGTGGGTGTGGG TTATGGTAAATTCCAGACCACTGAATACCCGACCTACAA ACACGACACCAGCGACTACGGTTTCTCCTACGGTGCGGG TCTGCAGTTCAACCCGATGGAAAACGTTGCTCTGGACTT CTCTTACGAGCAGAGCCGTATTCGTAGCGTTGACGTAGG CACCTGGATTTTGTCTGTTGGTTACCGCTTCGGGAGTAA ATCGCGTCGCGCGACTTCTACTGTAACTGGCGGTTACGC ACAGAGCGACGCTCAGGGCCAAATGAACAAAATGGGCG GTTTCAACCTGAAATACCGCTATGAAGAAGACAACAGCC CGCTGGGTGTGATCGGTTCTTTCACTTACACCGAGAAAA GCCGTACTGCAAGCGGTAGCGGCGGCCAGTCTGGCCGTA ATACTGCTGCACCAGTTCAAGAAACTTTACATGGTAGCC AACCAGTTACTCAAGAAGATGGTAAAGAATCTAGAATTT CTGTTCAGGAAAGACAATAATAAGGCCAAGCTTGGCTGT TTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATT AAATCAGAACGGGCCAAGGTGGCCCAGAAGCGGTCTGA TAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCC CACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTA GCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAG GGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTC GAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTG AACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGA TTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGG CAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAG CAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTA CAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATG TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAA TAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTC CGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCC TGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGA TGCTGAAGATCAGTTGGGTGCA
eCPX-TANK[a]	MKKIACLSALAAVLAFTAGTSVAGGQSGQSGDYNKNQYY GITAGPAYRINDWASIYGVVGVGYGKFQTTEYPTYKHDTS

	DYGFSYGAGLQFNPMENVALDFSYEQSRIRSVDVGTWILS VGYRFGSKSRRATSTVTGGYAQSDAQGQMNKMGGFNLK YRYEEDNSPLGVIGSFTYTEKSRTASGSGGQSGR**QSSVPIQ** **CTDKTDKQEALFK**GGSGGS
eCPX-TANK vector	GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT CGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC TCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCA ACTATGGATGAACGAAATAGACAGATCGCTGAGATAGG TGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT TTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAG CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTC CTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGG TTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA CTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCC TGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT TCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACT CAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATT TTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACG TTTACAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGA TAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGG TAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCG GCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCG AATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATC ACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAA TAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCA TCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACAT GGAAGCCATCACAGACGGCATGATGAACCTGAATCGCC AGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTG CCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATT GGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATC

	TTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTG GTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTC ATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA TCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCG GATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAG GCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTA AAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT TACGATGCCATTGGGATATATCAACGGTGGTATATCCAG TGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAA TCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTAT TTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATC AACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCC GGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAA GTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGC GTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGAT GGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAG CTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAA CGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTG TATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAG TGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACC GGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGC TTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTT CCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGA GAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCC CCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTG GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT TTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGC CTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGT TTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGC AGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT GAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGC AGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGA AGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGT TTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCA AAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCT GCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGC GCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATC AGATAAAATATTTGCTCATGAGCCCGAAGTGGCGAGCCC GATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAG CAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGC GTCCGGCGTAGAGGATCTGCTCATGTTTGACAGCTTATC ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGG GATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAAT

	TGTCTGATTCGTTACCAATTATGACAACTTGACGGCTAC ATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTC GGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAA TAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTG GCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCC TGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTA ATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGA CGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATAT CAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGAC AAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCG ACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCT CAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTT CCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGC TGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACC CCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCAT GCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGA TACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGA ATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGG CAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGAC CGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCA GCGGTCGGTCGATAAAAAAAATCGAGATAACCGTTGGCCT CAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACG AGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCA TACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATT GTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTA CTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGAC AAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAA AAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTA CCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATA CCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTACCTTT GAGGTGGTTATGAAAAAAATTGCATGTCTTTCAGCACTG GCCGCAGTTCTGGCTTTCACCGCAGGTACTTCCGTAGCT GGAGGGCAGTCTGGGCAGTCTGGTGACTACAACAAAAA CCAGTACTACGGCATCACTGCTGGTCCGGCTTACCGCAT TAACGACTGGGCAAGCATCTACGGTGTAGTGGGTGTGGG TTATGGTAAATTCCAGACCACTGAATACCCGACCTACAA ACACGACACCAGCGACTACGGTTTCTCCTACGGTGCGGG TCTGCAGTTCAACCCGATGGAAAACGTTGCTCTGGACTT CTCTTACGAGCAGAGCCGTATTCGTAGCGTTGACGTAGG CACCTGGATTTTGTCTGTTGGTTACCGCTTCGGGAGTAA ATCGCGTCGCGCGACTTCTACTGTAACTGGCGGTTACGC ACAGAGCGACGCTCAGGGCCAAATGAACAAAATGGGCG GTTTCAACCTGAAATACCGCTATGAAGAAGACAACAGCC CGCTGGGTGTGATCGGTTCTTTCACTTACACCGAGAAAA

	GCCGTACTGCAAGCGGTAGCGGCGGCCAGTCTGGCCGTC AGAGCTCTGTTCCAATTCAATGTACTGATAAAACAGACA AGCAAGAAGCTTTATTTAAAGGTGGCTCTGGCGGTAGCT AATAAGGCCAAGCTTGGCTGTTTTGGCGGATGAGAGAA GATTTTCAGCCTGATACAGATTAAATCAGAACGGGCCAA GGTGGCCCAGAAGCGGTCTGATAAAACAGAATTTGCCTG GCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGA ACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTG GGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATC AAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTT CGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTA GGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAG CAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATA AACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGA CGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTTAT TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGAC AATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA AGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCC CTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTG GGTGCA
Original eCPX constructs with FLAG tag:	
eCPX-CD40-FLAG[a]	MKKIACLSALAAVLAFTAGTSVAGGQSGQSGDYNKNQYY GITAGPAYRINDWASIYGVVGVGYGKFQTTEYPTYKHDTS DYGFSYGAGLQFNPMENVALDFSYEQSRIRSVDVGTWILS VGYRFGSKSRRATSTVTGGYAQSDAQGQMNKMGGFNLK YRYEEDNSPLGVIGSFTYTEKSRTASGSGGQSGR**NTAAPVQ** **ETLHGSQPVTQEDGKESRISVQERQ**GGSGGSGGGQSGQ**D** **YKDDDDK**
eCPX-CD40-FLAG vector	GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT CGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC TCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCA ACTATGGATGAACGAAATAGACAGATCGCTGAGATAGG TGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT TTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAG CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTC CTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGG TTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA CTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCC TGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT

	TCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACT CAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATT TTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACG TTTACAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGA TAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGG TAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCG GCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCG AATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATC ACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAA TAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCA TCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACAT GGAAGCCATCACAGACGGCATGATGAACCTGAATCGCC AGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTG CCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATT GGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATC TTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTG GTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTC ATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA TCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCG GATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAG GCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTA AAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT TACGATGCCATTGGGATATATCAACGGTGGTATATCCAG TGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAA TCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTAT TTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATC AACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCC GGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAA GTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGC GTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGAT GGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAG CTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAA CGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTG TATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAG

	TGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACC GGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGC TTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTT CCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGA GAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCC CCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTG GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT TTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGC CTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGT TTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGC AGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT GAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGC AGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGA AGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGT TTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCA AAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCT GCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGC GCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATC AGATAAAATATTTGCTCATGAGCCCGAAGTGGCGAGCCC GATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAG CAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGC GTCCGGCGTAGAGGATCTGCTCATGTTTGACAGCTTATC ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGG GATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAAT TGTCTGATTCGTTACCAATTATGACAACTTGACGGCTAC ATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTC GGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAA TAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTG GCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCC TGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTA ATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGA CGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATAT CAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGAC AAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCG ACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCT CAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTT CCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGC TGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACC CCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCAT GCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGA TACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGA ATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGG CAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGAC CGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCA

	GCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCT CAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACG AGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCA TACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATT GTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTA CTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGAC AAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAA AAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTA CCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATA CCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTACCTTT GAGGTGGTTATGAAAAAAATTGCATGTCTTTCAGCACTG GCCGCAGTTCTGGCTTTCACCGCAGGTACTTCCGTAGCT GGAGGGCAGTCTGGGCAGTCTGGTGACTACAACAAAAA CCAGTACTACGGCATCACTGCTGGTCCGGCTTACCGCAT TAACGACTGGGCAAGCATCTACGGTGTAGTGGGTGTGGG TTATGGTAAATTCCAGACCACTGAATACCCGACCTACAA ACACGACACCAGCGACTACGGTTTCTCCTACGGTGCGGG TCTGCAGTTCAACCCGATGGAAAACGTTGCTCTGGACTT CTCTTACGAGCAGAGCCGTATTCGTAGCGTTGACGTAGG CACCTGGATTTTGTCTGTTGGTTACCGCTTCGGGAGTAA ATCGCGTCGCGCGACTTCTACTGTAACTGGCGGTTACGC ACAGAGCGACGCTCAGGGCCAAATGAACAAAATGGGCG GTTTCAACCTGAAATACCGCTATGAAGAAGACAACAGCC CGCTGGGTGTGATCGGTTCTTTCACTTACACCGAGAAAA GCCGTACTGCAAGCGGTAGCGGCGGCCAGTCTGGCCGTA ATACTGCTGCACCAGTTCAAGAAACTTTACATGGTAGCC AACCAGTTACTCAAGAAGATGGTAAAGAATCTAGAATTT CTGTTCAGGAAAGACAAGGTGGCTCTGGCGGTAGCGGT GGAGGTCAGTCCGGCCAGGACTACAAAGACGACGATGA CAAATAATAAGGCCAAGCTTGGCTGTTTTGGCGGATGAG AGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGGG CCAAGGTGGCCCAGAAGCGGTCTGATAAAACAGAATTT GCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCAT GCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTA GTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAG GCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGG CCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCT GAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTG CGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCG CCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCAT CCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTG TTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATG AGACAATAACCCTGATAAATGCTTCAATAATATTGAAAA AGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTT

	ATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCA CCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC AGTTGGGTGCA
eCPX-TANK-FLAG[a]	MKKIACLSALAAVLAFTAGTSVAGGQSGQSGDYNKNQYY GITAGPAYRINDWASIYGVVGVGYGKFQTTEYPTYKHDTS DYGFSYGAGLQFNPMENVALDFSYEQSRIRSVDVGTWILS VGYRFGSKSRRATSTVTGGYAQSDAQGQMNKMGGFNLK YRYEEDNSPLGVIGSFTYTEKSRTASGSGGQSGR**QSSVPIQ** **CTDKTDKQEALFK**GGSGGSGGGQSGQ**DYKDDDDK**
eCPX-TANK-FLAG vector	GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT CGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC TCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCA ACTATGGATGAACGAAATAGACAGATCGCTGAGATAGG TGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT TTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAG CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTC CTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGG TTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA CTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCC TGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT TCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACT CAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATT TTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACG TTTACAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGA TAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTA GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA GTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGG TAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCG GCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCG AATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATC ACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAA TAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCA TCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACAT GGAAGCCATCACAGACGGCATGATGAACCTGAATCGCC

	AGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTG CCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATT GGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATC TTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTG GTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTC ATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA TCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCG GATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAG GCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTA AAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT TACGATGCCATTGGGATATATCAACGGTGGTATATCCAG TGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAA TCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTAT TTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATC AACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCC GGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAA GTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGC GTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGAT GGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAG CTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAA CGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTG TATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAG TGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACC GGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGC TTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTT CCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGA GAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCC CCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTG GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT TTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGC CTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGT TTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGC AGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT GAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGC AGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGA AGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGT TTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCA AAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCT GCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGC GCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATC AGATAAAATATTTGCTCATGAGCCCGAAGTGGCGAGCCC

	GATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAG CAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGC GTCCGGCGTAGAGGATCTGCTCATGTTTGACAGCTTATC ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGG GATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAAT TGTCTGATTCGTTACCAATTATGACAACTTGACGGCTAC ATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTC GGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAA TAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTG GCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCC TGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTA ATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGA CGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATAT CAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGAC AAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCG ACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCT CAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTT CCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGC TGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACC CCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCAT GCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGA TACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGA ATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGG CAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGAC CGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCA GCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCT CAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACG AGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCA TACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATT GTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTA CTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGAC AAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAA AAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTA CCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATA CCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTACCTTT GAGGTGGTTATGAAAAAAATTGCATGTCTTTCAGCACTG GCCGCAGTTCTGGCTTTCACCGCAGGTACTTCCGTAGCT GGAGGGCAGTCTGGGCAGTCTGGTGACTACAACAAAAA CCAGTACTACGGCATCACTGCTGGTCCGGCTTACCGCAT TAACGACTGGGCAAGCATCTACGGTGTAGTGGGTGTGGG TTATGGTAAATTCCAGACCACTGAATACCCGACCTACAA ACACGACACCAGCGACTACGGTTTCTCCTACGGTGCGGG TCTGCAGTTCAACCCGATGGAAAACGTTGCTCTGGACTT CTCTTACGAGCAGAGCCGTATTCGTAGCGTTGACGTAGG

	CACCTGGATTTTGTCTGTTGGTTACCGCTTCGGGAGTAA ATCGCGTCGCGCGACTTCTACTGTAACTGGCGGTTACGC ACAGAGCGACGCTCAGGGCCAAATGAACAAAATGGGCG GTTTCAACCTGAAATACCGCTATGAAGAAGACAACAGCC CGCTGGGTGTGATCGGTTCTTTCACTTACACCGAGAAAA GCCGTACTGCAAGCGGTAGCGGCGGCCAGTCTGGCCGTC AGAGCTCTGTTCCAATTCAATGTACTGATAAAACAGACA AGCAAGAAGCTTTATTTAAAGGTGGCTCTGGCGGTAGCG GTGGAGGTCAGTCCGGCCAGGACTACAAAGACGACGAT GACAAATAATAAGGCCAAGGTGGCCAAGCTTGGCTGTTT TGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA ATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGC CTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGC CGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGT GTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGC ATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCC TTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGA GTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCG AAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCC ATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCT GACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTT ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGA CAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG AAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTC CCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTT GGGTGCA
CL_GA_fwd	AGAATTTCTGTTCAGGAAAGACAATAATAAGGCCAAGCT TGGCTG
CL_GA_rev	CCAAGCTTGGCCTTATTATTGTCTTTCCTGAACAGAAATT CT
FLAG_cut_fwd	TAATAAGGCCAAGCTTGGCTGTTTTG
TL_rth_rev	GCTACCGCCAGAGCC ACC
Illumina preparation primers:	
Mme_fwd_TRAF	GGGACCACCACCTCCGACAGCGGTAGCGGCGGCCAGTC
3prime_rev_TANKLib	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTA TTAGCTACCGCCAGAGCCACC
3prime_rev_CD40Lib	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCAGGC CTTATTATTGTCTTTCCTGAACAGAAATTCT
FinalPCR_fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACG
FinalPCR_rev_TANKLib	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTA TTAGCTAC

FinalPCR_rev_CD40Lib	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCAGGC CTTATTA
rev_seq_TANKLib	CGGTCTCGGCATTATTAGCTACCGCCAGAGCCACC
rev_seq_CD40Lib	CGGTCTCGGCAGGCCTTATTATTGTCTTTCCTGAACAGA AATTCT

[a]Peptide sequence and FLAG-tag are in bold with the cysteine to serine mutation underlined.

