Systematic Analysis of Helical Protein Interfaces Reveals Targets for Synthetic Inhibitors

Andrea L. Jochim and Paramjit S. Arora*

Department of Chemistry, New York University, New York, New York 10003

Most modern pharmaceuticals are small molecules that target molecular pockets in enzymes or protein receptors, but in general they fail to achieve sufficient specificity and affinity to target extended and often flat interfaces common to protein−protein interactions (PPI). However, successful examples of small molecule PPI inhibitors are emerging (1). Analysis suggests that although protein interfaces are large, often a small subset of the residues contributes significantly to the free energy of binding (2, 3). Small molecules that reproduce the functionality of these "hot spot" residues have the potential to inhibit the relevant interfaces.

Alanine scanning mutagenesis offers a powerful approach for identifying hot spot residues (4). For example, in the well-studied p53/HDM2 interaction, three residues (Phe19, Trp23, and Leu26) from a helix in the p53 activation domain reside in a deep hydrophobic groove (Figure 1, panel a) (5). Mutation of any of these residues to alanine leads to a significant (>2 kcal mol^{-1}) decrease in the stability of the resulting complex (6). Similar alanine scanning results are obtained with pro-apoptotic partners of the anti-apoptotic protein Bcl-xL (Figure 1, panel b) (7). The complex between transcription factor p53 and its regulator HDM2 is inhibited by nutlins (Figure 1, panel c) (8, 9), and there are highly potent small molecule antagonists, including ABT-737 and A-385358, of the interactions be-

tween Bcl-xL and BH3 domains (Figure 1, panel d) (10, 11). We conjectured that these interactions can be inhibited with nanomolar affinity by small molecules because the critical residues lie within a small radius of each other on one of the partner proteins, allowing their arrangement on a low molecular weight scaffold. For instance, the two chlorobenzene groups in nutlin-3 span 6 Å (Figure 1, panel e) and occupy the binding pockets of the key tryptophan and leucine residues from the p53 helix (8). Similarly A-385358 targets same key pockets on Bcl-xL as the helical BH3 domains (12). Using these two examples of successfully inhibited protein−protein interactions as a guide, we surveyed the Protein Data Bank (PDB) (13) to identify protein−protein interactions as likely targets for small molecule inhibitors. Although a number of studies have focused on predicting the physico-chemical properties of small molecule protein−protein interaction inhibitors (14−17), we sought to develop a method to gauge the "inhibitability" of protein complexes.

Here we focus on protein complexes that feature α-helices at the interfaces. α-Helices constitute the largest class of protein secondary structure and mediate many protein interactions (18, 19). Helices located within the protein core are vital for the overall stability of protein tertiary structure, whereas exposed α-helices on protein surfaces constitute central bioactive regions for the

ABSTRACT Synthetic inhibitors of protein−protein interactions are being discovered despite the inherent challenge in targeting large contact surfaces with small molecules. An analysis of available examples identifies common features of complexes that make them tractable for small molecules. We deduced that relative disposition and energetic contributions of "hot spot" residues provide a predictive scale for the potential of protein−protein interactions to be inhibited by small molecules. On the basis of this model, we analyzed the full set of helical protein interfaces in the Protein Data Bank to identify those that are potentially suitable candidates for synthetic ligands.

