JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

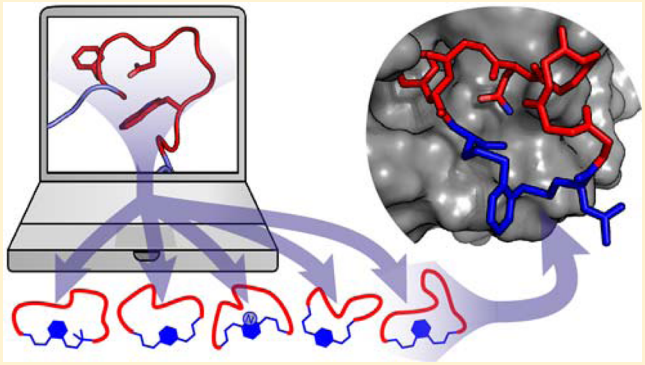
Analysis of Loops that Mediate Protein−Protein Interactions and Translation into Submicromolar Inhibitors

Timothy R. Siegert, Michael J. Bird, Kamlesh M. Makwana, and Joshua A. Kritzer*

Department of Chemistry, Tufts University, 62 Talbot Avenue, Medford, Massachusetts 02155, United States

Supporting Information

ABSTRACT: Effective strategies for mimicking α-helix and β-strand epitopes have been developed, producing valuable inhibitors for some classes of protein−protein interactions (PPIs). However, there are no general strategies for translating loop epitopes into useful PPI inhibitors. In this work, we use the LoopFinder program to identify diverse sets of "hot loops," which are loop epitopes that mediate PPIs. These include loops that are well-suited to mimicry with macrocyclic compounds, and loops that are most similar to variable loops on antibodies and ankyrin repeat proteins. We present data-driven criteria for scoring loop-mediated PPIs, uncovering a trove of potentially druggable interactions. We also use unbiased clustering to identify common structures among the hot loops. To translate these insights into real-world inhibitors, we describe a robust, diversity-oriented strategy for the rapid production and evaluation of cyclized loops. This method is applied to a computationally identified loop in the PPI between stonin2 and Eps15, producing submicromolar inhibitors. The most potent inhibitor is well-structured in water and successfully mimics the native epitope. Overall, these computational and experimental strategies provide new opportunities to design inhibitors for an otherwise intractable set of PPIs.



INTRODUCTION

Protein−protein interactions (PPIs) are increasingly important as therapeutic targets, but small molecules often have low affinity for PPI interfaces.[1,2] Constrained peptides are generally more capable of binding the large, flat surfaces associated with PPIs.[3,4] Modified peptides are increasingly sought after as drug leads due to their high potency, selectivity, and lower overall potential for toxicity.[5] Many design and molecular evolution strategies are available for producing peptides that act as PPI inhibitors, and many of these strategies require (or work better with) starting points that incorporate natural binding epitopes.[6−10] Identification of such epitopes has been facilitated by computational alanine scanning, which calculates the change in binding free energy when a specific residue is mutated to alanine ($\Delta\Delta G_{res}$). Such computational methods have been shown to be reasonably predictive and have led to the discovery of many new PPI epitopes within the Protein Data Bank (PDB).[11−15] Separate studies have scoured the PDB to identify and analyze PPIs involving α-helices,[16,17] helical dimers,[18] and β-strands.[19] Analysis of other PPI interfaces can be more challenging, because they do not depend on well-defined secondary structures and are not identified by existing programs.

In 2014, we introduced LoopFinder, a program which identifies PPIs mediated by loop structures.[20] We define loop structures not as segments lacking secondary structure, but as segments of variable structure that position their termini in close proximity (often called "omega loops").[21] LoopFinder allowed the identification of 1407 "hot loops" that computational energy analysis indicated were important for their respective PPIs. These PPIs were largely orthogonal to sets of PPIs identified by other PDB-wide analyses, demonstrating that many overlooked, potentially druggable PPIs are mediated by loops.

In this work, we improved LoopFinder energy calculations and benchmarked them to real-world values for peptide−protein interactions. Then, we applied LoopFinder more broadly to uncover hundreds of new hot loops, including common loop structures found in antibody complementarity-determining regions (CDRs) and ankyrin repeat proteins. Because loops are more varied in structure than helices or sheets, we wanted to ask whether recurring structures would be observed within the macrocycle loop set. Thus, we implemented a sophisticated clustering algorithm, and isolated 29 specific clusters that are commonly observed among hot loops of 4−8 amino acids. The improved LoopFinder analysis and database of hot loops provide unprecedented insight into this overlooked class of PPI.