Library sorting and FACS analysis

The general protocol for FACS sample preparation is as follows: Enough glycerol stock to oversample the library by at least 10-fold was used to inoculate 5 mL LB plus 0.2% glucose and 25 μg/mL chloramphenicol and grown overnight at 37 °C on a rotator wheel. Fresh 5 mL LB plus 25 μg/mL chloramphenicol cultures were started from a volume equivalent to 100 μL of an O.D.600 6.0 overnight culture. The 3, 7-mer, naïve libraries for each peptide were pooled equally according to optical density at 600 nm (O.D.600) on the first day of sorting. If the overnight culture was dilute such that a large inoculum volume was needed, the inoculum was first pelleted at 3000 relative centrifugal force (rcf) for 5 minutes and the pellet was resuspended in 100 μL LB to serve as the inoculum. Cultures were grown at 37 °C to an O.D.600 of 0.5-0.6 (~ 2hrs) and then induced to 0.04% w/v arabinose for 1.5 hours. 7 x 10^7 cells were pelleted at 3000 rcf for 5 min for library sorting samples (using the rule that a culture at an O.D.600 of 1.0 contains a concentration of *E. coli* of ~5 x 10^8 cells/mL) and resuspended in 100 μL PBS plus 0.5% w/v bovine serum albumin (BSA). 100 μL 2X TRAF protein in PBS + 4 mM DTT was added to the cells and incubated at ~23 °C for 1 hr. Samples were pelleted, washed with 200 μL PBS plus 0.1% w/v BSA (PBSA), pelleted again, and then 200 μL streptavidin-PE (Molecular Probes) at a 1:100 dilution in PBSA was added. Samples were incubated with the labeling reagent for 15 minutes on ice, and the wash step was repeated. Final samples for sorting were resuspended in 1.3 mL PBSA. FACS sorting was performed on BD FACSAria machines, and analysis was performed on BD FACSCalibur machines. Cells were collected in SOC and put on ice until all

sorting was finished. Samples were then recovered in 15-20 mL 37 °C SOC and incubated at 37 °C for ~ 30 minutes before 25 μg/mL chloramphenicol was added and then grown over night in 125 mL flasks with 250 rpm shaking. When possible, sorts were done on consecutive days, with the previous day's recovery culture serving as the inoculum for the next day's sort. This served to minimize extra growth steps and introduction of growth biases. The day after sorting, adequate volume of the recovery culture to oversample the library was pelleted for glycerol stocks and for Illumina sample prep.

Sorting was performed with relaxed gates to allow enrichment of moderate binders. The gates were set such that 0.5-1.0% of the negative control population fell in the gate. For TRAF2, TANK threonine-to-alanine mutation was used as the negative, and CD40 threonine-to-alanine was used for TRAF3 and TRAF5. Three sequential rounds of sorting on each library were performed at 10 μM TRAF2, 3, and 5. Four rounds were performed at 3 μM. Analysis of binding to the round 3 CD40 library 10 μM sort pools showed that these libraries had achieved some margin of "specificity for free" between TRAF3 vs. TRAFs 2 and 5. Therefore, I performed one round of negative sorting at the following conditions: CD40 library-10 μM TRAF2-round 3 at 10 μM TRAF3, CD40 library-10 μM TRAF5-round 3 at 10 μM TRAF3, CD40 library-10 μM TRAF 3-round 3 at 10 μM TRAF2, and CD40 library-10 μM TRAF3-round 3 at 30 μM TRAF5. Negative sort gates were set to include 100% of the events of an unlabeled control (non-binders). The CD40 library 3 μM sorts for TRAF3 and TRAF5 appeared to develop a growth bias, evidenced by fast-growing colonies and cultures; therefore, only the first and second rounds of the TRAF3 sorts and the first and fourth rounds of the TRAF5 sorts were sequenced.

Illumina sample preparation

Library pools were prepared for Illumina sampling as in Chapter 3. Briefly, cell pellets from recovery cultures were mini prepped, and 50-100 ng of DNA was used for the first PCR, which added on the Illumina 3' adapter region and the 5' MmeI site. The primers used for Illumina sample preparation and sequencing are shown in Table 4.1. Primers for the first PCR were MmeI_fwd_TRAF and 3prime_rev_TANKLib or 3prime_rev_CD40Lib. PCR products were digested with MmeI and then double stranded adapters with 5 base barcodes were ligated on as described. Twenty-four barcodes were used for multiplexing each library with the naïve TANK library pool receiving three barcodes, the naïve CD40 library pool two barcodes, and each sorted pool one barcode. Barcoded-products were amplified in a final round of PCR using primers FinalPCR_fwd and FinalPCR_rev_TANKLib or FinalPCR_rev_CD40Lib. Illumina sequencing was performed in the MIT Biomicro Center on an Illumina HiSeq2000 with each library in one lane. Paired-end reads of 85 bases were produced with the universal Illumina forward read primer and the reverse primers rev_seq_CD40Lib or rev_seq_TANKLib. A PhiX lane was run as a control for accurate base calling.

Sequencing data processing

Illumina sequencing yielded 2.48×10^8 reads for the TANK library and 1.80×10^8 reads for the CD40 library. Reads were split into their sorting pool by using an in-house Python script written by Vincent Xue, which required an exact match to the barcode and an average Phred score of 20 for the barcode. Paired-end reads were fused using Enrich 2.0 (Fowler et al., 2011). For the TANK library, all peptide positions were covered by both the forward and reverse reads. For the CD40 library, the last base of the last variable peptide position was only included in the reverse read. Enrich required an average Phred score of 20, no 'N' bases, and assigned the higher scoring base where the forward and reverse reads disagreed.

