Published in final edited form as:
J Mol Biol. 2013 January 9; 425(1): 171–185. doi:10.1016/j.jmb.2012.11.009.

Structure based re-design of the binding specificity of anti-apoptotic Bcl-x_L

T. Scott Chen, **Hector Palacios**, and **Amy E. Keating**[1]
Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139

Abstract

Many native proteins are multi-specific and interact with numerous partners, which can confound analysis of their functions. Protein design provides a potential route to generating synthetic variants of native proteins with more selective binding profiles. Re-designed proteins could be used as research tools, diagnostics or therapeutics. In this work, we used a library screening approach to re-engineer the multi-specific anti-apoptotic protein Bcl-x_L to remove its interactions with many of its binding partners, making it a high affinity and selective binder of the BH3 region of pro-apoptotic protein Bad. To overcome the enormity of the potential Bcl-x_L sequence space, we developed and applied a computational/experimental framework that used protein structure information to generate focused combinatorial libraries. Sequence features were identified using structure-based modeling, and an optimization algorithm based on integer programming was used to select degenerate codons that maximally covered these features. A constraint on library size was used to ensure thorough sampling. Using yeast surface display to screen a designed library of Bcl-x_L variants, we successfully identified a protein with ~1,000-fold improvement in binding specificity for the BH3 region of Bad over the BH3 region of Bim. Although negative design was targeted only against the BH3 region of Bim, the best re-designed protein was globally specific against binding to 10 other peptides corresponding to native BH3 motifs. Our design framework demonstrates an efficient route to highly specific protein binders and may readily be adapted for application to other design problems.

Keywords

computational library design; protein engineering; library screening; protein interaction specificity; Bcl-2

Introduction

Many proteins are capable of interacting with multiple protein partners, often leading to different functional consequences. The ability to manipulate interaction specificity by re-design could provide a valuable tool for investigating the roles of individual interactions. This concept has been illustrated in numerous studies. For example, Vidal and colleagues have promoted the idea of an “edgetic perturbation”, whereby a mutant of a protein loses interaction with one partner but maintains binding to others[1]. Because edgetic mutants can

[1]To whom correspondence should be addressed. keating@mit.edu.

be used to evaluate the biological consequences of losing just an edge in an interaction graph, rather than an entire node, this is potentially a much more precise and informative perturbation. From either a research or a therapeutic perspective, edgetic mutants of native proteins could be useful as selective dominant-negative inhibitors[2]. In addition to such practical benefits, attempts to rationally redesign interaction properties test our understanding of how protein sequence determines specific binding.

Some multi-specific proteins interact with different partners using different interfaces. In such instances, maintaining certain interactions while abrogating others can sometimes be achieved by simply introducing disruptive mutations to one of the binding sites. For example, signaling proteins often have multiple protein interaction domains and use them to interact with different partners[3], and disruptive mutation of each domain can remove interaction with the corresponding partner[4]. On the other hand, many proteins interact with multiple partners using the same interface, and residue contacts made between different complexes can be highly similar. For these proteins, selectively removing a subset of interactions without affecting others can be more difficult. For example, several conserved hydrophobic residues at the binding interface of the Src SH3 domain, when mutated, were shown to abolish or significantly weaken binding to multiple partners[5].

The anti-apoptotic Bcl-2 (B cell lymphoma) proteins are multi-specific[6]. Proteins in this family have a globular, helical fold and interact with many different partners. A large number of their partners contain short helical segments referred to as BH3 motifs. Peptides corresponding to BH3 motifs are competent for binding and are here referred to as BH3 peptides; in our notation, Bim BH3 is a peptide corresponding to the BH3 region of protein Bim. Native anti-apoptotic Bcl-2 proteins bind to many different BH3 peptides using the same interface with varying affinities[7]. The BH3-binding specificities of anti-apoptotic proteins are important for their biological roles regulating apoptosis, and differences between family members have presented an obstacle to developing pan-specific small molecule inhibitors as cancer therapeutics[8]. Mutants of Bcl-2 proteins with altered interaction properties have been used to elucidate the role of these proteins in cell-death processes[9,10]. For example, Andrews and colleagues created a mutant of Bcl-x_L that retained binding to Bid but not to Bax[9]. Use of this mutant in an *in vitro* membrane permeabilization system suggested that the anti-apoptotic activity of Bcl-x_L depended on its interactions with both Bid and Bax.

So far, little is known about which structural features confer the distinct binding profiles of different Bcl-2 family proteins. For example, studies that transplanted residues from one family member to another failed to switch binding specificity[11,12]. To investigate determinants of Bcl-2 family binding specificity, we sought to re-design anti-apoptotic protein Bcl-x_L so that it would lose the ability to strongly interact with Bim BH3 but retain tight binding to a BH3 peptide derived from Bad. This is an interesting problem because all known human anti-apoptotic Bcl-2 proteins interact strongly with Bim, which is proposed as an “activator” BH3 in some models of the regulation of apoptosis[13–16]. In contrast, the BH3-only protein Bad, proposed as a “sensitizer”, interacts with anti-apoptotic proteins in a more selective manner. The well-established specificity of Bcl-x_L for binding to Bad but not the related BH3 motif of Noxa demonstrates that selective binding can be achieved in some instances, and differences in the sequences of the Bim vs. Bad BH3 motifs make distinguishing these two partners appear feasible. In the longer term, a panel of re-designed selective proteins would provide useful reagents for deciphering the regulatory roles of Bcl-2 interactions, especially given that many assays in this area of research are done in extracts or with liposomes, where it would not be technically difficult to deploy engineered reagents[9, 17, 18].

Approaches commonly used to re-engineer proteins include computational protein design and experimental library screening[19–21]. The former offers great promise, but is still a maturing field. Efforts to computationally design protein-protein interaction specificity using structural information have been reported[19,21]. In a pioneering study, Havranek et al.[22] suggested the importance of explicitly considering targets and off-targets in the design process for this type of problem. Kortemme et al.[23] proposed a "computational second site suppressor" strategy to redesign both partners of a protein interface and showed that the redesigned interface retained specificity in a cellular context. Increasingly powerful and sophisticated algorithms have since been developed to facilitate multi-state design[24,25]. Despite these advances, it is worth noting that the scoring methods of modern design methodologies, which rely on computing terms based on physical interactions or the statistics of observed interactions in known protein structures, fall short of providing high accuracy for predicting binding specificity[26,27]. Thus, the risk of designed sequences not working as expected for this type of problem is high.

Experimental library screening is a powerful approach for identifying proteins with altered binding properties. However, the enormity of the possible sequence space can make screening of a completely random library an inefficient process. Efforts have been described that use computational modeling to design more focused libraries[21,28–37]. In these studies, an objective for the library is defined, such as the average of the predicted energies for all library sequences, and the library is optimized under different constraints such as the library size.[30,31,33,34] Alternatively, an ensemble of computationally designed sequences can be generated first, and the library designed to maximally cover these sequences.[29,32] Successes in application such as improving enzymatic activity[29,36] and discovering novel protein binders[32,37] have been reported. However, the use of structure-based design to generate focused libraries is still an emerging area where much remains to be optimized. In this study, we developed and tested a new framework that can be applied to problems of this type.

Our library design framework consisted of two stages. In the first, desired sequence features were predicted using both manual structure inspection and the modeling software Rosetta[38]. Desired features were defined permissively to include residues predicted to maintain binding to a target (here the BH3 region of protein Bad) and residues predicted to impart specificity for that target over an alternative (the BH3 region of Bim). In the second stage, we formulated the task of library optimization as an integer linear program[39] (ILP). The goal was to design a combinatorial, degenerate codon-based DNA library encoding the desired sequence features efficiently. ILP optimization is a flexible approach that can be use to optimally incorporate many different biases and restrictions into construction of a library[34].

Applying this approach, we designed libraries of Bcl-x_L variants and screened them using yeast surface display[40]. We successfully obtained proteins that showed a strong preference for binding Bad over Bim BH3 peptides. Detailed investigation of the sequence characteristics revealed that maintaining high library diversity was important for identifying high specificity sequences in this work. We further showed that our designed protein is globally specific against binding 10 other BH3-only peptides not considered in library design, with interesting implications for specificity design involving multiple undesired partners.

Results

Library design

Our two-stage library design process included two stages (Fig. 1A): In the first, we picked design positions to be randomized and candidate amino acids to be encoded at these positions. Guided by crystal structures of complexes between Bcl-x_L and Bim[41] or Bad[42],

we chose 9 Bcl-x_L sites where contacts are made to the central part of the Bim or Bad BH3 peptides. These 9 positions interact with BH3 peptide sites that are occupied by different amino acids in Bim vs. Bad (Fig. 1B, 1C). Including all 20 amino acids at these 9 positions using degenerate codons would lead to a library of size 32^9 (~4×10^{13}), greatly exceeding the limit of the yeast surface display method (~10^7–10^8). We used available crystal structures of complexes between Bcl-x_L and Bim or Bad, and a modeling procedure described below, to narrow the set of candidate amino acids to ones more likely to contribute to the desired binding preference.