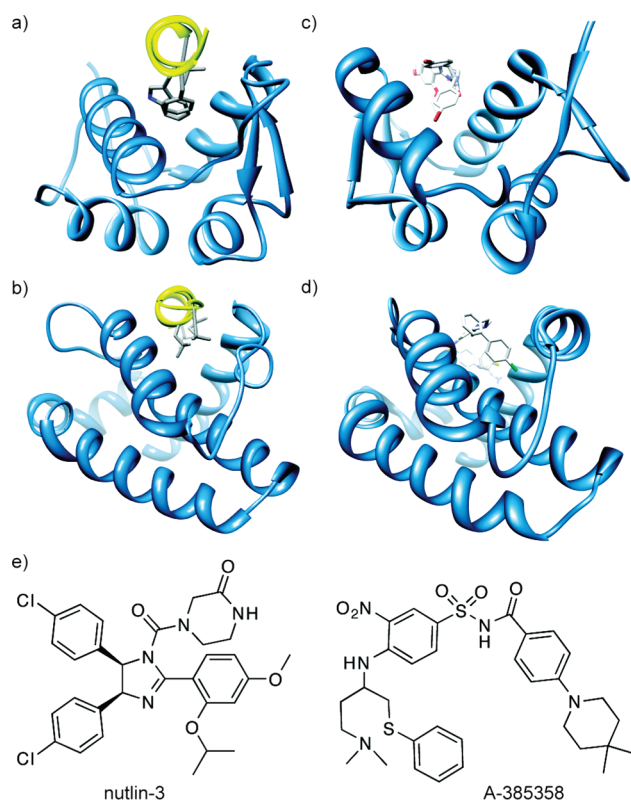*Corresponding author, arora@nyu.edu.

Figure 1. a) p53/HDM2 interaction (PDB code: 1YCR). A helix in the p53 activation domain resides in a deep hydrophobic groove. b) Pro-apoptotic protein partner Bak bound to the antiapoptotic protein Bcl-xL (PDB code: 1BXL). c) Nutlin-3 binds to HDM2 in the same hydrophobic groove occupied by the p53 helix (PDB code: 1rv1). d) ABT-785358 targets Bcl-xL at the site of its pro-apoptotic binding partners (PDB code: 2o22). e) Structures of nutlin-3 and A-385358.

recognition of numerous proteins, DNAs, and RNAs. Helix mimetics have emerged as a highly effective class of PPI inhibitors (*20–26*). A catalog of targetable helical interfaces should significantly enhance the utility of these helix mimetics.

We began by identifying the full set of α-helical interfaces in the PDB (Figure 2). The PDB (version 08/04/2009) was queried for structures containing more than one protein entity (Supporting Information) (*18*). This query extracted 9,339 complexes. We clustered these complexes according to sequence similarity of all protein chains in each complex using the CD-HIT (*27*) sequence alignment program at a 95% similar-

ity threshold. This yielded a data set of 4,143 unique protein complexes. For each <4 Å resolution structure, we extracted potential chain partners belonging to separate molecules as specified in the PDB file. Identification of secondary structure, interfacial residues, and hot spot residues was accomplished using the Rosetta suite of programs (*28–30*). Rosetta determines secondary structure by calculating the ϕ and φ angles of the protein backbone. We define a helical segment as one that contains at least four contiguous residues with ϕ and φ angles characteristic of an α- or the closely related 3_{10}-helix (Supporting Information) (*18*). An interfacial

residue is defined as a residue that has at least one atom within a 5 Å radius of an atom belonging to a binding partner in the protein complex. Hot spot residues were predicted using a computational alanine scan (*29, 30*). Hot spot residues were defined as residues that upon mutation to alanine are predicted to decrease the binding energy by a threshold value $\Delta\Delta G_{bind} \geq 1.0$ kcal mol^{-1}, as measured in Rosetta energy units. Our method identified 2,561 PDB entries possessing helix interfaces in protein−protein (HIPP) interactions and suggests that roughly 62% of the protein complexes in the PDB feature helical interfaces.