I calculated functional scores similarly to Starita et al. (Starita et al., 2015). First, I narrowed down the DNA and protein sequence files to only those sequences coding for protein sequences included in the theoretical libraries (which would include non-theoretical DNA sequences). I then calculated the frequencies of each sequence in these theoretical sequence-only pools. Enrich was also used to calculate enrichment ratios ($\log_2 (F_{selected}/F_{input})$) for each protein and DNA sequence. All enrichment ratios used the frequency of the sequence in the naïve library as F_{input}, and $F_{selected}$ was the frequency of the sequence in the selected pool. I extracted enrichment ratios for each sequence from sequential rounds (e.g. naïve library, CD40 Library 10-μM TRAF2 round 1, CD40 Library 10-μM TRAF2 round 2, CD40 Library 10-μM TRAF2 round 3). I then use the Scipy linear regression method to fit lines to the enrichment ratios across rounds and output the slope (S_{var}) for each line. To control for non-specific carryover, I calculated a correction factor (S_{stop}) based on stop codon sequences. The slopes for all sequences with stop codons in the first four variable positions were averaged to give S_{stop}. Finally, a functional score was calculated by taking the inverse log of the slope of each variant divided by the slope of the wild-type sequence (S_{wt}), both corrected by S_{stop}:

$$\text{Functional score} = \frac{2^{S_{var}} - 2^{S_{stop}}}{2^{S_{wt}} - 2^{S_{stop}}}$$

Variants enriched over wild type would have a functional score >1, and variants de-enriched would have a functional score <1. Protein level functional scores were used to make heatmaps of each library sort. Sequence logos were made using Weblogo (Crooks et al., 2004). All scripts were written in Python, and plots were made with Matplotlib.

References

Akhtar, J., Mallaredy, V., Dandapat, J., Maiti, P., Sahoo, S.K., and Singh, S. (2012). PEGylation of an osteoclast inhibitory peptide: suitable candidate for the treatment of osteoporosis. Int J Pharm *434*, 429–436.

Araya, C.L., Fowler, D.M., Chen, W., Muniez, I., Kelly, J.W., and Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci USA *109*, 16858–16863.

Arch, R.H., and Thompson, C.B. (1998). 4-1BB and Ox40 are members of a tumor necrosis factor (TNF)-nerve growth factor receptor subfamily that bind TNF receptor-associated factors and activate nuclear factor kappaB. Mol Cell Biol *18*, 558–565.

Arcipowski, K.M., Stunz, L.L., Graham, J.P., Kraus, Z.J., Vanden Bush, T.J., and Bishop, G.A. (2011). Molecular mechanisms of TNFR-associated factor 6 (TRAF6) utilization by the oncogenic viral mimic of CD40, latent membrane protein 1 (LMP1). J Biol Chem *286*, 9948–9955.

Arron, J.R., and Choi, Y. (2000). Bone versus immune system. Nature *408*, 535–536.

Cheng, G., Cleary, A.M., Ye, Z.S., Hong, D.I., Lederman, S., and Baltimore, D. (1995). Involvement of CRAF1, a relative of TRAF, in CD40 signaling. Science *267*, 1494–1498.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188–1190.

Devergne, O., Hatzivassiliou, E., Izumi, K.M., Kaye, K.M., Kleijnen, M.F., Kieff, E., and Mosialos, G. (1996). Association of TRAF1, TRAF2, and TRAF3 with an Epstein-Barr virus LMP1 domain important for B-lymphocyte transformation: role in NF-kappaB activation. Mol Cell Biol *16*, 7098–7108.

Fowler, D.M., Araya, C.L., Gerard, W., and Fields, S. (2011). Enrich: software for analysis of protein function by enrichment and depletion of variants. Bioinformatics *27*, 3430–3431.

Ha, H., Han, D., and Choi, Y. (2009). TRAF-mediated TNFR-family signaling. Curr Protoc Immunol *Chapter 11*, Unit11.9D.

Halgren, T.A. (2009). Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model *49*, 377–389.

Halgren, T. (2007). New method for fast and accurate binding-site identification and analysis. Chem Biol Drug Des *69*, 146–148.

Hauer, J., Püschner, S., Ramakrishnan, P., Simon, U., Bongers, M., Federle, C., and Engelmann, H. (2005). TNF receptor (TNFR)-associated factor (TRAF) 3 serves as an inhibitor of TRAF2/5-mediated activation of the noncanonical NF-kappaB pathway by TRAF-binding TNFRs. Proc Natl Acad Sci USA *102*, 2874–2879.

Häcker, H., Tseng, P.-H., and Karin, M. (2011). Expanding TRAF function: TRAF3 as a tri-faced immune regulator. Nat. Rev. Immunol. *11*, 457–468.

Hildebrand, J.M., Luo, Z., Manske, M.K., Price-Troska, T., Ziesmer, S.C., Lin, W., Hostager, B.S., Slager, S.L., Witzig, T.E., Ansell, S.M., et al. (2010). A BAFF-R mutation associated with non-Hodgkin lymphoma alters TRAF recruitment and reveals new insights into BAFF-R signaling. J Exp Med *207*, 2569–2579.

Ishida, T.K., Tojo, T., Aoki, T., Kobayashi, N., Ohishi, T., Watanabe, T., Yamamoto, T., and Inoue, J. (1996a). TRAF5, a novel tumor necrosis factor receptor-associated factor family protein, mediates CD40 signaling. Proc Natl Acad Sci USA *93*, 9437–9442.

Ishida, T., Mizushima, S.I., Azuma, S., Kobayashi, N., Tojo, T., Suzuki, K., Aizawa, S., Watanabe, T., Mosialos, G., Kieff, E., et al. (1996b). Identification of TRAF6, a novel tumor necrosis factor receptor-associated factor protein that mediates signaling from an amino-terminal domain of the CD40 cytoplasmic region. J Biol Chem *271*, 28745–28748.

Li, C., Ni, C.-Z., Havert, M.L., Cabezas, E., He, J., Kaiser, D., Reed, J.C., Satterthwait, A.C., Cheng, G., and Ely, K.R. (2002). Downstream regulator TANK binds to the CD40 recognition site on TRAF3. Structure *10*, 403–411.

Li, C., Norris, P.S., Ni, C.-Z., Havert, M.L., Chiong, E.M., Tran, B.R., Cabezas, E., Reed, J.C., Satterthwait, A.C., Ware, C.F., et al. (2003). Structurally distinct recognition motifs in lymphotoxin-beta receptor and CD40 for tumor necrosis factor receptor-associated factor (TRAF)-mediated signaling. J Biol Chem *278*, 50523–50529.

Marinis, J.M., Homer, C.R., McDonald, C., and Abbott, D.W. (2011). A Novel Motif in the Crohn's Disease Susceptibility Protein, NOD2, Allows TRAF4 to Down-regulate Innate Immune Responses. J Biol Chem *286*, 1938–1950.

Matsuzawa, A., Tseng, P.-H., Vallabhapurapu, S., Luo, J.-L., Zhang, W., Wang, H., Vignali, D.A.A., Gallagher, E., and Karin, M. (2008). Essential cytoplasmic translocation of a cytokine receptor-assembled signaling complex. Science *321*, 663–668.

McWhirter, S.M., Pullen, S.S., Holton, J.M., Crute, J.J., Kehry, M.R., and Alber, T. (1999). Crystallographic analysis of CD40 recognition and signaling by human TRAF2. Proc Natl Acad Sci USA *96*, 8408–8413.

Nakano, H., Oshima, H., Chung, W., Williams-Abbott, L., Ware, C.F., Yagita, H., and Okumura, K. (1996). TRAF5, an activator of NF-kappaB and putative signal transducer for the lymphotoxin-beta receptor. J Biol Chem *271*, 14661–14664.

Nakano, H., Sakon, S., Koseki, H., Takemori, T., Tada, K., Matsumoto, M., Munechika, E., Sakai, T., Shirasawa, T., Akiba, H., et al. (1999). Targeted disruption of Traf5 gene causes defects in CD40- and CD27-mediated lymphocyte activation. Proc Natl Acad Sci USA *96*, 9803–9808.