To choose which residues to prioritize for inclusion, we first looked for residues likely to retain binding to the target Bad peptide. We used criteria based on hydrophobicity and size to manually eliminate certain amino acids from consideration (Table 1). We then employed the structural modeling suite Rosetta to do further pruning. Modeled complexes between Bcl-x_L point mutants and Bad were generated and their Rosetta energy scores relative to that of the native amino acid, ΔE_{Bad}, were obtained (see Materials and Methods). It should be noted that predicting reliable differences in binding energies from structure is extremely challenging[43], so our scheme did not rely on high accuracy. Rather, we defined a maximum ΔE_{Bad} value as a cutoff to eliminate mutations with highly unfavorable Rosetta scores in the modeled complexes. Backbone flexibility was introduced by the backrub utility in the software suite. The remaining residues were defined as non-disruptive mutations. The protocol did not consider higher-order interactions among mutations at different designed positions. Although some such dependencies can potentially be modeled, they are difficult to experimentally encode in combinatorial libraries, so we left identification of appropriate coupling for the screening process. To introduce negative design, we attempted to identify mutations that could disfavor binding to Bim BH3. For each point substitution that was modeled, we tabulated the difference in Rosetta energy scores for each mutant Bcl-x_L interacting with Bim vs. Bad ($\Delta E_{Bim} - \Delta E_{Bad}$). Residues with a score difference above a certain threshold were predicted as specific mutations (Table 1). We reasoned that these mutations were more likely to contribute to the desired binding preference, compared to the predicted non-disruptive mutations, and should be prioritized in library optimization. Note that the predicted specific mutations were a subset of the predicted non-disruptive mutations.

A naïve, degenerate codon-based library designed to include all native, predicted non-disruptive and predicted specific mutations (Table 1, see Materials and Methods) had a library size of 7.2×10^7 (the total number of unique DNA sequences). Assuming the number of experimentally accessible library sequences to be 2×10^7 (the number of yeast transformants obtained in this study), the probability that a particular sequence would be sampled is around 0.25 (see Materials and Methods). Also note that among all library DNA sequences in this naïve library, only ~4.4% encoded protein sequences with all positions occupied by predicted non-disruptive mutations. This is because undesired amino acids were encoded in the library due to the limitation imposed by degenerate codon usage. We decided to compact the library further to a size below 10^7 and simultaneously improve the predicted fraction of potential binders.

We developed a framework for optimizing combinations of degenerate codons encoding diversity at designed positions under a constraint on library size. We formulated the optimization problem to be solved as an ILP, i.e. a system of equations that describes both the quantity to be optimized and an arbitrary number of constraints on the solution as linear functions of integer variables (see Materials and Methods)[39]. This is a convenient approach because many different linear constraints can be included, and existing software packages can solve this type of problem efficiently, providing a provably optimal solution. The objective to be maximized was the number of unique protein sequences in the library with all designed positions occupied by predicted non-disruptive mutations (including the native

amino acids). This objective can be loosely interpreted as the number of unique protein sequences predicted to bind the desired target Bad with high affinity (making several assumptions). We enforced two constraints in the ILP optimization. The first was on the library size in DNA sequence space, which was set to 10^7 for reasons described above. The second was that all predicted specificity mutations, as well as all native residues, were required to be included in the library. This reflected our willingness to enhance the probability of sampling native and predicted specific mutations at the expense of missing a few predicted non-disruptive mutations. The optimized library (referred to as library 1 in this study, Table 1) had a size of 8.9×10^6 and contained 2.2×10^5 unique protein sequences (encoded by 5.8×10^5 DNA sequences) without any disruptive residues. Compared to the library without optimization described in the previous paragraph, the probability that a library sequences is sample increased from 0.25 to 0.9 (see Materials and Methods), and the percentage of library DNA sequences encoding protein sequences whose positions were all occupied by predicted non-disruptive mutations (approximated as high-affinity Bad binders) increased from 4.4% to 6.6%. On the other hand, 3 predicted non-disruptive mutations (out of 41) were excluded from the optimized library. Note that all calculations concerning probabilities of sampling individual library sequences are estimates (see Materials and Methods).

Yeast surface display screening

We used yeast surface display for experimental screening. Native Bcl-x_L displayed on the yeast surface bound both Bim and Bad BH3 peptides strongly (Fig. 2A). The yeast library contained many clones that bound to Bad, as expected based on the design. More than 50% of the population of cells expressing Bcl-x_L variants showed binding at 1 μM Bad BH3, and more than 5% showed binding at only 10 nM Bad BH3 (Fig. 2B). The designed library also bound very well to Bim BH3 (even better than Bad BH3), which is not inconsistent with the design, given that most mutations encoded were not predicted to favor Bad over Bim (Table 1).

The designed library was subjected to 6 rounds of screening to identify Bcl-x_L variants that bound Bad BH3 in preference to Bim BH3. This included positive screening for Bad binding, explicit negative screening against Bim binding, and positive screening for Bad binding in the presence of excess unlabeled Bim (see Materials and Methods). The final population showed significantly enhanced specificity, with strong binding to Bad at 1 nM but strong binding to Bim only at 100 nM (Fig. 2C). Analysis of 48 yeast clones from this population gave 21 unique sequences for clones that showed stronger binding to 1 nM Bad compared to 10 nM Bim when tested individually (Fig. 2D, Table S1). The results revealed one Bcl-x_L designed position, 142, at which substitution of Ala to Gly (A142G) was found in all 21 sequences. Another 6 designed positions were occupied by both native and non-native amino acids, whereas 2 positions were occupied primarily with the native amino acid.

Based on these promising results, we designed a second library to identify sequences with further improved specificity. Using the same structural modeling protocol described above, we predicted non-disruptive mutations and specific mutations for 6 additional Bcl-x_L positions (Fig. 1B, Table 2). These new positions were mostly located at the edge of the BH3-binding interface and, not surprisingly, our very relaxed definition of non-disruptive mutations included almost all amino acids. Among the 9 designed positions screened in the previous library, we fixed position 142 as Gly (A142G) and reverted positions 97 and 112 back to native residues Phe and Leu. Non-disruptive mutations at the other 6 positions (101, 104, 108, 126, 129, 130) were redefined as amino acids with significant frequency in the first round of screening (Fig. 2D, Table 2). A total of 12 Bcl-x_L positions were randomized in the new library.

A naïve degenerate codon based library aiming to include all non-disruptive mutations as described above (see Materials and Methods) had a size of 2.0×10^{11}. For this library, the probability of a particular sequence being sampled was only 10^{-4}. Around 1.2% of library sequences encoded protein sequences with designed positions all occupied by non-disruptive mutations. To construct an optimized library, the same ILP library optimization procedure described previously was carried out to select degenerate codons for these positions. To increase efficiency in encoding amino-acid diversity, we introduced a slight modification to allow some designed positions to be encoded by a pair of degenerate codons rather than just one, subject to constraints imposed by the PCR assembly protocol (see Materials and Methods). The resulting library (referred to as library 2 in this study, Table 2) had a size of 1.0×10^{7}, the probability of a particular sequence being sampled was ~0.85, and around 79% of all library DNA sequences encoded protein sequences with designed positions all occupied by non-disruptive mutations. On the other hand, 26 non-disruptive mutations (out of 89) were excluded from consideration for the optimized library.

Significant improvement in specificity was observed after two rounds of screening the newly designed library (compare Fig. 2C and Fig. 3A). Sequencing results revealed strong biases at several designed positions (Fig. 3C, Table S2), as discussed below. We performed 5 additional rounds of screening, and the final population of yeast clones was highly specific for binding Bad over Bim (Fig. 3B), showing good binding to Bad BH3 at 1 nM but much lower binding to Bim BH3 at 1 μM. Only 2 sequences were present in this population, RX1 and RX2 (Table 3). Each contained 9 mutations from native Bcl-x_L, and the mutations were consistent with those observed at high frequency after two rounds of screening, as shown in Fig 3C. Five mutations were shared between RX1 and RX2, including a mutation (F105L) not present in the designed library or sequences identified from library 1. The effect of this mutation was investigated and is analyzed below.

Solution binding

To confirm that the specificity profiles of the selected Bcl-x_L variants seen on the yeast surface could be recapitulated in solution, we prepared and purified recombinant proteins. We chose to characterize RX1 rather than RX2 due to suspicions that a hydrophobic residue at position 96 might be associated with a tendency to oligomerize, based on analysis of designs from earlier rounds of screening (see Materials and Methods). Using circular dichroism spectroscopy, we determined that uncomplexed RX1 melts cooperatively at ~72 °C at 1 μM in phosphate buffer pH 7.4. The re-designed protein is slightly destabilized in comparison to Bcl-x_L, which melts at ~80 °C under the same conditions (Fig. S1).