After analyzing the energetic contributions and spatial arrangement of the hot spot residues in the helix of p53 in the p53/HDM2 interaction and the BH3 helix of Bak in the Bcl-xL/Bak interaction, we found common features that may provide insight into the reason HDM2 and Bcl-xL are tractable targets for inhibition by small molecules. For both complexes the calculated average $\Delta\Delta G_{bind}$ of the hot spot residues in the helix of the protein partner (p53 in the p53/HDM2 complex and Bak in the Bcl-xL/Bak complex) is greater than 2 kcal mol^{-1} and the radius between the hot spot residues in the helix of the protein partners was on the order of 7 Å. The HIPP data set was analyzed to screen for protein receptors that possessed these same features. Using these criteria we placed the HIPP interactions into three broad categories: (1) receptors that contain a *cleft* for helix binding (Figure 3, panel a), as in the p53/HDM2 complex, where at least two nearby residues contribute strongly to binding; (2) *extended* interfaces that require multiple contacts from two to five turn helices featuring two or more residues that contribute strongly to binding (Figure 3, panel b); and (3) receptors with clefts and extended interfaces characterized by weaker interactions (*31*). We defined hot spot residues as strong or weak contributors based on the change in free energy ($\Delta\Delta G_{avg}$) when these residues on a given helix are mutated to alanine, with a $\Delta\Delta G_{avg}$ cutoff of 2 kcal mol^{-1}. Receptors with clefts are targeted by helices with two or more hot spot residues within a 7 Å radius, while the extended interfaces category features a distribution of hot spot residues over a larger distance of 7−30 Å (Figure 3). Category 2 consists of interfaces where the helical segment spans 20 residues or roughly five helical turns; longer sequences were placed in category 3. We then calculated the proportion of hot spots residing on the helix versus the rest of the chain to determine if a simple mimic of an interfacial helix can inhibit a large interface.
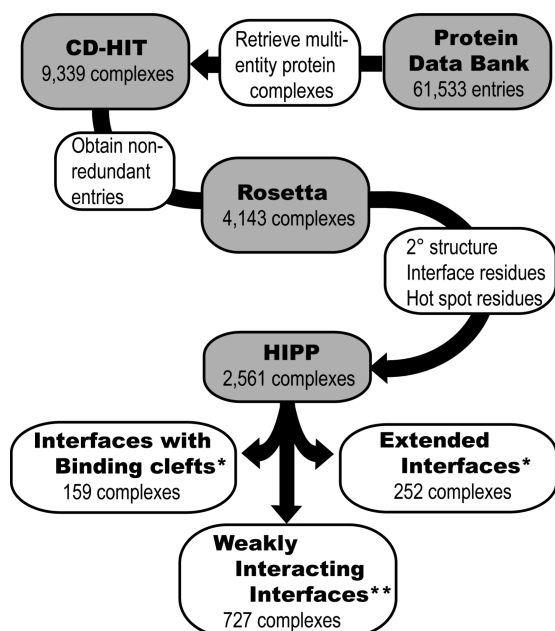
Figure 2. Evaluation of structures from the Protein Data Bank to identify and assess helical interfaces in protein–protein (HIPP) interactions. The helical interfaces were segregated on the basis of binding interfaces and computational alanine scanning mutagenesis analysis. $*\Delta\Delta G_{avg} \geq 2$ kcal mol^{-1}; $**\Delta\Delta G_{avg} = 1-2$ kcal mol^{-1}.

This analysis is complicated by the fact that a number of the complexes in the PDB do not have full-length proteins, and in fact a number of the relevant complexes consist of only a truncated helical domain.

This procedure provided 159 complexes that we predict are targets for small molecules and another 252 interfaces that may be inhibited by helix mimetics. The remainder of complexes that did not meet the criteria of the first two categories but were not eliminated from consideration for other reasons (Supporting Information) feature weaker interactions between the candidate helix and the protein receptor. Table 1 illustrates examples of interfaces with binding clefts and extended interfaces. Full lists of complexes that fall in each of the three aforementioned categories, along with the sequence information for each PDB entry, are included in the Supporting Information

along with a list of the redundant HIPP interactions. It is interesting to note that proteins from the same family do not necessarily fall into the same category. For example, when Bcl-xL is co-crystallized with the Bak peptide, this HIPP interaction is classified as a binding cleft target. On the other hand, the closely related Bad peptide, when co-crystallized with the same Bcl-xL protein, is classified as an extended HIPP interaction. These results reflect the distinct binding profiles observed with Bcl-2 family proteins and BH3 domains (32).