Ni, C.Z., Welsh, K., Leo, E., Chiou, C.K., Wu, H., Reed, J.C., and Ely, K.R. (2000). Molecular

basis for CD40 signaling mediated by TRAF3. Proc Natl Acad Sci USA *97*, 10395–10399.

Ni, C.-Z., Oganesyan, G., Welsh, K., Zhu, X., Reed, J.C., Satterthwait, A.C., Cheng, G., and Ely, K.R. (2004). Key molecular contacts promote recognition of the BAFF receptor by TNF receptor-associated factor 3: implications for intracellular signaling regulation. J Immunol *173*, 7394–7400.

Niu, F., Ru, H., Ding, W., Ouyang, S., and Liu, Z.-J. (2013). Structural biology study of human TNF receptor associated factor 4 TRAF domain. Protein Cell *4*(9), 687-694.

Park, Y.C., Burkitt, V., Villa, A.R., Tong, L., and Wu, H. (1999). Structural basis for self-association and receptor recognition of human TRAF2. Nature *398*, 533–538.

Poblenz, A.T., Jacoby, J.J., Singh, S., and Darnay, B.G. (2007). Inhibition of RANKL-mediated osteoclast differentiation by selective TRAF6 decoy peptides. Biochem Biophys Res Commun *359*, 510–515.

Pullen, S.S., Dang, T.T., Crute, J.J., and Kehry, M.R. (1999a). CD40 signaling through tumor necrosis factor receptor-associated factors (TRAFs). Binding site specificity and activation of downstream pathways by distinct TRAFs. J Biol Chem *274*, 14246–14254.

Pullen, S.S., Labadia, M.E., Ingraham, R.H., McWhirter, S.M., Everdeen, D.S., Alber, T., Crute, J.J., and Kehry, M.R. (1999b). High-affinity interactions of tumor necrosis factor receptor-associated factors (TRAFs) and CD40 require TRAF trimerization and CD40 multimerization. Biochemistry *38*, 10168–10177.

Pullen, S.S., Miller, H.G., Everdeen, D.S., Dang, T.T., Crute, J.J., and Kehry, M.R. (1998). CD40-tumor necrosis factor receptor-associated factor (TRAF) interactions: regulation of CD40 signaling through multiple TRAF binding sites and TRAF hetero-oligomerization. Biochemistry *37*, 11836–11845.

Régnier, C.H., Tomasetto, C., Moog-Lutz, C., Chenard, M.P., Wendling, C., Basset, P., and Rio, M.C. (1995). Presence of a new conserved domain in CART1, a novel member of the tumor necrosis factor receptor-associated protein family, which is expressed in breast carcinoma. J Biol Chem *270*, 25715–25721.

Régnier, C.H., Masson, R., Kedinger, V., Textoris, J., Stoll, I., Chenard, M.-P., Dierich, A., Tomasetto, C., and Rio, M.-C. (2002). Impaired neural tube closure, axial skeleton malformations, and tracheal ring disruption in TRAF4-deficient mice. Proc Natl Acad Sci USA *99*, 5585–5590.

Rothe, M., Wong, S.C., Henzel, W.J., and Goeddel, D.V. (1994). A novel family of putative signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor receptor. Cell *78*, 681–692.

Rousseau, A., McEwen, A.G., Poussin-Courmontagne, P., Rognan, D., Nominé, Y., Rio, M.-C., Tomasetto, C., and Alpy, F. (2013). TRAF4 is a novel phosphoinositide-binding protein modulating tight junctions and favoring cell migration. PLoS Biol *11*, e1001726.

Shiels, H., Li, X., Schumacker, P.T., Maltepe, E., Padrid, P.A., Sperling, A., Thompson, C.B., and Lindsten, T. (2000). TRAF4 deficiency leads to tracheal malformation with resulting alterations in air flow to the lungs. Am. J. Pathol. *157*, 679–688.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.

Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics, doi: 10.1534/genetics.115.175802.

Tsitsikov, E.N., Laouini, D., Dunn, I.F., Sannikova, T.Y., Davidson, L., Alt, F.W., and Geha, R.S. (2001). TRAF1 is a negative regulator of TNF signaling. enhanced TNF signaling in TRAF1-deficient mice. Immunity *15*, 647–657.

Vallabhapurapu, S., Matsuzawa, A., Zhang, W., Tseng, P.-H., Keats, J.J., Wang, H., Vignali, D.A.A., Bergsagel, P.L., and Karin, M. (2008). Nonredundant and complementary functions of TRAF2 and TRAF3 in a ubiquitination cascade that activates NIK-dependent alternative NF-kappaB signaling. Nat Immunol *9*, 1364–1370.

Wu, S., Xie, P., Welsh, K., Li, C., Ni, C.-Z., Zhu, X., Reed, J.C., Satterthwait, A.C., Bishop, G.A., and Ely, K.R. (2005). LMP1 protein from the Epstein-Barr virus is a structural CD40 decoy in B lymphocytes for binding to TRAF3. J Biol Chem *280*, 33620–33626.

Xie, P. (2013). TRAF molecules in cell signaling and in human diseases. J Mol Signal *8*, 7.

Ye, H., Park, Y.C., Kreishman, M., Kieff, E., and Wu, H. (1999). The structural basis for the recognition of diverse receptor sequences by TRAF2. Mol Cell *4*, 321–330.

Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaia, M., Yim, M., et al. (2002). Distinct molecular mechanism for initiating TRAF6 signalling. Nature *418*, 443–447.

Yoon, J.H., Cho, Y.-J., and Park, H.H. (2014). Structure of the TRAF4 TRAF domain with a coiled-coil domain and its implications for the TRAF4 signalling pathway. Acta Crystallogr D Biol Crystallogr *70*, 2–10.

Zapata, J.M., Lefebvre, S., and Reed, J.C. (2007a). Targeting TRAfs for therapeutic intervention. Adv Exp Med Biol *597*, 188–201.

Zapata, J.M., Martínez-García, V., and Lefebvre, S. (2007b). Phylogeny of the TRAF/MATH domain. Adv Exp Med Biol *597*, 1–24.

Zepp, J.A., Liu, C., Qian, W., Wu, L., Gulen, M.F., Kang, Z., and Li, X. (2012). Cutting edge: TNF receptor-associated factor 4 restricts IL-17-mediated pathology and signaling processes. J Immunol *189*, 33–37.

Zhang, P., Reichardt, A., Liang, H., Aliyari, R., Cheng, D., Wang, Y., Xu, F., Cheng, G., and Liu, Y. (2012). Single Amino Acid Substitutions Confer the Antiviral Activity of the TRAF3 Adaptor Protein onto TRAF5. Sci Signal *5*, ra81–ra81.

Zheng, C., Kabaleeswaran, V., Wang, Y., Cheng, G., and Wu, H. (2010). Crystal structures of the TRAF2: cIAP2 and the TRAF1: TRAF2: cIAP2 complexes: affinity, specificity, and regulation. Mol Cell *38*, 101–113.

Chapter 5

Conclusions and future directions

The ability to characterize protein-peptide interaction determinants has been greatly improved by the combination of peptide library screening and deep sequencing technology. My thesis work on the Bcl-2 family utilized models built on small-scale mutational datasets and structure-based models, leveraging these models to design and screen peptide libraries for selective inhibitors. My work on the TRAF family has begun to build a more comprehensive mutational dataset on peptide interaction preferences that will yield valuable information for models of TRAF-peptide interactions. The work done on these families of peptide recognition domain paralogs has emphasized how divergent their interaction preferences can be and the scope of the data needed to provide accurate models of binding specificity. This chapter discusses lessons from my work and future directions in three areas related to protein-peptide interaction specificity. First, I discuss approaches to the initial characterization of binding preferences. Second, I compare strategies for design and screening of peptide libraries for specific binding partners. Finally, I propose biological applications of peptides that specifically bind viral Bcl-2 and TRAF family members.