In many cases, translating α-helix and β-strand epitopes into potent PPI inhibitors has required the development of specific intramolecular cross-links that stabilize the desired secondary

Received: June 2, 2016
Published: September 9, 2016

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884

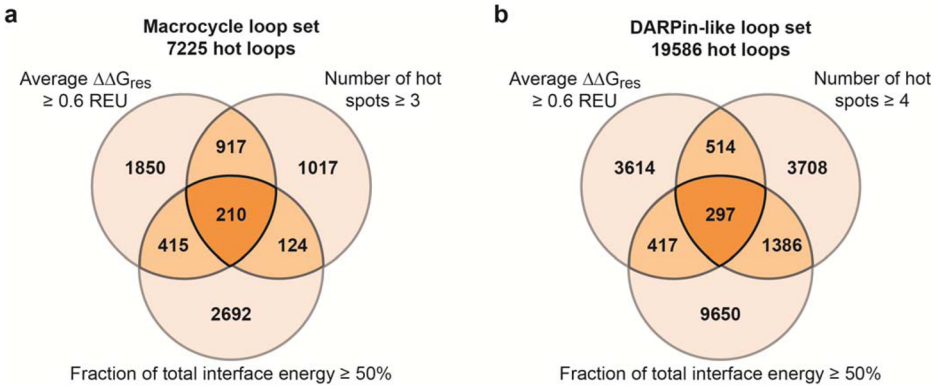


Figure 1. Defining and identifying hot loops. Venn diagrams showing hot loops as identified for the (a) macrocycle and (b) DARPin-like loop sets. Loop counts for each of the three specific hot loop criteria, and loop counts for all overlapping regions, are shown. The central region consists of hot loops that fit all three criteria, and represents the optimal loops for designing inhibitors of their associated PPIs. Similar data for additional loop sets are shown in Figure S2.

structure.[22−24] Because loops are so varied in structure, we anticipated that translating hot loops into macrocyclic PPI inhibitors would not be amenable to a similar "one-size-fits-all" approach. Instead, we devised a strategy that could systematically vary conformational constraints using a robust, sequence-tolerant macrocyclization reaction. Here, we report the successful application of this diversity-oriented approach to the LoopFinder-identified hot loop between stonin2 and Eps15, identifying the first submicromolar inhibitors for this PPI. This methodology is independent of structure, and should be useful for the general translation of hot loops into useful PPI inhibitors.

RESULTS AND DISCUSSION

Revisiting Energy Calculations for the Macrocycle Loop Set. In the first analysis of loop-mediated PPIs,[20] some common assumptions caused many loop energies to be underestimated. In order to capture a complete picture of loop-mediated PPIs, we revisited several of these assumptions. First, the criteria for picking out "hot loops" had been arbitrarily defined, and it would be highly desirable to benchmark these to known peptide−protein interactions. Also, we found that loops were enriched in negative $\Delta\Delta G_{res}$ values as calculated by Rosetta-based alanine scanning. While previous applications of Rosetta-based alanine scanning had ignored these occurrences,[14−19] we found that our loop sets required more nuanced interpretation. These improvements, described below, have dramatically increased the number and quality of hot loops identified with LoopFinder.

For the production of more extensive loop sets, 27 938 PPI structures were downloaded from the PDB in May 2015. Loops are defined by three customizable parameters: a range for loop length in residues, a maximum distance between terminal C_α atoms, and the percent of residues judged to be at the interface (within 6.5 Å of the partner chain). The first new loop set we produced was an updated set of "macrocycle loops": these loops are judged to be most suitable for direct translation into macrocycles that inhibit the associated PPI. The macrocycle loop set consists of continuous stretches of 4−8 amino acids with the terminal C_α atoms ≤6.2 Å apart (shorter cutoffs were used for loops of 4 and 5 residues as described[20,25]) and with at least 80% of their residues at the interface.

Once loops were identified, they were subjected to computational alanine scanning using Rosetta energetics, producing $\Delta\Delta G_{res}$ values for each amino acid (measured in Rosetta Energy Units or REUs; 1 REU is commonly equated to roughly 1 kcal/mol). As noted by us and others, alanine scanning using Rosetta can produce negative $\Delta\Delta G_{res}$ values.[14,15] These are commonly ignored as neutral interactions since stabilizing alanine substitutions that have been experimentally validated are rare, and these seldom have free energy differences more negative than −0.9 kcal/mol.[14,15] Loop-mediated interactions appear to have an unusually large proportion of negative calculated $\Delta\Delta G_{res}$ values. We attribute this to Rosetta energy functions being optimized for protein structures at large, which are primarily composed of regular secondary structures. It may also be that loop structures have greater cooperativity among individual residues, a feature known to be underestimated in Rosetta-based alanine scanning.[15] Overall, this may suggest that current energy models could be updated to describe more accurately energies of residues within loop structures. This is consistent with recent efforts to model accurately macrocycle structure and dynamics, which typically require custom force fields and advanced sampling algorithms.[26−28] In loops containing amino acids with negative $\Delta\Delta G_{res}$ values, we observed that residues with the largest negative $\Delta\Delta G_{res}$ values typically had large extents of buried surface area at the interface and/or atoms participating in poorly aligned hydrogen bonds. On the basis of these observations, we used a conservative cutoff of −2.0 REU as the $\Delta\Delta G_{res}$, below which a residue was highly likely to be a hot spot. For subsequent analysis, the $\Delta\Delta G_{res}$ of these residues were adjusted to 1.0 REU. All $\Delta\Delta G_{res}$ values between −2 and 0 REU were adjusted to 0 REU to negate any effects on subsequent calculations and hot loop identification.