We propose that receptors with binding clefts are candidates for high-throughput screening efforts with small molecule libraries currently available to chemical biologists (1, 33). It is expected that several of these targets would not be of interest for inhibitor design owing to their biological function and other criteria (15). The overall numbers of targets in each category are expected to increase as the PDB is further populated with helical protein–protein interactions. The aim of this inquiry is to devise an algorithm to gauge the "inhibitability" of protein–protein interactions. Existing examples of potent small molecules disrupting protein–protein interfaces as predicted in Category 1 are

listed in Supplementary Table S1 (14, 15). Our analysis suggests that stabilized helices and other structured oligomers are potentially better candidates for targeting extended interfaces (Category 2) (21, 22), although these helix mimetics can also effectively modulate Category 1 interactions (24, 25, 34). It is likely that direct mimics of helices from Category 3 interfaces, where the hot spot residues do not contribute strongly, will not target the cognate protein receptor with high affinity, although utilization of non-natural residues or use of covalent cross-links with protein receptor could overcome the inherent weak affinities at these interfaces.
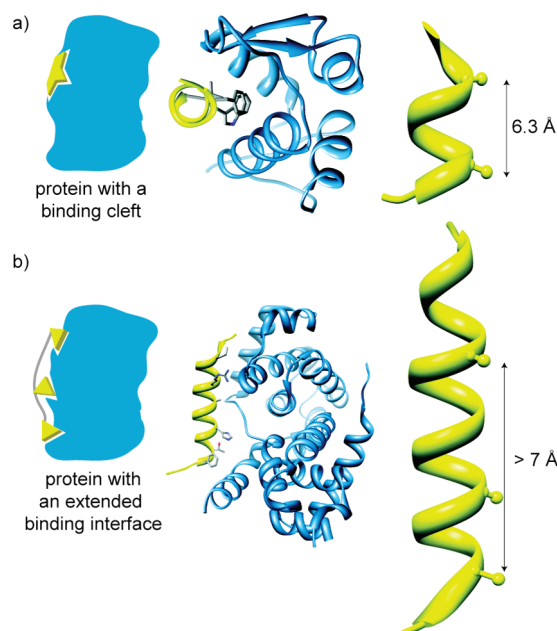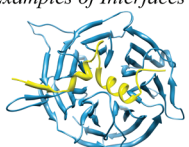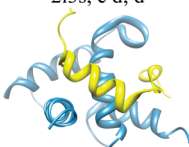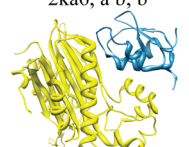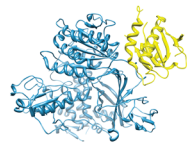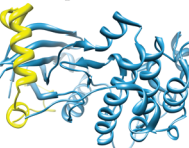

Figure 3. Helical interfaces. We have divided helical protein–protein interactions between those that feature clefts for binding (a) and those with extended interfaces (b). The p53/MDM2 (PDB code: 1YCR) (a) and cyclin-dependent kinase6/D-type viral cyclin (PDB code: 1G3N) (b) complexes are representative examples of binding cleft and extended interfaces, respectively. The distance between flanking hot spot residues in the helix of the protein partner of a binding cleft target spans a radius of 7 Å or less (a) and greater than 7 Å but less than 30 Å for an extended interface target.

TABLE 1. Predicted helical protein—protein interactions as targets for synthetic ligands

PDB code, interface chains; candidate helix[a]	$\Delta\Delta G_{\text{avg/helix}}$ (kcal/mol)[b]	Helix contribution[c]	Hot spot residue helix position[d]	Chain name/Function[e]
Examples of Interfaces with Binding Clefts				
2i3s, c d; d	2.8	55%	i $i+1$ $i+3$	c: Cell cycle arrest protein d: Checkpoint serine/threonine-protein kinase Function: cell cycle
2ka6, a b; b	2.6	54%	i $i+1$ $i+3$ $i+4$	a: CREB-binding protein b: Signal transducer and activator of transcription 1-α/β Function: transcription and gene regulation
2pop, c d; c	2.3	100%	i $i+1$ $i+4$	c: Mitogen-activated protein kinase kinase kinase 7-interacting protein 1 d: Baculoviral IAP repeat-containing protein Function: signaling
Examples of Extended Interfaces				
2nup, b c; c	2.5	72%	i $i+4$ $i+5$ $i+8$	b: Protein transport protein Sec24A c: Vesicle-trafficking protein SEC22b Function: protein transport
2q0o, a c; c	2.4	66%	i $i+10$ $i+13$ $i+14$	a: Probable transcriptional activator protein traR c: Probable transcriptional repressor tram Function: transcription and gene regulation
2vgo, a d; d	2.5	24%	i $i+8$	a: Serine/threonine-protein kinase 12-A d: Inner centromere protein A Function: enzyme (transferase)