Characterization and modeling of peptide binding preferences

A thorough characterization of binding preferences for peptide recognition domains (PRDs) provides the best basis for future design of specific binding partners or interactome prediction. Historically, binding preferences have been assayed in relatively small-scale mutagenesis or peptide array experiments. This was true for the characterization of the BH3 binding preferences of the anti-apoptotic Bcl-2 homologs. Boersma et al. performed alanine and hydrophile (lysine and glutamate) scanning of 18 positions in the Bim BH3 and tested binding to the human homologs Mcl-1 and Bcl-x_L by fluorescence polarization assays (Boersma et al.,

2008). Stefano Gullá used SPOT arrays to assay binding of the five main human Bcl-2 homologs to mutants of the Bim BH3 at 10 peptide positions to 18 amino acids (DeBartolo et al., 2012; Dutta et al., 2010; London et al., 2012). He did the same SPOT array experiment for two viral Bcl-2 homologs, KSBcl-2 and BHRF1 (Chapter 3). These mutagenesis and SPOT array experiments provided valuable datasets with which to compare the BH3 binding preferences of viral and human Bcl-2 homologs. However, the datasets did not cover all positions of the BH3 peptide that influence binding, which includes at least 20-25 positions. $PSSM_{SPOT}$ models have been built from the Bim SPOT array data, and these have demonstrated predictive power in screening the proteome for new BH3 sequences (DeBartolo et al., 2014). These models have also been used by myself and others to design libraries enriched in specific binding partners of Bfl-1, Bcl-x_L, KSBcl-2, or BHRF1 (Dutta et al., 2013; 2014) (and Chapter 3). However, the success of the $PSSM_{SPOT}$ models for interactome prediction or specificity design is limited by their incomplete coverage of peptide positions. This was exemplified by the discovery by Dutta et al. (2014) and myself in Chapter 3, that the BH3 position 4e, located near the C-terminus of the peptide, is an important Bcl-x_L vs. Bcl-2 and KSBcl-2/BHRF1 vs. Mcl-1 specificity determinant. This position was not varied in the libraries targeting Bcl-x_L and the viral Bcl-2 homologs due to the fact that mutational preferences of all of the homologs were not known for this position because it was not varied on the Bim BH3 SPOT arrays.

Mutagenesis and peptide array experiments are valuable for their direct, quantitative (or semi-quantitative) measurement of binding, but are limited in their coverage by the effort required or cost of array synthesis. These practical considerations lead researchers to test only the peptide positions they think most likely to influence binding based on the available structural data. More recent approaches using peptide library display paired with deep sequencing allow a

less biased coverage of the sequence space because they are higher throughput. The challenge with library methods lies in extracting a quantitative metric of affinity from the sequencing results. Two library approaches have the potential to provide quantitative affinity estimates of PRDs for comprehensive single and double mutant libraries of peptides: Sortcery and deep mutational scanning.

The Sortcery method was developed to rank peptides according to affinity by quantifying their distribution across gates set in FACS experiments (Reich et al., 2014). It was used to rank ~1000 BH3 variants according to their affinity for the Bcl-2 homolog Bcl-x_L. The BH3 peptide variants were displayed by on the yeast surface and they were sorted into pools corresponding to different regions of the FACS plot, which correspond to different degrees of affinity. These individual pools were then Illumina sequenced, and the relative rank of each variant was computed from its distribution across the FACS regions as quantified by the number of copies of the sequence in the pools. The Sortcery ranking of 19 variants correlated well with their K_D values measured on the surface of yeast, allowing the extrapolation of K_D values from rankings. Thus far, this method has only been used with small yeast surface display libraries of combinatorial variants. In theory, it could be used with other display systems that can be screened by FACS, such as bacterial surface display. With sufficient sampling, it could also be used to affinity-rank larger libraries of tens-of-thousands of variants. The ability to affinity rank small libraries makes Sortcery ideal for quantifying binding to single and double mutant libraries. As only one round of FACS is needed for a Sortcery experiment, the binding of several paralogs to a peptide library could be assayed in the space of a week. The sample preparation for Illumina sequencing and subsequent data analysis would be much more time intensive.

Significant experimental testing or simulation would be necessary for adaptation of Sortcery to different protein-peptide interaction systems, display methods, and affinity ranges.

Deep mutational scanning (DMS) is a second approach to correlating sequencing results to binding. Conceptually, it is a simpler approach than Sortcery, and it has been used to quantify binding to libraries screened by a variety of methods including bacterial two-hybrid, phage display, and yeast surface display (Fowler et al., 2010; McLaughlin et al., 2012; Whitehead et al., 2012). Several different ways of analyzing DMS data have been published, but at their core, they all use the relative frequency of variants in the selected pool versus the naïve library as a readout of the function assayed (Araya et al., 2012; Fowler et al., 2011; McLaughlin et al., 2012; Starita et al., 2015). Testing of the correlation of sequence enrichment with affinity has been lacking in most DMS experiments. As with Sortcery, more rigorous analysis of this correlation is needed for different library display systems, protein-protein interaction families, and affinity ranges.

Characterization of TRAF-peptide interactions is in its infancy. Therefore, large-scale, unbiased screening of preferences is needed. For this reason, I chose to perform a deep mutational scanning experiment to assay TRAF binding preferences. In Chapter 4, I assayed binding of TRAFs 2, 3, and 5 to single and double mutants of 17- and 21-mer regions of TANK and CD40 by bacterial surface display and sequencing enrichment analysis. Future work on this project will include analysis of the correlation of binding affinity with enrichment. However, preliminary analysis showed agreement of trends in single mutant enrichment data with previously published SPOT array results (Pullen et al., 1999). These experiments only tested binding to mutants of the peptide core motif and regions C-terminal to the core. Future experiments will be needed to explore binding determinants N-terminal to the core. Data from single and double mutant libraries are easily interpretable and can be incorporated into models of

binding preferences. However, they only cover a small fraction of the possible peptide sequence space. One way to address this is to test single mutant preferences in several peptide backbones. Comparison of such results would allow exploration of covariation between peptide positions. DMS experiments performed for multiple TRAF paralogs on multiple peptide backgrounds will provide a valuable resource to compare and model TRAF binding preferences.

Design and screening of peptide libraries to achieve specific peptides

In chapters 2 and 3, I screened libraries of BH3 peptides for binding specificity to the viral Bcl-2 homologs KSBcl-2 and BHRF1. The library screened in chapter 2 for KSBcl-2 specificity was designed by Stefano Gulla based on the Bim BH3 SPOT arrays. This library screening only yielded partial specificity for KSBcl-2, in that the resulting peptides showed modest specificity over the human homologs Bfl-1, Bcl-x_L, Bcl-2, and Bcl-w, but bound tightly to Mcl-1. This first-generation library only varied six positions along the peptide. In chapter 3, I used the SPOT array data and structure based STATIUM models to design libraries that varied nine peptide positions. These libraries were somewhat more successful in that the resulting peptides had large margins of specificity over most human homologs, and a modest amount of specificity over Mcl-1. However, I only achieved large margins of specificity over Mcl-1 when I mutated a further position that was not varied in the libraries.