We used a fluorescence polarization (FP) assay to measure binding of different peptides (Table S3) to Bcl-x_L and RX1 (see Materials and Methods). Direct binding of fluoresceinated Bim vs. Bad BH3 confirmed a strong preference for RX1 binding Bad over Bim (Fig. S2 and S3, Table S4). However, experimental uncertainties due to changes in the anisotropy signal from fluoresceinated Bim BH3 over time led us to develop a competition assay with fluoresceinated Bad BH3 for quantitative comparisons (see Materials and Methods). In this assay, Bcl-x_L interacted very strongly with both Bim BH3 and Bad BH3 28-mer recombinant peptides (Bim-28 and Bad-28, Table S3), with K_i values below 0.1 nM (Fig. 4A, Table S5). In contrast, the fitted K_i values for RX1 interacting with Bim-28 or with Bad-28 were 2.3 μM and 0.26 nM, respectively (Fig. 4B, Table S5). The tightest binding that can be reliably quantified with this assay is ~0.1 nM. Thus, to more reliably measure the increase in specificity from native Bcl-x_L to RX1 in solution, we used shorter BH3 peptides (22 residues), which were expected to be of lower affinity. The shorter peptides (Bim-22 and Bad-22, Table S3) maintained interactions with all of the designed Bcl-x_L positions, based on crystal structures. The fitted K_i values of RX1 for Bim-22 and Bad-22 were > 50 μM and 33 nM, respectively, and the fitted K_i values of Bcl-x_L for Bim-22 and Bad-22 were ~3 nM

and 2 nM, indicating a specificity increase of ~1,000-fold or more for the designed protein (Table S6).

We also evaluated interactions between RX1 and 10 other peptides derived from the BH3 regions of human Bcl-2 family proteins not included in the design/screening experiments (Fig. 5, Table S5). In contrast to Bcl-x_L, which interacts strongly with several other BH3s (Fig. 5A), significant interaction was observed only between RX1 and PUMA (K_i = 35 nM, Fig. 5B). The interaction of RX1 with PUMA was significantly weaker than that with Bad (K_i = 0.26 nM), and also much weaker than that between Bcl-x_L and PUMA (K_i < 0.1 nM). In summary, RX1 displayed global specificity against the other BH3s not included in specificity screening. In cells, we anticipate that RX1 would be less efficient than Bcl-x_L in preventing apoptosis, and that it would not effectively protect cells from pro-death stimuli that signal through BH3-only proteins other than Bad and PUMA.

Dissection of residues important for specificity

To analyze how individual mutations at each designed position contributed to the binding specificity of RX1, we made point mutations in Bcl-x_L and also individually reverted selected residues of RX1 back to the native Bcl-x_L amino acid. We examined binding of these variants to Bim and Bad BH3 peptides. In the context of Bcl-x_L (Fig. 4C, S4, Table S6), mutations of Ser 122 to Ile (denoted S122I), V126A and A142G preferred binding Bad over Bim. Several other mutations led to a preference for binding Bim over Bad, especially Q111G. When examined in the context of RX1 (Fig. 4D, S5, S6, Table S7), reverting each of L105, I122, A126, G142 and A146 individually back to their native residues all caused significant loss of Bad over Bim specificity. Two different constructs of Bad and Bim BH3 peptides were used for this analysis, with highly similar results (compare Fig. 4D and Fig. S6B). The loss in specificity for RX1-A146F was particular interesting as it likely explained why the F146A mutation was present in all specific sequences in library 2 (Fig. 3C), even though Ala at this position did not confer specificity when measured in the context of Bcl-x_L (Fig. 4C). The loss of specificity for RX1-L105F was also interesting, because F105L was not included in the library and actually favored Bim binding over Bad when made in the context of Bcl-x_L (Fig. 4C). Overall, the analysis suggested that although some of the influences of the designed residues were relatively independent of the sequence context, non-additive effects also contributed to the observed specificity.

We measured interactions between RX1 and five mutant Bad peptides judged highly likely to disrupt binding to Bcl-x_L based on the crystal structure of the complex between Bad and Bcl-x_L (Fig. S7, Table S8). These included Bad-G2eL, in which a Gly residue was mutated to Leu to generate unfavorable steric clashes, Bad-L3aD, in which a Leu residue deeply buried in the hydrophobic pocket of Bcl-x_L was mutated to Asp, and Bad-D3fK, in which an Asp residue that is strongly conserved in native BH3 sequences was mutated to Lys. Bad-S3eL and Bad-F4aE were made following similar logic. As expected, interactions between these Bad mutant peptides and both Bcl-x_L and RX1 were significantly destabilized compared to the native Bad peptide, providing evidence that RX1 employed the same binding interface to engage Bad.

Discussion

In this work, we designed a library of Bcl-x_L variants using a new computation-guided framework, and screened the library to obtain proteins that bound Bad BH3 in preference to Bim BH3. Below we discuss the rationale for our approach and provide a retrospective analysis of our library design given the experimental data. We also discuss what we learned from analyzing the sequence determinants of the observed interaction specificity of one of our most specific sequences, RX1.

There were two stages to our library design method. The first focused on generating a list of desirable sequence features and the second on constructing a library that optimized the sampling of those features. This provides a compromise between rational design and random screening, and as methods for structure-guided protein design improve, we anticipate that this type of focused-library approach will become increasingly valuable. Indeed, several other laboratories have been exploring related methods[30–32, 36,37]. Importantly, the second design stage in our framework, involving formal ILP library optimization, can provide significant advantages in design regardless of what method is chosen to prioritize sequence features in stage 1. This makes it a very general and flexible method, as discussed below.

In optimizing the library to enrich for desired sequence features, while maintaining high diversity, we made two decisions. The first was to impose a constraint on the library size, and the second was to mandate the inclusion of native and predicted specific mutations. The constraint on library size enhanced the likelihood that every library sequence would be sampled, which is appropriate if it is possible to pre-define a space that includes solutions to the design problem. Our hypothesis was that such a space could be defined by enforcing the inclusion of native and predicted specificity mutations and maximizing diversity by including as many predicted non-disruptive mutations as possible.

In assessing our results, we found that our methods for designing a library including many Bad binders appeared to work well. More than 50% of library clones showed signal for binding Bad at 1 μM, and more than 5% showed signal for binding Bad at 10 nM (Fig. 2B). This is consistent with our effort to prioritize inclusion of sequences with a high percentage of native residues and mutations denoted as non-disruptive. We were less successful at identifying mutations that gave rise to binding specificity. Using our mutational analysis to compare the library design to experimental results, we found only partial agreement between modeling and experimental screening. For example, mutations A104M (not in RX1) and S122I were predicted and confirmed to be specific. But Q111G and L130I were predicted to be specific and shown not to be. Specific mutations V126A and A142G were included in the library only as predicted non-disruptive mutations. F146A, included as a non-disruptive mutation, did not confer specificity in the Bcl-x_L sequence context but did contribute to specificity in the context of RX1. The many assumptions that we used in our models, e.g. considering point substitutions only in the wild-type context and including only limited backbone flexibility, surely contributed to this. However, modeling specificity is challenging even with state-of-the-art methods[21], which is why we adopted a combined modeling/screening approach to begin with. Similar observations about the limits of computational modeling and the importance of a combined approach have been noted in other library design studies. For example, in attempting to re-engineer the ubiquitin ligase E6AP to bind to a non-natural partner, Guntas et al.[32] found that although structural modeling proved useful for identifying design positions more tolerant of mutations, several aspects of the experimental amino-acid profiles discovered by screening were not anticipated by modeling.

The power of ILP can provide an advantage even if desirable features are defined by structure-gazing or chemical intuition rather than computation in stage 1. In this project, the benefit was particularly evident in the second round of design, where the structure-based computation did little to reduce the sequence space that we wanted to sample because most mutations were classified as non-disruptive. Nevertheless, ILP optimization allowed us to construct a library where 79% of all sequences were predicted to include only non-disruptive mutations, compared to 1.2% of sequences in a larger library designed to span all non-disruptive mutations. The benefit of strategic library optimization increases as the size of the candidate substitution space grows and difficult choices need to be made about what to include/exclude from sampling, because the optimization framework allows a user to systematically explore tradeoffs between library size and content. It is also possible to

introduce a wide range of constraints into the optimization; in other work, we used this framework to include constraints that maintained chemical diversity at variable sites (Dutta, Chen and Keating, under review). Note that it is possible that ILP might not always find a solution in a reasonable amount of time, e.g. if there are a large number of design positions. However, our work suggests this is not likely to be an issue for the type of combinatorial library design problems typically encountered; our optimization calculations were completed within minutes on a standard machine with a quad-core processor.