New, Data-Driven Criteria for Identifying Hot Loops. Our previous analysis of LoopFinder results used arbitrary cutoffs for defining important epitopes, as have previous analyses of other PPI interfaces.[16−20] To benchmark hot loop criteria to real-world binding data, we identified well-characterized peptide−protein interactions within the macrocycle loop set. Twelve such peptide−protein interactions were identified, 8 of which had reported K_d or K_i values (Supporting Information, Table S1). Most of the interactions had affinities in the submicromolar range. The values for the average $\Delta\Delta G_{res}$ for the hot loops identified within these peptides were mainly between 0.6 and 1.0 REU per residue. Thus, to capture hot loops similar to these, hot spots were redefined as residues with

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884

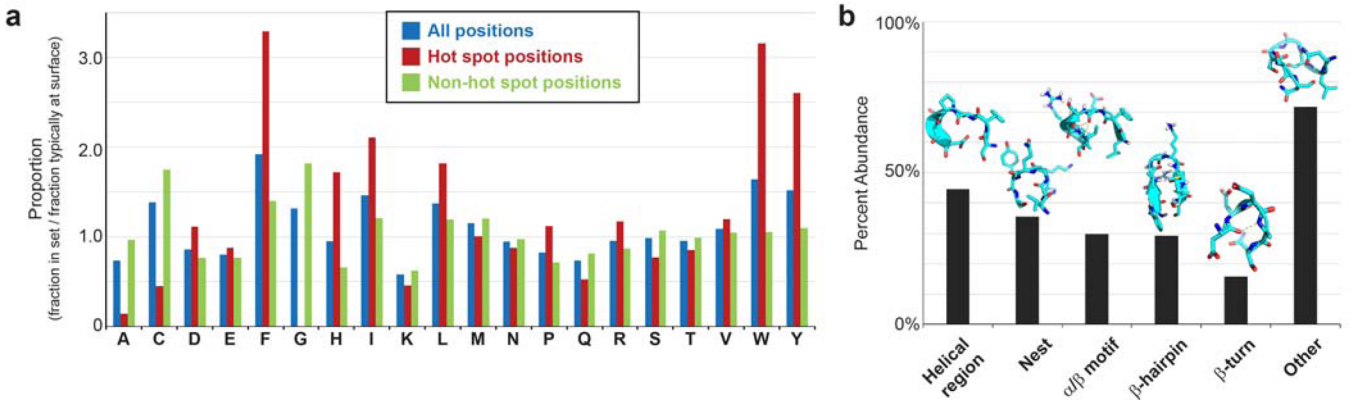


Figure 2. Global analysis of the macrocycle loop set. (a) The 7225 hot loops in the macrocycle loop set were analyzed with respect to amino acid composition. Amino acid abundances in all positions, in hot-spot positions, and in non-hot-spot positions were calculated as relative fraction of the abundance typically observed at the protein surface. Similar data for additional loop sets are shown in SI, Figures S3−S6. (b) The 7225 hot loops in the macrocycle loop set were analyzed with respect to the presence of structural motifs as judged by PDBeMotif. Examples of each type of structural motif are shown above each bar. These data and similar data for additional loop sets are provided in SI, Tables S4−S8.

$\Delta\Delta G_{res} \geq 0.60$ REU. We also used this benchmark to adjust our criterion for classifying hot loops based on average $\Delta\Delta G_{res}$ over the whole loop, specifying that hot loops should have an average $\Delta\Delta G_{res}$ of at least 0.60 REU per residue.

A second criterion for identifying hot loops is number of hot spots. Loops that mediate PPIs using three or more hot spots were categorized as hot loops. We note that loops with only one or two hot spots may be more readily identified using structure-independent hot spot analysis. They may also be more readily mimicked by small molecules rather than macrocycles or other molecules with larger surface area.

The third criterion for selecting hot loops is percent of the overall interface energy represented within the loop. Interface energies were calculated as the sum of $\Delta\Delta G_{res}$ values for all residues at the interface, consistent with prior work.[18] The cutoff for percent interface energy was chosen at 50% so that hot loops would contain a majority of the binding energy of their interfaces.

Using these new criteria, 7225 loops from the macrocycle loop set were identified as satisfying one or more criteria. All together, these data-driven criteria identified over 5-fold more hot loops than were found previously.[20] Of these hot loops, 1666 fulfill multiple hot loop criteria, and 210 fulfill all three (Figure 1a). These represent the best starting points for macrocycle design.

Additional Loop Sets with Diverse Parameters. Identification of hot loops has applications beyond macrocycle design. For instance, antibodies and ankyrin repeat proteins (natural and engineered) are among the most versatile molecules for protein recognition. Both these scaffolds use loop epitopes to mediate PPIs, and replacement of loop epitopes in antibodies and designed ankyrin repeat proteins (DARPins) is a common strategy for the development of proteins with novel fuctions.[29−31] To identify hot loops consistent with structures of antibody CDR loops and DARPin loops, we ran a comprehensive PDB-wide search using additional sets of loop definition parameters (SI, Table S2). These parameters defined four groups of loops: H1/L1-like, H2/L2/L3-like, H3-like, and DARPin-like. Each of these loop sets were larger in size (each with 300 000 to 1.6 million nonredundant loops, see SI, Table S3) and more diverse in structure than the macrocycle loop set. To pull out only the most important of these loops, we applied the same hot loop criteria as the macrocycle loop set, except we required hot loops to have at least 4 hot spots (rather than 3) due to the longer loop lengths. These criteria identified between 11166 and 25141 hot loops for each of these four loop sets, and a total of 1638 loops across all five loop sets that meet all three hot loop criteria (Figure 1 and SI, Figure S2). Overall, these represent a wealth of new, potentially druggable PPIs, each with an associated hot loop suitable for macrocycle engineering and protein grafting.