[a]Chains in the complex featuring a helix at the interface; candidate helix to be mimicked is part of the indicated chain. [b]$\Delta\Delta G_{\text{avg/helix}}$ is derived from Rosetta computational alanine mutagenesis studies and indicates the average free energy penalty for mutating two or more key residues at the interface to alanine [c]Helix contribution refers to the proportion of key contact residues positioned on the candidate helix as compared to the chain (see text for a detailed explanation). [d]Relative positioning of the hot spot residues on a helix. [e]Description of chains featuring the helical interface and cellular function.

We sorted the helical interactions in the HIPP data set according to function as defined in the PDB (Supplementary Figure S2). Although some HIPP interactions could fall into more than one functional group, we limited each HIPP interaction to one group. Helical interfaces are involved in a broad range of functions from enzymatic activity to gene regulation. Interfaces with binding clefts are involved in a comparable range of functions as extended interfaces. We anticipate that the methods and results described here will guide development of next generation protein–protein interaction inhibitors.

Supporting Information Available: This material is available free of charge *via* the Internet at http://pubs.acs.org.

REFERENCES

1. Wells, J. A., and McClendon, C. L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces, *Nature 450*, 1001–1009.

2. Bogan, A. A., and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces, *J. Mol. Biol. 280*, 1–9.

3. Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007) Hot spots—a review of the protein-protein interface determinant amino-acid residues, *Proteins 68*, 803–812.

4. Cunningham, B. C., and Wells, J. A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis, *Science 244*, 1081–1085.

5. Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., and Pavletich, N. P. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain, *Science 274*, 948–953.

6. Picksley, S. M., Vojtesek, B., Sparks, A., and Lane, D. P. (1994) Immunochemical analysis of the interaction of p53 with MDM2; fine mapping of the MDM2 binding site on p53 using synthetic peptides, *Oncogene 9*, 2523–2529.

7. 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., Thompson, C. B., and Fesik, S. W. (1997) Structure of Bcl-x(L)-Bak peptide complex: Recognition between regulators of apoptosis, *Science 275*, 983–986.

8. Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., and Liu, E. A. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, *Science 303*, 844–848.

9. Shangary, S., and Wang, S. (2009) Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy, *Annu. Rev. Pharmacol. Toxicol. 49*, 223–241.

10. 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.

11. Shoemaker, A. R., Oleksijew, A., Bauch, J., Belli, B. A., Borre, T., Bruncko, M., Deckwirth, T., Frost, D. J., Jarvis, K., Joseph, M. K., Marsh, K., McClellan, W., Nellans, H., Ng, S., Nimmer, P., O'Connor, J. M., Oltersdorf, T., Qing, W., Shen, W., Stavropoulos, J., Tahir, S. K., Wang, B., Warner, R., Zhang, H., Fesik, S. W., Rosenberg, S. H., and Elmore, S. W. (2006) A small-molecule inhibitor of Bcl-XL potentiates the activity of cytotoxic drugs in vitro and in vivo, *Cancer Res. 66*, 8731–8739.

12. Bruncko, M., Oost, T. K., Belli, B. A., Ding, H., Joseph, M. K., Kunzer, A., Martineau, D., McClellan, W. J., Mitten, M., Ng, S. C., Nimmer, P. M., Oltersdorf, T., Park, C. M., Petros, A. M., Shoemaker, A. R., Song, X., Wang, X., Wendt, M. D., Zhang, H., Fesik, S. W., Rosenberg, S. H., and Elmore, S. W. (2007) Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL, *J. Med. Chem. 50*, 641–662.

13. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235–242.

14. Higueruelo, A. P., Schreyer, A., Bickerton, G. R., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database, *Chem. Biol. Drug Des. 74*, 457–467.

15. Bourgeas, R., Basse, M. J., Morelli, X., and Roche, P. (2010) Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database, *PLoS One 5*, e9598.