Given our highly selective Bad binder, RX1, we used mutational analysis to probe the origins of binding specificity (Fig. 4C, 4D). Reverting mutations F105L, S122I, V126A, A142G, and F146A to wild type in RX1 caused significant loss in specificity. Of these, S122I, V126A and A142G also conferred specificity when examined individually in the sequence context of Bcl-x_L. The behaviors of some of the mutations could be rationalized. For example, when modeled on the crystal structure of either Bcl-x_L/Bim or Bcl-x_L/Bad (Fig. S8), the position on the BH3 peptide contacting A142G is occupied by Ser in Bad and by Gly in Bim (Fig. S8B). Bcl-x_L-A142G might therefore confer Bad over Bim specificity by better tolerating the larger residue. Evidence for coupling between design sites came from the observation that some mutations behaved differently depending on the context in which they were made (e.g. F105L, Q111G, F146A). Examining available structures of complexes between Bcl-x_L and Bad or other BH3s does not provide obvious reasons for the non-additive contributions. However, the α2-α3 region of Bcl-x_L has been shown to be conformationally variable among different structures[44], and mutations Y101H, F105L and Q111G are located in this region. Residue 146 is also connected to this region via interaction with Tyr in the 2d position of Bad BH3 (Fig. S8C). Based on these observations, and the fact that our data support a binding mode that is overall similar for Bcl-x_L and RX1, it is probable that the designed protein adopts a locally altered structure in this region that better accommodates Bad over Bim. Efforts to solve a high-resolution structure of the complex between RX1 and Bad are underway, to examine if this is indeed the case. Better ways to describe structural flexibility could prove beneficial for modeling these positions[45]. For example, residues deemed disruptive for binding Bad when modeled onto one backbone structure may be compatible on another structural template. However, extensive sampling of diverse predicted binding modes in design is computationally expensive and not yet routine. If reliable predictions of structural relaxation could be made, including them in library design could increase efficiency. For example, Lippow et al. suggested the importance of explicitly looking at higher-order interactions in the context of enzyme design[46].

Interestingly, design RX1 is not only specific against Bim, but also against all other peptides derived from human BH3-only proteins tested in this study. The only other BH3 peptide that showed significant interaction with RX1 was PUMA, which shares some features with Bad, such as Met at position 3d and Ala at 3e (similar to Ser in Bad in terms of being larger than Gly) (Table S3). Further investigating the mechanism for the observed specificity could be interesting. For example, the A142G mutation, which favors binding Bad over Bim as discussed above, may be responsible for specificity against other BH3 peptides as well, because many of these peptides (Bax, Bak, Bid, Bik and Hrk) have a conserved Gly at position 3e (Fig. S8B). It is also possible that some locally altered conformation of the designed protein better accommodates Bad over not only Bim (as suggested in the previous paragraph), but other BH3 peptides as well. Design examples where specificity was obtained "for free", i.e. without explicit consideration, have been reported previously[47,48]. In the present case, specificity against Bim had to be introduced by competition screening; the original library 1 bound strongly to both Bad and Bim. Elements that destabilized interaction with Bim apparently also destabilized interaction with many other BH3 peptides. Thus, for challenging multi-specificity design problems where it is impractical to screen against all relevant competitors, it may be reasonable to use just one or several competitors that span a

relevant sequence space. In contrast, design studies targeting bZIP coiled coils showed that ignoring some competitors in design calculations could lead to undesired binding[26]. The degree to which negative design is required appears to depend critically on the particular problem being addressed[49].

Materials and Methods

Cloning, protein expression and purification

For yeast surface display, the human Bcl-x_L gene (1–209), followed by a GGGGSG linker and a C-terminal myc tag (EQKLISEEDL), was cloned into the pCTCON2 vector via NheI and BglII sites, with the gene fused in frame to the C-terminus of Aga2p with a $(GGGGS)_3$ linker. PCR amplification of the Bcl-x_L gene was performed using a previously made MBP Bcl-x_L fusion as the template. For recombinant proteins used in the fluorescence polarization assay, the Bcl-x_L gene and variants obtained from screening were cloned into a modified pDEST17 vector via BamHI and XhoI sites. A BamHI cut site was present in the Bcl-x_L gene, and therefore either a BglII or a BclI site, both compatible for ligation to a BamHI cut vector, was included in the primers for PCR amplification. Mutants of either the Bcl-x_L gene or the RX1 design were made using PCR based sited directed mutagenesis followed by blunt end ligation[50], or Quick change (Agilent). Recombinant human BH3 peptides (Bim-28, residues 142–169; Bad-28, residues 104–131; other BH3 28-mers, Table S3), with a C-terminal GG linker followed by a Flag tag (DYKDDDDK), were constructed by gene synthesis. Primers were designed using DNAWorks[51], and a two-step PCR procedure was used for annealing and amplification. The genes were then cloned into a modified pDEST17 vector via BamHI and XhoI sites. Recombinant Bcl-x_L proteins and BH3 peptides (with a His_6 tag) were expressed in *E. coli* RP3098 cells. Cultures were grown at 37 °Cto OD ~0.4–0.9, and expression was induced by addition of 1 mM IPTG. Purification of Bcl-x_L proteins was performed under native conditions using Ni-NTA. An additional step of gel-filtration purification with a HiLoad Superdex™ 75 column (GE) was performed for the mutants and the designed proteins because protein oligomerization was observed for some of them. For selected examples examined, the monomer fractions were stable as monomers upon repeat analysis. Purification of BH3 peptides was performed under denaturing conditions using Ni-NTA and followed by reverse-phase HPLC. Masses were verified by MALDI spectrometry.

Structural modeling

Structural models of Bcl-x_L point mutants interacting with Bim or Bad were generated using Rosetta 3.0[38]. The crystal structure of human Bcl-x_L in complex with Bim (PDB ID: 3FDL)[41] was used to model interactions between Bcl-x_L mutants with Bim and Bad, and that of mouse Bcl-x_L in complex with Bad (PDB ID: 2BZW)[42] was used to model interactions between Bcl-x_L mutants with Bad only. An ensemble of 100 structures was derived separately from each of 3FDL and 2BZW, with fixed native sequence, using the backrub flexible-backbone modeling utility in Rosetta[52]. For the backrub simulation, residues spanning the α2 helix to the α5 helix (86 to 157) of the Bcl-x_L protein and the entire BH3 peptide were picked as “pivot residues”. To generate each individual structure in the ensemble, 10,000 backrub moves were attempted. Each Bcl-x_L mutant interacting with Bim or Bad was then modeled on all members of the structural ensemble using the fixed backbone design mode in Rosetta. Side-chain repacking was allowed for residues at the binding interface (residues within 5 Å of any residue of the interacting partner, defined by Cα-Cα distances), and extra sampling of chi 1 and chi 2 angles for the rotamers was used. A 50-step conjugate-gradient based minimization was performed for each ensemble member, and the Rosetta energy for each minimized structure within the ensemble was obtained. The scoring was based on the default energy weights in Rosetta 3.0. The lowest energy was used

as the score for interaction between the Bcl-x_L mutant being modeled and Bim or Bad, and the difference relative to the score of native Bcl-x_L interacting with Bim or Bad was calculated (ΔE_{Bim} or ΔE_{Bad}). The unbound states were not modeled; the 20 single amino-acid reference energies in Rosetta served as the reference. The score should therefore not be viewed as an attempt to predict changes in binding energies. We instead used it to identify mutations that were predicted not to be well tolerated in the structure of the complex (even if a mutation would have no impact on binding affinity due to a similar destabilizing effect in the unbound state). Because interactions between mutants with Bad were modeled using both 3FDL and 2BZW as templates, two values of ΔE_{Bad} were generated and the lower one was chosen. Residues with ΔE_{Bad} lower than 3 (for the first designed library) or 2 Rosetta energy unit (for the second designed library) were defined as non-disruptive mutations. Residues with ΔE_{Bim} - ΔE_{Bad} greater than 2 (for the first designed library) or 3 Rosetta energy units (for the second designed library) were defined as specificity mutations.

Position Y195 was not subjected to structural modeling as it was missing in the human Bcl-x_L/Bim structure (3FDL). The corresponding position (Y195) was observed in the mouse Bcl-x_L/Bim structure (1PQ1) and formed a hydrogen bond with N102 (position 4b in the BH3 alignment) of Bim (occupied by Val in Bad). Phe was included in library 2 to explore whether removing this hydrogen bond would provide specificity.

Selecting degenerate codons for the designed library

At each designed position i, we defined two quantities for each degenerate codon j: (1) the size, s_{ij}, which is the number of unique trinucleotides within codon j, and (2) n_{ij}, the number of predicted non-disruptive mutations (including the native amino acid) encoded by codon j. The codons were pre-filtered by the following criteria: (1) The native amino acid and all predicted specific mutations at the position must be encoded by the codon, (2) codons encoding only the native amino acid were eliminated, and (3) among the pool of degenerate codons passing the first two criteria, any codon with a larger s_{ij} but a smaller n_{ij} than another within the pool at position i was eliminated. This process was repeated until no further elimination was possible. Optimization of degenerate codon combinations, out of the remaining pool of codons D_i at each designed position i, was performed by solving the following integer linear programming problem:

$$\text{Max} \sum_i \sum_{j \epsilon Di} c_{ij} \log_{10}(n_{ij}), \text{ under} \sum_i \sum_{j \epsilon Di} c_{ij} \log_{10}(S_{ij}) <= 7, \text{ and} \sum_{j \epsilon Di} c_{ij} = 1 \text{ for each position } i$$

Where $c_{ij} = 1$ if codon j was picked at position i, and 0 otherwise. For the codons j picked at each position i, $\Sigma_i \log_{10}(n_{ij}) = \log_{10}(\prod_i n_{ij})$ is the logarithm of the number of unique protein sequences encoded with all designed positions occupied by predicted non-disruptive mutations, and $\Sigma_i \log_{10}(s_{ij}) = \log_{10}(\prod_i s_{ij})$ is the logarithm of the library size (or the number of unique DNA sequences in the library) as described in the text. The above efforts allowed us to formulate the problem as an ILP[39]. Solving the ILP, which is not a trivial process, was performed by the glpsol solver in the GLPK package (GNU MathProg). Perl scripts to write the necessary input files for the glpsol solver and to interpret the output files are available upon request. Note that occasionally multiple codons at one position could have identical statistics and all be optimal under this formulation, and in this case we manually examined the choices and selected one codon.