Amino Acid Composition of the Loop Sets. In motifs that mediate PPIs, specific amino acids have been observed in excess to the proportion typically found at protein surfaces.[20,32,33] These relative differences in amino acid usage have been compared among helix, strand and loop motifs that mediate PPIs.[16,19,34] To see if these trends hold up for diverse hot loops, we compared the relative amino acid abundances among all five sets of hot loops. These are all highly similar (Figure 2a and SI, Figures S3−S6), revealing that hot loops use common residues as hot spots regardless of loop size or structure. Charged residues, which have high abundance on protein surfaces, are all under-represented in the hot loop sets. This is particularly notable for lysine, which is among the most common surface-exposed amino acids but is highly under-represented in hot loops, even at non-hot-spot positions. Despite their moderate relative abundance, aspartate, arginine, and histidine appear often as hot spots. Nonpolar residues are over-represented in hot loops, and as hot spots within those loops. These results are consistent with the established concept that buried regions of PPI interfaces closely resemble the protein interior. Phenylalanine, tryptophan, and tyrosine are all highly over-represented, and are the most common hot-spot residues. As suggested by previous results,[20,34] phenylalanine is more prevalent as a hot spot within hot loops, and less prevalent as a hot spot within strands or helices that mediate PPIs. Interestingly, glycine (which cannot be assigned a $\Delta\Delta G_{res}$ value because it is not mutated during the computational alanine scan) is overrepresented in the macrocycle loop set, but not in any of the other four loop sets. We interpret this as a need for the expanded conformational tolerance of glycine within the shorter loop structures of the macrocycle set, whereas the other four sets consist of longer loops that rely less heavily on glycine for tight turns.

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884

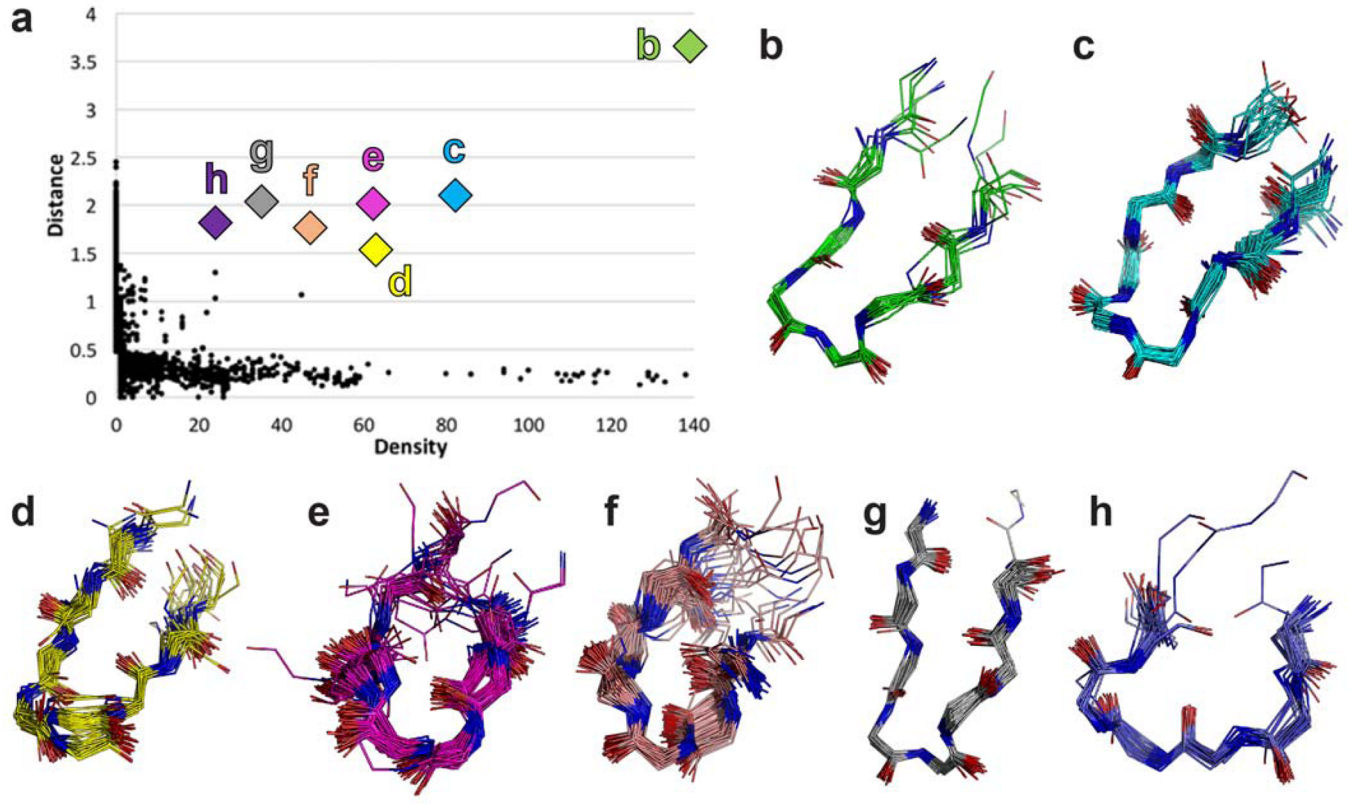


Figure 3. Unbiased clustering of hot loops reveals common structural clusters. (a) Rodriguez-Laio decision graph[36] for all loops of length 6 from the 1666-member set of macrocycle hot loops that fulfill multiple hot loop criteria. This decision graph allows identification of optimal cluster centers in an unbiased manner. For each individual structure, it plots the local density (how many other structures are relatively close in RMSD) versus the distance (in RMSD) to the nearest neighbor with higher density. Structures with high local density, but also a large distance between it and another structure with higher density, are objectively defined as cluster centers.[36] Individual loops that are not cluster centers are shown as black dots, and loops identified as cluster centers are shown as colored diamonds. (b–h) Seven clusters of length 6 are shown, with letter and colors corresponding to the cluster nodes shown in panel a. Longer loops are shown in full, though only 6 residues were used for clustering. Side chains are omitted for clarity. Similar decision graphs and all 29 clusters are shown in SI, Figures S7–S11.