Several lessons can be learned from the work done by myself and others to design specific peptide inhibitors of Bcl-2 family members. First, and most obviously, the success of a library screening experiment in resulting in peptides with the desired degree of affinity and specificity depends in large part on what mutations go into the library. Given the large amounts of structural and binding data available for Bcl-2:BH3 interactions, it has been tempting to try to

design libraries that are enriched in peptides with the desired binding properties. Dutta et al. have use this approach to design BH3 peptides libraries for selectivity to Bfl-1 and Bcl-x_L (Dutta et al., 2013; 2014). Both of these libraries varied seven positions along the BH3, and were restricted to positions for which there was information from the Bim BH3 SPOT arrays. The Bfl-1 library screen resulted in peptides that discriminated against Bcl-2, Bcl-w, and Bcl-x_L, but still bound Mcl-1 with low nanomolar affinity. Likewise, the Bcl-x_L screen discriminated against Mcl-1 and Bfl-1 binding, but the peptides showed moderate-to-tight binding to Bcl-2 and Bcl-w. This is in agreement with my viral Bcl-2 library results and emphasizes that discriminating between homologs with similar binding preferences is a challenging problem.

An unanswered question in the field is what are the most important considerations in designing and screening libraries. Is a carefully designed library more likely to be successful than a random library? Do computational approaches to library design improve success over the rational library design of a knowledgeable biochemist? How large of a library is needed, and how many positions should be varied for a given objective? Which is more likely to result in a high affinity, high specificity peptide: starting with a peptide backbone of moderate affinity and specificity, or a peptide with high affinity and no specificity? A systematic examination of these questions would be a boon to the large field of peptide and protein library screening. Random libraries have the potential to sample mutations at many positions. However, the sequence space is large and existing display and screening methods cannot sample all combinations of mutations. Therefore, screening a random library may identify individual beneficial mutation, but is unlikely to sample all possible combinations of beneficial mutations. For example, a random library might have sampled the Y4eK mutation and shown that it was good for Bcl-x_L and viral Bcl-2 specificity. However, as my mutational analysis of the Mcl-1-specific peptides in chapter 2

shows, achieving a peptide with broad specificity against several close paralogs requires the combined effects of mutations at several positions. Any one position is unlikely to provide specificity against all paralogs.

The goal of including more sequences with the desired binding properties can be addressed with larger libraries and better information upon which to base the library design. In chapter 3, I adapted a bacterial surface display system for BH3 peptide library display. Cell surface display methods offer real-time affinity resolution of library members when paired with FACS. The size of a typical yeast surface display library is limited to ~10^7 by the transformation efficiency of yeast. Bacterial surface display on *E. coli* allows libraries of 10^{10} because of the higher transformation efficiency of gram negative bacteria. However, initial screening of such large libraries is limited by the throughput of FACS to ~10^8, so larger bacterial display libraries must first be enriched by magnetic bead sorting. The *E. coli* eCPX system has been used in the past without a control for expression. By adding an epitope tag, I enabled affinity-based screening of peptide libraries, creating a system combining the best properties of the pre-existing yeast and bacterial surface display systems. The larger libraries will allow sampling of larger sequence spaces. The bacterial display system would also be a good approach for comparison of several library design strategies due to the speed with which libraries can be screened.

To improve the quality of sequences going into the library, better models of binding preferences are needed. Two approaches to this are possible. One, more exhaustive experimental sampling of the peptide sequence space with single or double mutant libraries would be beneficial in creating more comprehensive models of binding preferences. The BH3 peptide libraries in chapters 1 and 2 and the work by Dutta et al. were primarily limited to the 10 positions covered on the Bim BH3 SPOT arrays. A deep mutational scanning experiment on

several BH3 peptides would provide a more comprehensive dataset upon which to model Bcl-2 binding preferences. The approach of using deep mutational scanning results to inform combinatorial library design has been used with success to improve the affinity of protein-protein interactions (Fujino et al., 2012; Whitehead et al., 2012). It would be interesting to combine deep mutational scanning results for several paralogs to design combinatorial libraries directed at specificity. A second approach to improving library design would be to rely more heavily on computational models of binding preferences. STATIUM predictions were correct for position 2a and 4e, which were mutated in the final viral Bcl-2-specific peptides, but not present on the SPOT arrays. However, further validation of such computational models is needed in order to know what types of interactions they are best at modeling, and where they might be incorrect. This argues for the continued combination of experiment-based and structure-based models for library design, as I did in chapter 3.

The deep mutational scanning results for TRAFs 2, 3and 5 showed differences between the binding preferences of TRAF3 versus TRAFs 2 and 5, suggesting that, in the future, combinatorial peptide libraries could be designed to target specificity for individual TRAFs. My DMS datasets could be used to inform the design of such combinatorial libraries. Computational modeling of TRAF-peptide interactions could also provide leads on specificity mechanisms. However, the possibility of flexibility and conformational diversity in the manner in which the peptide termini bind to TRAFs presents a challenge for creating computational models of binding preferences. Therefore, the DMS datasets will also be important in benchmarking possible computational approaches. Modeling approaches for flexible peptides include the Rosetta FlexPepDock and Rosetta FloppyTail methods, among others (Kleiger et al., 2009; London et al., 2011). The combinatorial libraries targeting viral Bcl-2 binding specificity in

chapters 2 and 3 built upon many years of mutational analysis and structural characterization of Bcl-2:BH3 interactions. Likewise, a thorough understanding of TRAF-peptide interaction preferences may be needed before highly specific peptides can be obtained.

Applications of specific peptides

Peptide reagents that specifically bind only one member of a paralogous family are useful for a variety of applications. In chapter 2, I created a peptide, MS1 that bound with low nanomolar affinity to Mcl-1 and micromolar affinity to the other four human Bcl-2 homologs. The *in vitro* binding specificity of MS1 was reflected in cellular BH3 profiling assays. Research groups are now using MS1 as a reagent to detect anti-apoptotic dependence on Mcl-1 in primary tumor samples. Specific peptides are excellent reagents for use in diagnostic assays. The barriers to the use of peptides as therapeutics (e.g. lack of cell permeability and protease resistance) are not issues in diagnostic assays, which can be performed in a short time-span in permeabilized cells. Other groups are also using MS1 as a competitor in small molecule screens for specific inhibitors of Mcl-1. With further modification, such as chemical crosslinking to improve protease resistance and cell penetrance, inhibitory peptides also have the potential to serve as therapeutics. The sequence of MS1 has served as a starting point for the design of chemically cross-linked peptide inhibitors of Mcl-1.

Recently, BH3 profiling assays have been used to investigate the dependency of herpesvirus-infected cell lines on viral and human anti-apoptotic Bcl-2 homologs (Cojohari et al., 2015). The picture of Bcl-2 dependence in virally infected cells is complicated due to several factors including the up-regulation of human anti-apoptotic Bcl-2 members by viral infection, the utilization of the viral Bcl-2 homologs, and differential Bcl-2 expression levels depending on the

stage of viral infection and the cell type. Deng et al. have shown that the mitochondrial permeabilization response measured in BH3 profiling assays is a better indicator of dependence on Bcl-2 homologs than the levels of Bcl-2 expression (Deng et al., 2007). BH3 profiling assays can be rapidly performed on many different cell lines including cell lines in different stages of viral infection, thus offering an advantage over genetic knockout studies, with which it would be laborious to monitor the dependence on all human and viral Bcl-2 homologs.