To design the 2nd library, a “degenerate codon pair” was considered in addition to individual degenerate codons at each designed position. This provided greater flexibility in sampling desired sequence features within a fixed library size. A “degenerate codon pair” was defined as two degenerate codons that share no overlap between the trinucleotides encoded.

Experimentally, a designed position can be constructed as a codon pair by mixing oligonucleotides, or by mixing the PCR products generated by using each of two individual degenerate codons at a site (as done in this study). The size, s_{ij}, for a pair j is the sum of its two codon components, and the experimental mixing ratio is simply the ratio of the sizes for the two codons. The metric n_{ij} is the total number of unique predicted non-disruptive mutations from the two components. The filtering process as described for optimization of the first designed library can be performed for a "codon pair" separately from normal codons. We imposed an additional filtering criterion to exclude any stop codons. For the optimization process, each designed position could be encoded as a single degenerate codon or a pair of codons. However, to avoid an explosion of steps in the library construction protocol, additional constraints were imposed to ensure that no more than one designed position was encoded by a codon pair within the same oligonucleotide in the PCR based assembly procedure (Table S9):

$$\sum_i \sum_{j \in Di} c_{ij}\, p_{ij} <= 1 \text{ for all position } i \text{ randomized on the same oligonucleotide}$$

where p_{ij} is 1 if j is a codon pair at position i and 0 otherwise. The criterion for deciding which designed positions were encoded in the same oligonucleotide is described later. The same ILP optimization procedure as described for the first designed library was then solved with the above constraints to obtain the second designed library.

When designing naïve libraries intended to include all predicted non-disruptive mutations, the degenerate codon with the smallest size among those encoding all desired mutations was picked at each designed position. The size of the library was calculated accordingly.

Construction of combinatorial libraries

The oligonucleotides used to introduce diversity for the two designed libraries are listed in Table S9. PAGE-purified oligonucleotides were ordered from Integrated DNA Technology. Two randomized positions were encoded in the same oligonucleotide if the length of the constant region between them was shorter than 15 nucleotides. The first library was constructed by PCR overlap extension joining two PCR fragments, #1-1 and #1-2. Fragment #1-2 was PCR amplified from PCR fragment #1-2a. PCR amplification for fragment #1-1 introduced diversity for positions 97, 101, 104, 108 and 112 using Bcl-x_L in pCTCON2 as the template. PCR amplification for fragment #1-2a randomized positions 126, 129 and 130 using Bcl-x_L in pCTCON2 as the template, and subsequent PCR amplification for fragment #1-2 randomized position 142 using fragment #1-2a as the template.

The second designed library was made similarly using PCR overlap extension joining two PCR fragments, #2-1 and #2-2. Fragment #2-1 was amplified from fragment #2-1a, and fragment #2-2 was made using PCR overlap extension joining PCR fragments #2-2a and #2-2b. PCR amplification for fragment #2-1a introduced diversity for positions 96, 101, 104, 108 and 111 using Bcl-xL in pCTCON2 as the template. PCR amplification for fragment #2-1 introduced diversity for positions 122, 125, 126, 129 and 130 using fragment #2-1a as the template. PCR amplification for fragment #2-2a introduced diversity for positions 146 and 195 using clone C1 from the first designed library (in the pCTCON2 vector) as the template. Fragment #2-2b was PCR amplified from the fragment #2-2c, which was in turn amplified using clone C1 from the first designed library (in the pCTCON2 vector) as the template.

The final PCR products were co-transformed into yeast with pCTCON2 vector that had been cut with NheI/XhoI, following the procedure of Chao et al.[53] using a BioRad Gene Pulser.

Yeast surface display, flow cytometry analysis and cell sorting

Yeast strain EBY100 and the plasmid for yeast surface display (pCTCON2) were a gift from Dr. K. D. Wittrup (Massachusetts Institute of Technology). Combinatorial DNA libraries were transformed into yeast, and cells were grown/induced following protocols described by Chao et al.[53] Briefly, yeast cells were grown overnight at 30 °C in SDCAA media, and display of the Bcl-x_L protein was induced by switching to SGCAA media for > 12 hr. Induced cells were washed with TBS (50 mM Tris, 100 mM NaCl, pH 8.0), and incubated with different concentrations of Bim-28 or Bad-28 for 1–2 hr in TBS at ~25 °C. Cells were then washed with cold TBS and labeled with primary antibodies (anti-c-*myc* rabbit and anti-His mouse, Sigma) at 1:67 (anti-c-*myc*) or 1:100 (anti-His) dilution for 30 min - 2 hr in BSS (TBS with 1 mg/mL bovine serum albumin) at 4 °C. After washing in cold BSS, cells were labeled with secondary antibodies (PE conjugated anti-rabbit, Sigma and APC conjugated anti-mouse, BD Bioscience) at 1:100 dilution for 30 min – 2 hr in BSS at 4 °C. Cells were then washed again in cold BSS prior to analysis or sorting. The analysis was performed on BD FACSCalibur-HTS1 (BD Bioscience), and the sorting on BD FACSAria (BD Bioscience) or MoFlo (Beckman Coulter). Cells were gated by forward light scattering to avoid the analysis/sorting of clumped cells. Data were analyzed using FlowJo (Tree Star, Inc.).

Below we describe sorting of the first and the second designed libraries. The first designed library was subjected to one round of positive sorting (gating for positive expression and binding) for binding to 10 nM Bad-28, one round of positive sorting for binding to 10 nM Bad-28 in the presence of 1 μM unlabeled Bim-26, two rounds of negative sorting (gating for expression without binding) using 10 nM Bim-28, one round of negative sorting using 100 nM Bim-28, and finally one round of positive sorting for cells binding 10 nM Bad. The second designed library was subjected to two rounds of positive sorting for binding to 1 nM Bad-28 in the presence of 1 μM unlabeled Bim-26, two rounds of positive sorting for binding to 1 nM Bad-28 in the presence of 5 μM Bim-26, one round of negative sorting using 3 μM Bim-28, one round of negative sorting using 5 μM Bim-28, and finally one round of positive sorting for binding to 1 nM Bad-28.

Generation of sequence frequency plots

For the first designed library, 48 individual clones from the final sorted population were examined for binding to 1 nM Bad-28 or 10 nM Bim-28. Twenty-one clones with unique sequences showed stronger binding signal for 1 nM Bad-28 over 10 nM Bim-28, and the sequence frequency plot shown in Fig. 2D was generated from these sequences using WebLogo[54] (Table S1). For the second designed library, 48 individual clones from the population after two rounds of sorting were examined for binding to 1 nM Bad-28 or 500 nM Bim-28. Twenty-eight clones with unique sequences (Table S2) showed stronger binding signal for 1 nM Bad-28 over 500 nM Bim-28 and were used to generate the frequency plot in Fig. 3C.

Circular dichroism spectroscopy

Circular dichroism experiments were performed on AVIV 400 or 202 spectrometers. Thermal unfolding curves were performed using a 1-cm pathlength cuvette and determined by monitoring ellipticity at 222 nm with an averaging time of 30 seconds, an equilibration time of 1.5 minutes, and a scan interval of 2 °C. The melting temperature was estimated as the mid-point of the unfolding transition after manually picking the baselines. The protein concentration was 1 μM, in PBS buffer containing 12.5 mM potassium phosphate (pH 7.4), 150 mM KCl, 0.25 mM EDTA and 1 mM DTT.