Global Structural Characterization of the Hot Loops. Next, we sought to identify common structural motifs within the hot loops. In previous work, structural motifs were identified using PDBeMotif.[20,35] PDBeMotif is the most suitable system since it distinguishes among many subtly different turn motifs, allowing for a highly descriptive analysis of loop structures. To optimize this process for larger loop sets, we wrote an automated assignment program that recognizes structural motifs (as defined by PDBeMotif) that are at least partially within a hot loop. An average of 2.2 structural motifs per loop were identified for the macrocycle loop set (Figure 2b), and analysis of the longer loops identified even more motifs per loop (SI Tables S4–S8). Although complete helical turns are excluded from the macrocycle loop set, 45% of these loops contain three or more contiguous residues in a helical conformation. These often represent N- or C-terminal helix caps. Within the larger loop sets, between 45% and 67% of loops have single helical turns or helix caps. The biophysical roles of helix caps in stabilizing helical structure are well-established, but these striking findings implicate helix caps as common recognition elements for PPIs.

When β-turns are defined using a characteristic hydrogen bond, only 15.8% of hot loops in the macrocycle set have a discrete β-turn. However, over 6000 of the 7225 hot loops contain β-turn-like torsions, an unsurprising result considering the loops are just 4 to 8 residues. The remainder of these loops can be sorted into various turn-like structures including ASX motifs, ASX turns, ST motifs, ST staples, ST turns, β-bulges, and γ-turns. This structural categorization left only 8.6% of macrocycle hot loops without a known structural motif, and this percentage was only 3.5–6.2% for the larger hot loops. These results reveal that the vast majority of hot loops are not poorly structured. Rather, they contain recognizable turn motifs that could be stabilized in the context of macrocyclic molecules.

Clustering of the Macrocycle Hot Loops. Global analyses of amino acid usage and structural motifs are informative, but we sought higher-resolution analysis of hot loops based on structure. Thus, we performed unbiased, structure-based clustering of all 1666 macrocycle hot loops that meet multiple hot loop criteria. Backbone atoms of each loop were optimally overlaid onto every other loop, and also loop segments with the same number of residues. The backbone RMSD values for the resulting overlays were used as a distance metric, and cluster centers were identified in an unbiased manner using the methodology of Rodriguez and Laio (Figure 3a and SI, Figures S7–S11).[36] A total of 29 clusters were identified, and selected clusters are shown in Figure 3b–h and Figure 4 (all clusters are shown in SI, Figures S7–S11).

There were some surprising findings among the 29 clusters. Even though only 29.3% of the loops adopt hairpin structures as categorized by PDBeMotif, 17 of the 29 clusters are centered around β-hairpins and hairpin-like turns (for example, clusters

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876–12884

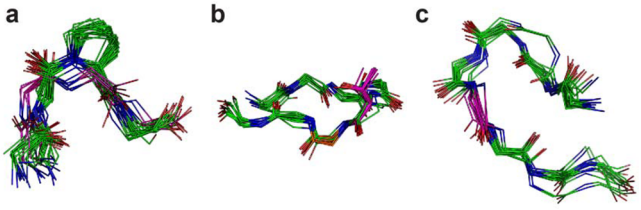


Figure 4. Selected clusters with conserved positions. (a) A cluster of 33 four-residue hot loops that conform to a β-beta-turn-like structure. Proline is present in 78.8% of these loops at position 3. A distinct subcluster (magenta) is observed for the proline-free loops. (b) A cluster of 17 eight-residue hot loops. Threonine is present in 52.9% of these loops at position 5 (magenta), and glycine is present in 52.9% of these loops at position 6 (orange), though these are not correlated. (c) A cluster of 20 eight-residue loops with glycine 95% conserved at position 5 (magenta). Side chains, except for the conserved proline and threonine, are omitted for clarity.

in Figure 3b,c,d,g). Given the unbiased clustering method, this indicates that these hairpins have distinct features that separate them from each other, but relatively tight structural similarity within each cluster. Thus, each of these clusters represents a distinct class of hairpin that commonly mediates PPIs.