The BH3 profiling assay performed on an Epstein Barr virus (EBV) infected cell line, Akata cells, showed a signature indicative of a dependence upon Bcl-x_L, but expression of the EBV Bcl-2 homolog BHRF1 was not detected in the cells (Cojohari et al., 2015). BHRF1 only binds to a small subset of natural BH3-only proteins that are also bound by all human Bcl-2 homologs. Thus, my BHRF1-specific peptide, BL6-22_Y4eK provides a unique ability to detect BHRF1 dependence. This peptide could be used to profile a panel of EBV-infected cell lines in different stages of viral infection in order to determine under what conditions BHRF1 is important for prevention of apoptosis.

BH3 profiling on a Kaposi sarcoma herpesvirus (KSHV) infected cell line, Bcbl-1, showed a profile that suggested that the cells were dependent upon both Mcl-1 and the KSHV Bcl-2 homolog KSBcl-2 (Cojohari et al., 2015). KSBcl-2 dependence can be detected using existing natural BH3 peptides, but as none specifically bind KSBcl-2, detection requires a panel of peptides. My KSBcl-2-specific peptide, KL6-7_Y4eK provides the ability to specifically detect KSBcl-2 dependence with one peptide. The high affinity of KL6-7_Y4eK for KSBcl-2 and the large margin of specificity over human Bcl-2 homologs makes this peptide a promising candidate for therapeutic development.

There is also precedent for using peptides corresponding to TRAF binding sites (TBSs) is inhibitors and modulators of TRAF signaling pathways. Peptides corresponding to the TRAF6 binding site in RANK have been used to prevent osteoclast differentiation, which is desired to block osteoporosis (Ye et al., 2002). While numerous examples of natural TRAF6-specific TBSs exist, there are fewer examples of TBSs that differentiate between TRAFs 1, 2, 3 and 5. Many natural TBSs have not been profiled for binding to all TRAFs, so their specificity is unknown. One possible application of specific TBSs would be to insert the specific sequences in place of TBSs in full-length TRAF binding partners in order to study the dependence of downstream signaling events on individual TRAFs. This approach has been taken with mutants of CD40 to study the dependence of NF-κB, JNK, and p38 MAPK activation on TRAF2, TRAF3, and TRAF6 binding to CD40 (Pullen et al., 1999). An important consideration for this type of experiment would be the level of specificity needed to prevent off-target binding in the cell. TRAFs have been demonstrated to form high-order oligomers in the cell, allowing signaling even from complexes that show undetectable binding in monomeric binding assays (Graham et al., 2009). Weak interactions could be tested *in vitro* in a highly avid system such as a plate-based assay with peptides immobilized on the plate and trimeric TRAFs further oligomerized by streptavidin or beads. Considering that TRAFs function in numerous signaling pathways, this TBS-swapping approach would offer a more directed and less disruptive strategy for studying TRAF function than genetic knockout. Beyond design of specific peptide reagents, an understanding of TRAF binding specificity will allow prediction of new interaction partners from the proteome and the formation of functional hypotheses about the specificity of signaling through individual TRAF paralogs.

References

Araya, C.L., Fowler, D.M., Chen, W., Muniez, I., Kelly, J.W., and Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci USA *109*, 16858–16863.

Boersma, M., Sadowsky, J., and Tomita, Y. (2008). Hydrophile scanning as a complement to alanine scanning for exploring and manipulating protein-protein recognition: Application to the Bim BH3 domain. Protein Science *17*, 1232-1240.

Cojohari, O., Burrer, C.M., Peppenelli, M.A., Abulwerdi, F.A., Nikolovska-Coleska, Z., and Chan, G.C. (2015). BH3 Profiling Reveals Selectivity by Herpesviruses for Specific Bcl-2 Proteins to Mediate Survival of Latently Infected Cells. J Virol, doi:10.1128/JVI.00236-15.

DeBartolo, J., Dutta, S., Reich, L., and Keating, A.E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. J Mol Biol *422*, 124–144.

DeBartolo, J., Taipale, M., and Keating, A.E. (2014). Genome-wide prediction and validation of peptides that bind human prosurvival Bcl-2 proteins. PLoS Comput Biol *10*, e1003693.

Deng, J., Carlson, N., Takeyama, K., Dal Cin, P., Shipp, M., and Letai, A. (2007). BH3 profiling identifies three distinct classes of apoptotic blocks to predict response to ABT-737 and conventional chemotherapeutic agents. Cancer Cell *12*, 171–185.

Dutta, S., Chen, T.S., and Keating, A.E. (2013). Peptide ligands for pro-survival protein Bfl-1 from computationally guided library screening. ACS Chem Biol *8*, 778–788.

Dutta, S., Gullá, S., Chen, T.S., Fire, E., Grant, R.A., and Keating, A.E. (2010). Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. J Mol Biol *398*, 747–762.

Dutta, S., Ryan, J., Scott Chen, T., Kougentakis, C., Letai, A., and Keating, A.E. (2014). Potent and specific peptide inhibitors of human pro-survival protein Bcl-xL. J Mol Biol *427*, 1241-1253.

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat Methods *7*, 741–746.

Fowler, D.M., Araya, C.L., Gerard, W., and Fields, S. (2011). Enrich: software for analysis of protein function by enrichment and depletion of variants. Bioinformatics *27*, 3430–3431.

Fujino, Y., Fujita, R., Wada, K., Fujishige, K., Kanamori, T., Hunt, L., Shimizu, Y., and Ueda, T. (2012). Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. Biochem Biophys Res Commun *428*, 395–400.

Graham, J.P., Moore, C.R., and Bishop, G.A. (2009). Roles of the TRAF2/3 binding site in differential B cell signaling by CD40 and its viral oncogenic mimic, LMP1. J Immunol *183*, 2966–2973.

Kleiger, G., Saha, A., Lewis, S., Kuhlman, B., and Deshaies, R.J. (2009). Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. Cell *139*, 957–968.

London, N., Gullá, S., Keating, A.E., and Schueler-Furman, O. (2012). In silico and in vitro elucidation of BH3 binding specificity toward Bcl-2. Biochemistry *51*, 5841–5850.

London, N., Lamphear, C.L., Hougland, J.L., Fierke, C.A., and Schueler-Furman, O. (2011). Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS Comput Biol *7*, e1002170.

McLaughlin, R.N., Jr, Poelwijk, F.J., Raman, A., Gosal, W.S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–142.

Pullen, S.S., Dang, T.T., Crute, J.J., and Kehry, M.R. (1999). CD40 signaling through tumor necrosis factor receptor-associated factors (TRAFs). Binding site specificity and activation of downstream pathways by distinct TRAFs. J Biol Chem *274*, 14246–14254.

Reich, L.L., Dutta, S., and Keating, A.E. (2014). SORTCERY-A High-Throughput Method to Affinity Rank Peptide Ligands. J Mol Biol *427*, 2135-2150.

Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics, doi: 10.1534/genetics.115.175802.

Whitehead, T.A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S.J., De Mattos, C., Myers, C.A., Kamisetty, H., Blair, P., Wilson, I.A., et al. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol *30*, 543–548.

Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaia, M., Yim, M., et al. (2002). Distinct molecular mechanism for initiating TRAF6 signalling. Nature *418*, 443–447.