Fluorescence polarization binding assays

All unlabeled and FITC-labeled peptides (Table S3) were synthesized by the MIT Biopolymers Facility at the Koch Institute for Integrative Cancer Research unless otherwise noted. A purified 21-mer Bad peptide with a FITC-labeled lysine[55] was ordered from Calbiochem (now EMD Biosciences). The fBim-27 peptide was ordered from CHI scientific. Labeled peptides were ordered with free C-termini, and unlabeled peptides were ordered with free N and C-termini for enhanced solubility (except for the Bim-26 used for yeast surface display, which was synthesized with N-acetylated and C-amidated ends). Synthesized peptides were purified by reverse phase HPLC using a C18 column. All assays were performed in assay buffer (50 mM NaCl, 20 mM Na_2HPO_4, 1 mM EDTA, 0.01% Triton X-100, and 5% DMSO)[55] at ~25 °C. For direct binding assays, the concentration of the fluorescently labeled peptide was fixed at 5 nM. Serial dilution of Bcl-x_L or its variants was performed before mixing with the fluorescently labeled peptide. The reaction was allowed to equilibrate for at least 1 hr. For competition assays, the concentration of the fluorescently labeled peptide was fixed at 15 nM, and Bcl-x_L or its variant was fixed at 50 nM. Serial dilution of the unlabeled peptide was performed, before adding the mixture of fluorescently labeled peptide and the Bcl-x_L protein or its variant. The reaction was allowed to equilibrate for at least 3 hr. Different fluorescently labeled peptides were used for experiments involving different Bcl-x_L protein variants in order to obtain K_i values that could be fit reasonably (Fig S4–S6, Table S4). Non-binding 96 well plates (Corning Incorporated) were used for all assays. Anisotropy measurements were performed on a SpectraMax M5 (Molecular Devices) plate reader. All experiments were done in duplicate. The averaged values and error bars were plotted. Error bars were calculated using the standard deviation formula, but are based on only two independent measurements. Complete models for fitting K_d (or K_i) values for both direct binding and competition experiments were described previously[56], and the K_d (or K_i) values were fit using Matlab (Mathworks). For direct binding, the average and standard deviation of individual K_d values fitted from each of the duplicate experiments are shown in Table S4. For competition assays, each of two duplicate experiments was fit using each of the two protein/labeled-peptide K_i values determined from direct binding. This generated a total of 4 possible K_i values for the competitor, and the average and standard deviation of the highest and lowest values were calculated and are shown in Table S5–S8. Values plotted in Fig 4C, 4D and S6B were normalized by the averaged K_i values of Bcl-x_L and RX1. A lower baseline corresponding to the measured anisotropy value of the free fluorescently labeled peptide in solution was enforced in fitting for competition experiments with competitor peptides failing to reach near-complete inhibition at the highest concentration.

Estimating library sampling probabilities

We calculated the probability that any individual DNA sequence will be sampled when y number of DNA sequence are randomly drawn from a library of size x as $(1-(1/x))^y$. We set x to be 2×10^7, which is roughly the number of yeast transformants obtained for libraries 1 and 2 in this study. It should be noted that these calculations were all greatly simplified. Codon bias in the library, differences in the expression/display of different proteins, selective growth advantages/disadvantages conferred by the expressed sequence, the survival rate of individual yeast clones after each round of FACS screening, and other factors can make the above numbers an under- or over-estimate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the staff at the M.I.T. flow cytometry facility for assistance, and R. Cook and the M.I.T. Biopolymers facility for peptide synthesis. We thank M. Berrondo for help with running structural minimization in Rosetta 3. We thank members of the Keating lab for helpful discussions and the group of S. Bell and K. D. Wittrup for use of equipment. This work was funded by NIH award GM084181.

Abbreviations used

Bcl	B cell lymphoma
BSA	bovine serum albumin
CD	circular dichroism
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
HPLC	high-performance liquid chromatography
ILP	integer linear program
IPTG	isopropyl β-d-1-thiogalactopyranoside
PBS	phosphate-buffered saline
Ni-NTA	nickel nitrilotriacetic acid
T_m	melting temperature
TRIS	tris(hydroxymethyl)aminomethane

References

1. Dreze M, Charloteaux B, Milstein S, Vidalain PO, Yildirim MA, Zhong Q, Svrzikapa N, Romero V, Laloux G, Brasseur R, Vandenhaute J, Boxem M, Cusick ME, Hill DE, Vidal M. ‘Edgetic’ perturbation of a C. elegans BCL2 ortholog. Nat Methods. 2009; 6:843–9. [PubMed: 19855391]
2. Steed PM, Tansey MG, Zalevsky J, Zhukovsky EA, Desjarlais JR, Szymkowski DE, Abbott C, Carmichael D, Chan C, Cherry L, Cheung P, Chirino AJ, Chung HH, Doberstein SK, Eivazi A, Filikov AV, Gao SX, Hubert RS, Hwang M, Hyun L, Kashi S, Kim A, Kim E, Kung J, Martinez SP, Muchhal US, Nguyen DH, O’Brien C, O’Keefe D, Singer K, Vafa O, Vielmetter J, Yoder SC, Dahiyat BI. Inactivation of TNF signaling by rationally designed dominant-negative TNF variants. Science. 2003; 301:1895–8. [PubMed: 14512626]
3. Pawson T, Nash P. Protein-protein interactions define specificity in signal transduction. Genes Dev. 2000; 14:1027–47. [PubMed: 10809663]
4. Rozakis-Adcock M, Fernley R, Wade J, Pawson T, Bowtell D. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. Nature. 1993; 363:83–5. [PubMed: 8479540]
5. Erpel T, Superti-Furga G, Courtneidge SA. Mutational analysis of the Src SH3 domain: the same residues of the ligand binding surface are important for intra- and intermolecular interactions. Embo J. 1995; 14:963–75. [PubMed: 7534229]
6. Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. Nat Rev Mol Cell Biol. 2008; 9:47–59. [PubMed: 18097445]
7. Chen L, Willis SN, Wei A, Smith BJ, Fletcher JI, Hinds MG, Colman PM, Day CL, Adams JM, Huang DC. Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. Mol Cell. 2005; 17:393–403. [PubMed: 15694340]
8. van Delft MF, Wei AH, Mason KD, Vandenberg CJ, Chen L, Czabotar PE, Willis SN, Scott CL, Day CL, Cory S, Adams JM, Roberts AW, Huang DC. The BH3 mimetic ABT-737 targets selective Bcl-2 proteins and efficiently induces apoptosis via Bak/Bax if Mcl-1 is neutralized. Cancer Cell. 2006; 10:389–99. [PubMed: 17097561]

9. Billen LP, Kokoski CL, Lovell JF, Leber B, Andrews DW. Bcl-XL inhibits membrane permeabilization by competing with Bax. PLoS Biol. 2008; 6:e147. [PubMed: 18547146]
10. Lee EF, Czabotar PE, van Delft MF, Michalak EM, Boyle MJ, Willis SN, Puthalakath H, Bouillet P, Colman PM, Huang DC, Fairlie WD. A novel BH3 ligand that selectively targets Mcl-1 reveals that apoptosis can proceed without Mcl-1 degradation. J Cell Biol. 2008; 180:341–55. [PubMed: 18209102]
11. Day CL, Chen L, Richardson SJ, Harrison PJ, Huang DC, Hinds MG. Solution structure of prosurvival Mcl-1 and characterization of its binding by proapoptotic BH3-only ligands. J Biol Chem. 2005; 280:4738–44. [PubMed: 15550399]
12. Goreshnik I, Brock AM, Maly DJ. Biochemical and pharmacological profiling of the pro-survival protein Bcl-xL. Bioorg Med Chem Lett. 21:4951–5. [PubMed: 21807512]
13. Gelinas C, White E. BH3-only proteins in control: specificity regulates MCL-1 and BAK-mediated apoptosis. Genes Dev. 2005; 19:1263–8. [PubMed: 15937216]
14. Letai A, Bassik MC, Walensky LD, Sorcinelli MD, Weiler S, Korsmeyer SJ. Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as prototype cancer therapeutics. Cancer Cell. 2002; 2:183–92. [PubMed: 12242151]
15. Willis SN, Fletcher JI, Kaufmann T, van Delft MF, Chen L, Czabotar PE, Ierino H, Lee EF, Fairlie WD, Bouillet P, Strasser A, Kluck RM, Adams JM, Huang DC. Apoptosis initiated when BH3 ligands engage multiple Bcl-2 homologs, not Bax or Bak. Science. 2007; 315:856–9. [PubMed: 17289999]
16. Leber B, Lin J, Andrews DW. Still embedded together binding to membranes regulates Bcl-2 protein interactions. Oncogene. 2010; 29:5221–30. [PubMed: 20639903]
17. Antonsson B, Conti F, Ciavatta A, Montessuit S, Lewis S, Martinou I, Bernasconi L, Bernard A, Mermod JJ, Mazzei G, Maundrell K, Gambale F, Sadoul R, Martinou JC. Inhibition of Bax channel-forming activity by Bcl-2. Science. 1997; 277:370–2. [PubMed: 9219694]
18. Certo M, Del Gaizo Moore V, Nishino M, Wei G, Korsmeyer S, Armstrong SA, Letai A. Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. Cancer Cell. 2006; 9:351–65. [PubMed: 16697956]
19. Karanicolas J, Kuhlman B. Computational design of affinity and specificity at protein-protein interfaces. Curr Opin Struct Biol. 2009; 19:458–63. [PubMed: 19646858]
20. Jackel C, Kast P, Hilvert D. Protein design by directed evolution. Annu Rev Biophys. 2008; 37:153–73. [PubMed: 18573077]
21. Chen TS, Keating AE. Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods. Protein Sci. 2012; 21:949–63. [PubMed: 22593041]
22. Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. Nat Struct Biol. 2003; 10:45–52. [PubMed: 12459719]
23. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein-protein interaction specificity. Nat Struct Mol Biol. 2004; 11:371–9. [PubMed: 15034550]
24. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009; 458:859–64. [PubMed: 19370028]
25. Leaver-Fay A, Jacak R, Stranges PB, Kuhlman B. A generic program for multistate protein design. PLoS One. 2011; 6:e20937. [PubMed: 21754981]
26. Grigoryan G, Keating AE. Structure-based prediction of bZIP partnering specificity. J Mol Biol. 2006; 355:1125–42. [PubMed: 16359704]
27. Sammond DW, Eletr ZM, Purbeck C, Kuhlman B. Computational design of second-site suppressor mutations at protein-protein interfaces. Proteins. 2010; 78:1055–65. [PubMed: 19899154]
28. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH. Protein building blocks preserved by recombination. Nat Struct Biol. 2002; 9:553–8. [PubMed: 12042875]
29. Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI. Combining computational and experimental screening for rapid optimization of protein properties. Proc Natl Acad Sci U S A. 2002; 99:15926–31. [PubMed: 12446841]