Despite the fact that only backbone atoms were used for cluster overlays, many clusters contained highly conserved residues at specific positions. In the four-residue β-turn cluster, proline was conserved in 80% of hot loops at the third position, consistent with common β-turn-stabilizing motifs (Figure 4a).[37−39] One cluster of loops of length 8 had threonine present at the fifth position in over 50% of sequences (Figure 4b). Glycine was the only residue with greater than 50% conservation at individual positions within multiple clusters. In fact, 10 of the 29 clusters had a single position with 50% or greater conservation of glycine. One of these clusters is the same cluster for which threonine was conserved at the fifth position (Figure 4b): glycine is conserved at the sixth position, though the two residues were not correlated. In the most extreme case, one eight-residue cluster consists of 20 loops, 19 of which contain glycine at position 5 (Figure 4c). All together, these observations agree with the global amino acid and structure motif analyses, but provide a much higher degree of resolution. We have identified 29 different common loop structures used by nature to mediate PPIs, and conserved features of each at the residue-by-residue level. These will directly inform the development of specific macrocycle and small-molecule scaffolds aimed at inhibiting separate clusters of loop-mediated PPIs.

Web Site for Public Use of LoopFinder. At LoopFinder.Tufts.edu, we provide online tools for applying LoopFinder to user-initiated projects. Users can perform a loop search within any multimeric PDB entry using custom parameters for defining loops and custom criteria for classifying hot loops. Raw $\Delta\Delta G_{res}$ data are provided so users can analyze the data using methods beyond hot loop identification. The Web site also provides access to databases of the updated macrocycle loop set, the antibody-CDR-like loop sets, and the DARPin-like loop set, which can be searched online or downloaded in full. Loop sets based on additional parameters will be added, and loop sets will be periodically updated as the PDB continues to grow.

Targeting the Eps15-EH2 Domain Using a Computationally Identified Loop From stonin2. Most of the newly identified hot loops are within PPIs that have no known high-affinity inhibitors. One such target was the adapter protein Eps15 (epidermal growth factor substrate 15), which is bound by a hot loop on stonin2 (Figure 5a).[40] Eps15 is part of a

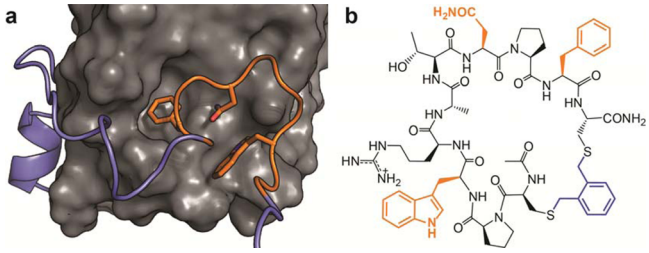


Figure 5. Hot loop within the stonin2-Eps15 interaction. (a) A hot loop on stonin2 (ribbon) was identified with three hot spots (orange side chains) predicted to be important for binding to Eps15-EH2 (surface). Figure rendered from PDB 2JXC.[40] The orange ribbon corresponds to peptide **ST-lin**. (b) Structure of **ST1-oxy**. Predicted hot spots are shown in orange and the *ortho*-xylene (**oxy**) linker is shown in blue.

multiprotein complex that recruits the clathrin-mediated endocytosis machinery to the cell membrane.[41] Eps15 interacts with proteins including epsin, synaptojanin-p170 and stonin2 to recruit clathrin to the cell surface using its three EH (Eps15-homology) domains.[41−45] Upon binding to Eps15 through the second of the EH domains (Eps15-EH2), stonin2 recruits the adapter complex AP-2 to clathrin-coated pits.[40,43] Genetic disruption of this interaction results in impaired endocytosis and reduced Ebola infectivity, demonstrating Eps15-EH2 as an intriguing target for selective control of endocytosis.[46−48]

Prior work had shown that a 39-residue stonin2 fragment binds Eps15-EH2 with a K_d of 0.15 μM, and an unrelated, phage-display-derived peptide binds Eps15-EH2 in the 500 μM range.[49−51] These are the only Eps15-EH2 ligands reported to date. LoopFinder identified an 8-residue loop on stonin2 that fulfills all three hot loop criteria (Figure 5a). Notably, although this hot loop contains the known Asn-Pro-Phe (NPF) turn motif that mediates EH domain interactions,[40,50] we hypothesized that the loop structure, plus the additional hot spot at W309, would make it a better starting point for producing peptide-based inhibitors. To test this hypothesis, a 10-residue peptide from S308 to L316 (**ST-lin**) was synthesized and assessed for binding to Eps15-EH2 using isothermal titration calorimetry (ITC). ITC was necessary because no other biochemical binding assays had been developed for Eps15-EH2. **ST-lin** had a K_d of 18.2 ± 3.4 μM for Eps15-EH2 (Table 1), verifying that LoopFinder had indeed identified a critical loop for the stonin2-Eps15 interaction.

Diversity-Oriented Macrocyclization Strategy. We sought to translate the stonin2 hot loop into an inhibitor, but rational design of small, well-structured cyclic peptides is currently an intractable problem.[26−28] Instead, we devised a strategy to test systematically diverse conformational constraints. We chose to adapt methods for peptide cyclization via dithiol bis-alkylation due to the robustness of the reaction and the variety of structurally diverse linkers available (Figure 6).[52−57] First, we synthesized parent peptides containing the hot loop flanked by the thiol-containing amino acids cysteine or penicillamine. Unpurified, crude peptides were divided into separate vessels and individually cyclized using a panel of dibromomethyl aryl linkers. The linkers used were dibromo-*o*-xylene (**oxy**), dibromo-*m*-xylene (**mxy**), dibromo-*p*-xylene