16. Davis, F. P., and Sali, A. (2010) The overlap of small molecule and protein binding sites within families of protein structures, *PLoS Comput. Biol. 6*, e1000668.

17. Reynes, C., Host, H., Camproux, A. C., Laconde, G., Leroux, F., Mazars, A., Deprez, B., Fahraeus, R., Villoutreix, B. O., and Sperandio, O. (2010) Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods, *PLoS Comput. Biol. 6*, e1000695.

18. Jochim, A. L., and Arora, P. S. (2009) Assessment of helical interfaces in protein-protein interactions, *Mol. BioSyst. 5*, 924–926.

19. Jones, S., and Thornton, J. M. (1995) Protein-protein interactions—a review of protein dimer structures, *Prog. Biophys. Mol. Bio. 63*, 31–65.

20. Henchey, L. K., Jochim, A. L., and Arora, P. S. (2008) Contemporary strategies for the stabilization of peptides in the alpha-helical conformation, *Curr. Opin. Chem. Biol. 12*, 692–697.

21. Moellering, R. E., Cornejo, M., Davis, T. N., Del Bianco, C., Aster, J. C., Blacklow, S. C., Kung, A. L., Gilliland, D. G., Verdine, G. L., and Bradner, J. E. (2009) Direct inhibition of the NOTCH transcription factor complex, *Nature 462*, 182–188.

22. Horne, W. S., Johnson, L. M., Ketas, T. J., Klasse, P. J., Lu, M., Moore, J. P., and Gellman, S. H. (2009) Structural and biological mimicry of protein surface recognition by alpha/beta-peptide foldamers, *Proc. Natl. Acad. Sci. U.S.A. 106*, 14751–14756.

23. Davis, J. M., Tsou, L. K., and Hamilton, A. D. (2007) Synthetic non-peptide mimetics of alpha-helices, *Chem. Soc. Rev. 36*, 326–334.

24. Kritzer, J. A., Lear, J. D., Hodsdon, M. E., and Schepartz, A. (2004) Helical β-peptide inhibitors of the p53-hDM2 interaction, *J. Am. Chem. Soc. 126*, 9468–9469.

25. Henchey, L. K., Kushal, S., Dubey, R., Chapman, R. N., Olenyuk, B. Z., and Arora, P. S. (2010) Inhibition of hypoxia inducible factor 1-transcription coactivator interaction by a hydrogen bond surrogate α-helix, *J. Am. Chem. Soc. 132*, 941–943.

26. Harrison, R. S., Shepherd, N. E., Hoang, H. N., Ruiz-Gomez, G., Hill, T. A., Driver, R. W., Desai, V. S., Young, P. R., Abbenante, G., and Fairlie, D. P. (2010) Downsizing human, bacterial, and viral proteins to short water-stable alpha helices that maintain biological potency, *Proc. Natl. Acad. Sci. U.S.A. 107*, 11686–11691.

27. Li, W., Jaroszewski, L., and Godzik, A. (2001) Clustering of highly homologous sequences to reduce the size of large protein databases, *Bioinformatics 17*, 282–283.

28. Kuhlman, B., and Baker, D. (2000) Native protein sequences are close to optimal for their structures, *Proc. Natl. Acad. Sci. U.S.A. 97*, 10383–10388.

29. Kortemme, T., and Baker, D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes, *Proc. Natl. Acad. Sci. U.S.A. 99*, 14116–14121.

30. Kortemme, T., Kim, D. E., and Baker, D. (2004) Computational alanine scanning of protein-protein interfaces, *Sci. STKE* l2.

31. Murray, J. K., and Gellman, S. H. (2007) Targeting protein-protein interactions: Lessons from p53/MDM2, *Biopolymers 88*, 657–686.

32. 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.

33. Arkin, M. R., and Wells, J. A. (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream, *Nat. Rev. Drug Discov. 3*, 301–317.

34. Cummings, C. G., and Hamilton, A. D. (2010) Disrupting protein-protein interactions with non-peptidic, small molecule alpha-helix mimetics, *Curr. Opin. Chem. Biol. 14*, 341–346.