30. Treynor TP, Vizcarra CL, Nedelcu D, Mayo SL. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. Proc Natl Acad Sci U S A. 2007; 104:48–53. [PubMed: 17179210]
31. Allen BD, Nisthal A, Mayo SL. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. Proc Natl Acad Sci U S A. 2010; 107:19838–43. [PubMed: 21045132]
32. Guntas G, Purbeck C, Kuhlman B. Engineering a protein-protein interface using a computationally designed library. Proc Natl Acad Sci U S A. 107:19296–301. [PubMed: 20974935]
33. Pantazes RJ, Saraf MC, Maranas CD. Optimal protein library design using recombination or point mutations based on sequence-based scoring functions. Protein Eng Des Sel. 2007; 20:361–73. [PubMed: 17686879]
34. Parker AS, Griswold KE, Bailey-Kellogg C. Optimization of combinatorial mutagenesis. J Comput Biol. 2011; 18:1743–56. [PubMed: 21923411]
35. Wang W, Saven JG. Designing gene libraries from protein profiles for combinatorial protein experiments. Nucleic Acids Res. 2002; 30:e120. [PubMed: 12409479]
36. Chen MM, Snow CD, Vizcarra CL, Mayo SL, Arnold FH. Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. Protein Eng Des Sel. 2012; 25:171–8. [PubMed: 22334757]
37. Azoitei ML, Correia BE, Ban YE, Carrico C, Kalyuzhniy O, Chen L, Schroeter A, Huang PS, McLellan JS, Kwong PD, Baker D, Strong RK, Schief WR. Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. Science. 2011; 334:373–6. [PubMed: 22021856]
38. Das R, Baker D. Macromolecular modeling with rosetta. Annu Rev Biochem. 2008; 77:363–82. [PubMed: 18410248]
39. Cormen, TH.; Leiserson, CE.; Rivest, RL.; Stein, C. Introduction to Algorithms. Vol. Ch 29. MIT Press; 2001. p. 819
40. Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JM, Yeung YA, Cochran JR, Heinzelman P, Colby D, Swers J, Graff C, Wiley HS, Wittrup KD. Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. Nat Biotechnol. 2003; 21:163–70. [PubMed: 12536217]
41. Lee EF, Sadowsky JD, Smith BJ, Czabotar PE, Peterson-Kaufman KJ, Colman PM, Gellman SH, Fairlie WD. High-resolution structural characterization of a helical alpha/beta-peptide foldamer bound to the anti-apoptotic protein Bcl-xL. Angew Chem Int Ed Engl. 2009; 48:4318–22. [PubMed: 19229915]
42. Lee KH, Han WD, Kim KJ, Oh BH. unpublished.
43. Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng Des Sel. 2009; 22:553–60. [PubMed: 19561092]
44. Feng W, Huang S, Wu H, Zhang M. 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. 2007; 372:223–35. [PubMed: 17659302]
45. Mandell DJ, Kortemme T. Backbone flexibility in computational protein design. Curr Opin Biotechnol. 2009; 20:420–8. [PubMed: 19709874]
46. Lippow SM, Moon TS, Basu S, Yoon SH, Li X, Chapman BA, Robison K, Lipovsek D, Prather KL. Engineering enzyme specificity using computational design of a defined-sequence library. Chem Biol. 17:1306–15. [PubMed: 21168766]
47. Potapov V, Reichmann D, Abramovich R, Filchtinski D, Zohar N, Ben Halevy D, Edelman M, Sobolev V, Schreiber G. Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. J Mol Biol. 2008; 384:109–19. [PubMed: 18804117]
48. Yosef E, Politi R, Choi MH, Shifman JM. Computational design of calmodulin mutants with up to 900-fold increase in binding specificity. J Mol Biol. 2009; 385:1470–80. [PubMed: 18845160]
49. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. Curr Opin Struct Biol. 2011; 21:50–61. [PubMed: 21071205]

50. Fisher CL, Pei GK. Modification of a PCR-based site-directed mutagenesis method. Biotechniques. 1997; 23:570–1. 574. [PubMed: 9343663]

51. Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. 2002; 30:e43. [PubMed: 12000848]

52. Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. J Mol Biol. 2008; 380:742–56. [PubMed: 18547585]

53. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nat Protoc. 2006; 1:755–68. [PubMed: 17406305]

54. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–90. [PubMed: 15173120]

55. Zhang H, Nimmer P, Rosenberg SH, Ng SC, Joseph M. Development of a high-throughput fluorescence polarization assay for Bcl-x(L). Anal Biochem. 2002; 307:70–5. [PubMed: 12137781]

56. Fu X, Apgar JR, Keating AE. Modeling backbone flexibility to achieve sequence diversity: the design of novel alpha-helical ligands for Bcl-xL. J Mol Biol. 2007; 371:1099–117. [PubMed: 17597151]

Highlights

- Proteins with re-designed interaction profiles can be useful for many purposes
- Bcl-x_L binds many partners containing BH3 motifs to regulate cell death
- A guided library screening approach was used to re-design Bcl-x_L interactions
- Re-designed Bcl-x_L was highly specific for Bad BH3 over other BH3 motifs
- Focused library design can be adapted for use on many problems

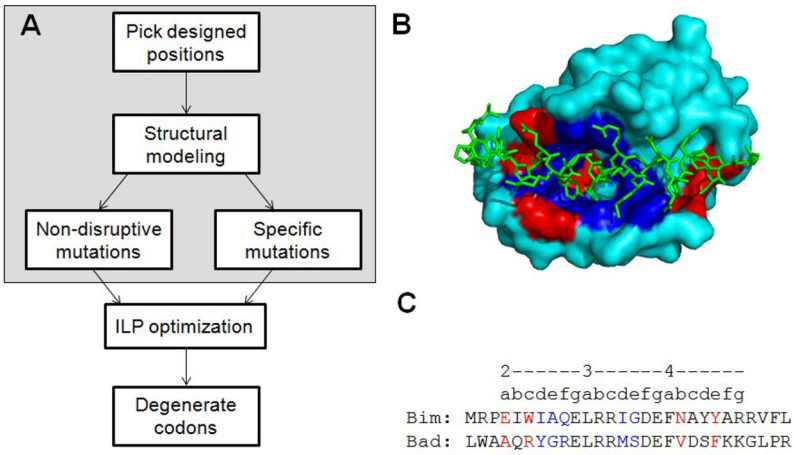

Figure 1.

Library design. (A) The computational library design protocol. In stage 1, non-disruptive and specific mutations were predicted at every design position using structural modeling with Rosetta; in stage 2, ILP optimization was performed to select degenerate codons that maximized the inclusion of predicted non-disruptive mutations while enforcing the inclusion of predicted specific mutations, under a library size constraint. (B) The interface between Bcl-x_L and a Bad BH3 peptide (PDB ID: 2BZW), with Bcl-x_L in cyan and the peptide in green sticks (PyMol, Delano Scientific). Designed positions for the first library are in blue and those for the second library are in red. (C) Sequence alignment of the BH3 regions of Bim (residues 142 to 169) and Bad (residues 104 to 131). Non-identical peptide positions used to select Bcl-x_L sites for library design are colored in blue (for the first library) or red (for the second library). BH3 positions were numbered using a heptad convention as shown at the top.

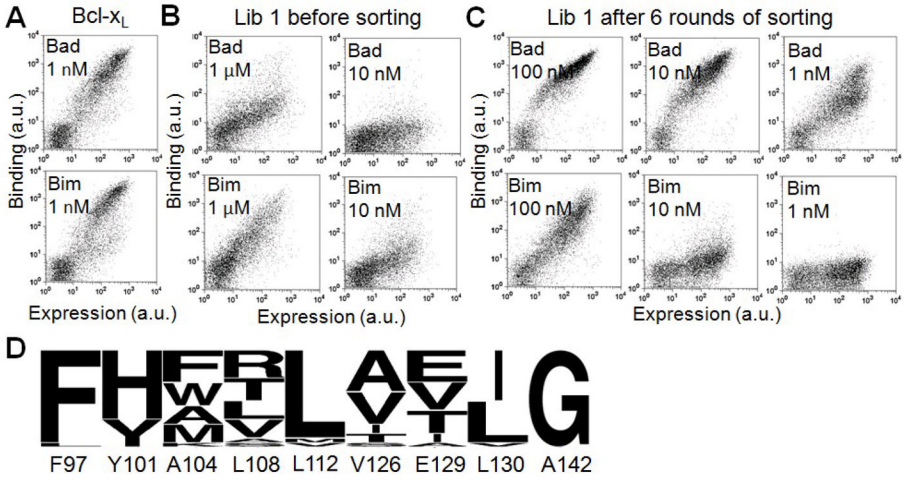

Figure 2.