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884

Table 1. Selected Peptides and Their Binding Affinities and Inhibitory Potencies[a]

peptide, linker	sequence	K_d (μM)	IC_{50} (μM)
ST-lin	SPWRATNPFL	18.2 ± 3.4	>100
ST1-oxy	CPWRATNPFC	0.33 ± 0.01	2.2 ± 0.3
ST1-mxy	CPWRATNPFC	0.37 ± 0.02	2.3 ± 0.2
ST1-pxy	CPWRATNPFC	3.67 ± 0.17	41.3 ± 7.9
ST1-pyr	CPWRATNPFC	1.53 ± 0.05	9.1 ± 3.3
ST1-dbz	CPWRATNPFC	N.B.	27.5 ± 8.3
ST1-oxy-W3A	CPARATNPFC	1.7 ± 0.28	8.3 ± 0.7
ST1-oxy-N7A	CPWRATAPFC	N.B.	>100
ST1-oxy-F9A	CPWRATNPAC	N.B.	>100

[a]All peptides were N-terminally acetylated and C-terminally amidated. Linkers between the two cysteines are indicated in peptide name (see SI, Figure S1 for structures). K_d values were obtained from three independent ITC experiments (SI, Figures S4–S12), and IC_{50} values were obtained from three independent FP experiments. N.B. indicates no binding observed to the limit of solubility.

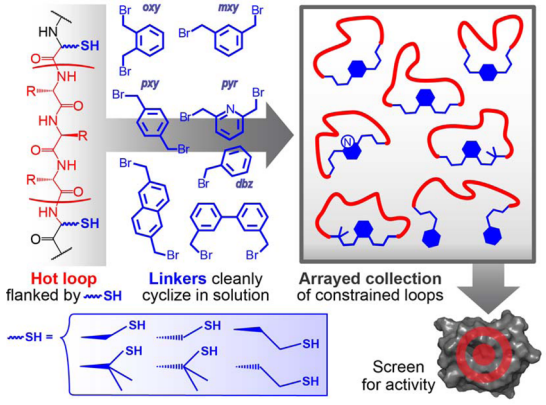


Figure 6. Strategy for parallel synthesis and screening of constrained loops. The robustness of the thiol bis-alkylation reaction allows complete freedom with respect to different loop lengths, sequences, and incorporation of non-natural amino acids. It also allows broad diversification of the linker at a late synthetic stage due to the ability to incorporate a variety of thiol-containing amino acids and dibromo linker reagents. Specific linker reagents used in this work include dibromo-*o*-xylene (**oxy**), dibromo-*m*-xylene (**mxy**), dibromo-*p*-xylene (**pxy**), and 2,6-di(bromomethyl)pyridine (**pyr**). A linear, dibenzylated control peptide was also synthesized using benzyl bromide (**dbz**).

(**pxy**), and 2,6-di(bromomethyl)pyridine (**pyr**) (SI, Figure S12). A linear, dibenzylated control peptide was also synthesized using benzyl bromide (**dbz**). Solution-phase cyclization reactions went to >90% completion in 2 h or less at room temperature, and showed minimal cyclodimer formation even without high dilution (SI, Figure S13).[58] This procedure provides late-stage conformational diversification using a robust, solution-phase cyclization reaction that is easily run in parallel.

Application of the Diversity-Oriented Strategy to Cyclization of the stonin2 Hot Loop. A small initial library of cyclized loops was synthesized and analyzed for binding to Eps15-EH2 via ITC. **ST1-oxy** (chemical structure shown in Figure 5b) and **ST1-mxy** showed similar binding, with K_d values of 0.33 and 0.37 μM, respectively (Table 1 and SI, Table S10). The isomeric peptide with the *para*-xylene linker, **ST1-pxy**, had 10-fold poorer binding affinity. The linear peptide **ST1-dbz**, in which both cysteines are benzylated, had no measurable binding up to 40 μM. We attribute these results to conformational restriction within **ST1-oxy** and **ST1-mxy** that promotes higher-affinity binding to Eps15-EH2. The pyridine-containing cyclic peptide **ST1-pyr** has a 5-fold lower affinity than its closest analog **ST1-mxy**, demonstrating that linker electronics can also play a role in peptide conformation and/or target recognition. In this case, the pyridine group in the linker may destabilize the favored structure by engaging in unwanted dipole-mediated interactions, such as repulsion with another dipole within the peptide or the binding site, intramolecular hydrogen bonding, or hydrogen bonding to water.

To test the computational prediction of specific hot spots within the loop, three analogs of **ST1-oxy** were prepared that substituted each predicted hot-spot residue with alanine. Substitution of the tryptophan (**ST1-W3A-oxy**) resulted in a 5-fold loss in binding affinity (Table 1). This residue originally had a negative $\Delta\Delta G_{res}$ as calculated by Rosetta-based alanine scanning, supporting our methodological changes for energy analysis. Mutation of either the asparagine or phenylalanine hot spots (**ST1-N7A-oxy** and **ST1-F9A-oxy**) resulted in a complete loss of binding. These residues reside within the conserved asparagine-proline-phenylalanine (NPF) motif that mediates most EH domain interactions.[50] Overall, these results confirm the importance of the NPF motif and also demonstrate a substantial contribution by the rest of the hot loop.