Screening results for the first designed library. (A) Flow cytometry plots showing binding of native Bcl-x_L to Bad (top) and Bim (bottom) at 1 nM. Expression and binding signals are plotted on the horizontal and vertical axes, respectively. (B) Flow cytometry plots showing binding of the first designed library to Bad and Bim at the concentrations indicated. (C) Flow cytometry plots showing binding of the population obtained from 6 rounds of sorting of the first designed library to Bad and Bim at the concentrations indicated. (D) Sequence frequency plot for 21 unique sequences identified as specific for Bad over Bim BH3 in the first designed library (Table S1), with the native Bcl-x_L residue shown below each column. Plots were generated using WebLogo[54].

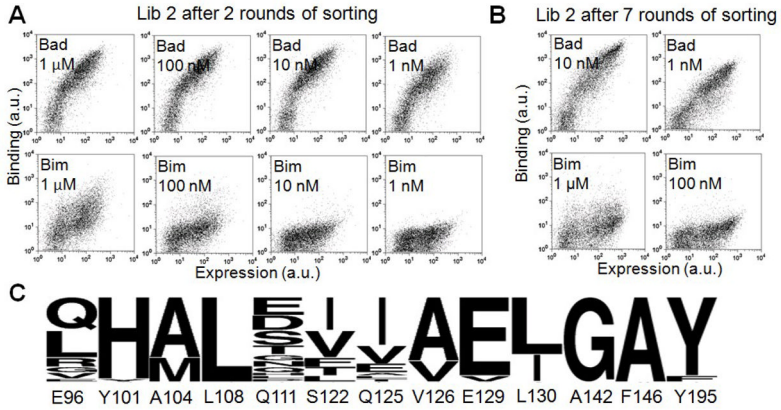

Figure 3.

Screening results for the second designed library. (A) Flow cytometry plots showing binding of the population obtained from 2 rounds of sorting of the second designed library to Bad (top) and Bim (bottom) at the concentrations indicated. Expression and binding signals are plotted on the horizontal and vertical axes, respectively. (B) Flow cytometry plots showing binding of the population obtained from 7 rounds of sorting of the second designed library to Bad and Bim at the concentrations indicated. (C) Sequence frequency plot for 28 unique sequences (Table S2) identified as specific for Bad over Bim BH3 from the second designed library after two rounds of screening. Note that position 142 was not varied in this library but is a change from wild-type Bcl-x_L.

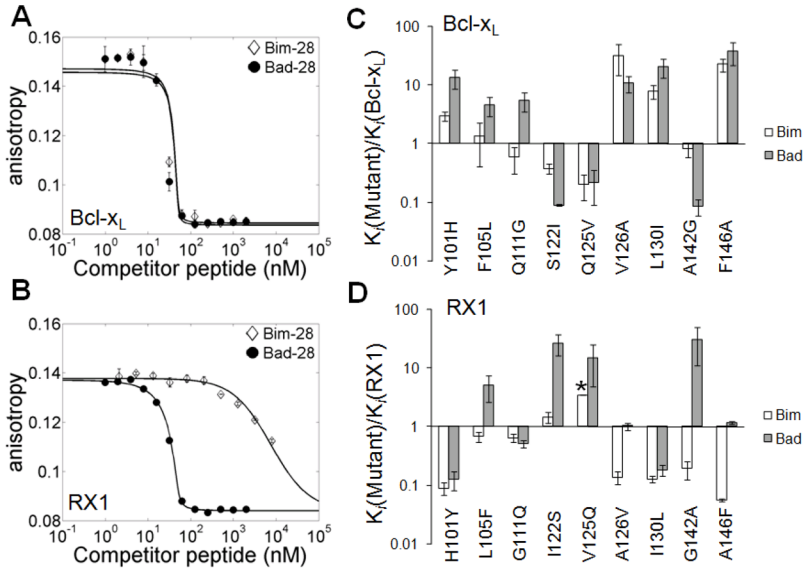

Figure 4.

Fluorescence polarization experiments characterizing Bcl-x_L and its variants binding to BH3 peptides derived from Bim or Bad. (A, B) Competition of Bim-28 or Bad-28 with fBad-23 for binding to native Bcl-x_L (A) or RX1 (B). The average values of two independent measurements are plotted as a function of competitor peptide concentration; error bars are the standard deviations for 2 values. (C) The influence of point mutations in Bcl-x_L on binding to Bim or Bad BH3. Ratios of the K_i values for Bcl-x_L point mutants binding Bim-22 (white) or Bad-22 (gray) to the K_i values for Bcl-x_L binding the same BH3 peptide are shown. (D) Mutational effects in RX1, as in (C). Peptides used were Bim-28 (white) and Bad-28 (for RX1 mutants L105F, I122S, V125Q, G142A) or Bad-22 (for RX1 mutants H101Y, G111Q, A126V, I130L, A146F) (gray). The ratio for RX1-V125Q binding to Bim-28 is marked with an asterisk (*), because only a lower bound on the K_i value could be determined (Table S7). Binding conditions are described in Materials and Methods, and peptide sequences are given in Table S3.

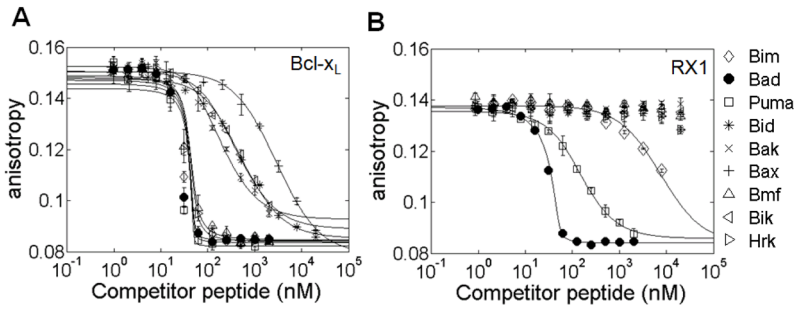

Figure 5.

Fluorescence polarization experiments characterizing Bcl-x_L and RX1 binding to 10 BH3 peptides from human proteins. (A, B) Competition of BH3s with fBad-23 for binding to native Bcl-x_L (A) or RX1 (B). The average values as well as the standard deviations of two independent measurements are plotted as a function of competitor peptide concentration. In (B), competition curves are shown only for competitor peptides that bound Bcl-x_L significantly. K_i values are listed in Table S5.

Table 1

The first designed library

position	amino acids modeled[a]	amino acids encoded[b]
F97	AFGILMV	FILM (WTK)
Y101	AFGILMTVY	FHLY (YWT)
A104	AFGILMSTVY	ACDEFGIKLMNRSTVWYZ (DNK)
L108	AFGILMV	AGILPRSTV (VBT)
L112	AFGILMV	LMV (DTG)
V126	AFGILMV	AGIMRSTV (RBK)
E129	AEITV	AEIKTV (RHA)
L130	AFGILMV	ILV (VTC)
A142	AGSTV	AGST (RSC)

[a]Amino acids modeled at each position. Underlined amino acids were predicted to be non-disruptive, shaded ones were predicted to be specific.

[b]Amino acids included in the designed library (encoded by the degenerate codon in parentheses). A stop codon is indicated by “Z”. The IUB codes for mixtures of nucleotides were adopted when representing the degenerate codons (M: A/C, R: A/G, W: A/T, S: C/G, Y: C/T, K: G/T, V: A/C/G, H: A/C/T, D: A/G/T, B: C/G/T, N: A/G/C/T)

Table 2

The second designed library[a]

position	amino acids modeled	amino acids encoded
E96	ADEFGHIKLMNQRSTVY	EGLQRV (SDA)
Y101	HY	HY (YAT)
A104	AFMW	AM (GCA and ATG)
L108	LRTV	LV (STG)
Q111	ADEFGHIKLMNQRSTVY	ADEGHIKLMNPQRSTV (VNK)
S122	ADEFGHIKLMNQRSTVY	ADFHILNPSTVY (NHC)
Q125	ADEFGHIKLMNQRSTVY	ADEFGILNQRSTVY (DHT and SDA)
V126	AV	AV (GYA)
E129	ETV	EV (GWA)
L130	LI	LI (MTC)
F146	AFGILMV	AFL (TTS and GCT)
Y195	FY	FY (TWC)

[a]See descriptions for Table 1. All sequences also had mutation A142G, relative to native Bcl-x_L.

Table 3

Sequences of clones from the final sorted population of the 2nd designed library

	E96	Y101[a]	F105[a]	Q111	S122I	Q125	V126	L130	A142	F146
RX1	E	H	L	G	I	V	A	I	G	A
RX2	L	H	L	D	T	E	A	L	G	A

[a]Y101H was not included in the modeling but was included in library 1 due to codon choice. Position 105 was not included in the designed library. All other mutations were predicted to be non-disruptive.