Using a Competition Binding Assay To Provide Further Structure–Activity Data for ST1-oxy. With a submicromolar ligand in hand, we next developed a higher-throughput assay to assess binding to Eps15-EH2. We adapted the **ST1-oxy** peptide into a probe for fluorescence polarization (FP) assays (**Flu-ST1-oxy**, SI Figure S14). FP assays measured a K_d of 0.51 μM for the probe binding to Eps15-EH2 (Figure 7a). Eps15-EH2 is only one example in a wide class of EH

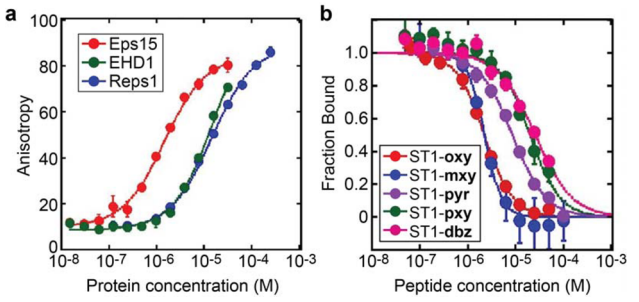


Figure 7. Binding and competitive inhibition data for cyclic peptides based on the stonin2 hot loop. (a) Direct binding of **Flu-ST1-oxy** to EH domains Eps15-EH2 (red), EHD1-EH (blue) and Reps1-EH (green). Curve fits shown are consistent with K_d values of 0.51 ± 0.06 μM, 14.8 ± 1.7 μM, and 15.8 ± 1.1 μM, respectively (SI, Figures S24–S26). (b) **Flu-ST1-oxy** was used as an FP competition probe to obtain valuable structure–activity relationships for **ST1-oxy**. Shown are inhibitors **ST1-oxy** (red), **ST1-mxy** (blue), **ST1-pyr** (purple), **ST1-pxy** (green), and **ST1-dbz** (magenta). IC_{50} curve fits shown reflect values reported in Table 1 (SI, Figures S28–S32).

domains, all of which recognize NPF and DPF motifs with native affinities in the 10–100 μM range under physiological conditions.[59] The specificity of our probe was assessed using EHD1-EH and Reps1-EH domains. **Flu-ST1-oxy** was 30-fold more selective for Eps15-EH2 over these related domains (Figure 7a).

The FP probe was then applied in a competition assay to screen further compounds. Results for the initial panel of cyclic peptides directly paralleled the ITC results (Table 1, Figure

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876–12884

7b), verifying that binding affinity depends on macrocycle conformation. This was further supported by results with a series of similar cyclic peptides that lack the proline in position 2, based on linear peptide CWRATNPFC. This panel showed roughly a 10-fold decreased activity in competition assays, but a similar dependence on the linker in modulating activity (SI, Table S11).

The FP competition assay enabled more rapid exploration of structure–activity relationships for **ST1-oxy**. We tested 27 analogs of **ST1-oxy** that varied key aspects of the linker and hot spots (SI, Figures S24–S48 and Table S11). Replacing the acetyl cap with a pentynyl cap yielded **pST1-oxy**, with a 2-fold improvement in IC_{50}. The added alkyne group will be critical for ongoing studies to assess biological activity and cell penetration. In addition, the C-terminal carboxamide was omitted via synthesis of the linear peptide using side-chain-loaded cysteamine resin. This yielded **pST1b-oxy**, which had an IC_{50} comparable to **ST1-oxy**. Replacing the phenylalanine with the larger 1-naphthylalanine reduces inhibitory potency by 10-fold, whereas substitution of the tryptophan with 1-naphthylalanine results in 2-fold improvement in inhibitory potency. Additional structure–activity data are provided in the SI (Table S11). Overall, these extensive structure–activity data support a binding mode similar to that of the native stonin2 domain (*vide infra*).

This strategy for late-stage conformational diversification can not only take advantage of different linkers but also different thiol-containing amino acids. D-Cysteine, L-homocysteine, and D-homocysteine are straightforward options for diversifying macrocycle size and conformation. We found that penicillamine can be also used with no effect on the robustness of the cyclization reaction. Penicillamine represents an excellent addition to this strategy, adding torsional restriction relative to cysteine (Figure 6). In this case, substitution of one or two penicillamines within **ST1-oxy** resulted in peptides with slightly higher IC_{50} values (SI, Table S11).

Solution Structure of ST1-oxy in Water. To determine whether ST1-oxy was truly recapitulating the structure of the native hot loop, we solved the solution structure of **ST1-oxy** in water using data from 2D-NMR. Chemical shifts were well-dispersed and deviated consistently from values expected for random-coil conformation, indicating a high degree of overall structure (SI, Table S12 and Figures S49–S54). Several medium- and long-range NOEs were observed (Figure 8a and SI, Table S13 and Figure S55). These NOEs plus 12 dihedral angle restraints were used in molecular dynamics simulated annealing to calculate solution structures. The ensemble of 30 lowest-energy structures, shown in Figure 8b, has an overall backbone RMSD of 0.37 Å. Although simulation results must be interpreted carefully, the chemical shifts and the NOEs observed for residues R4 through C10 independently indicate that **ST1-oxy** is well-structured in the NPF turn and the residues immediately flanking the turn. To confirm that **ST1-oxy** displays an unusually high degree of structure in aqueous solution, we used an iodide-mediated fluorescence quenching assay (SI, Figure S56).[60−62] These results showed substantially lower quenching of intrinsic Trp fluorescence for **ST1-oxy** compared to **ST1-dbz**, providing an independent confirmation that **ST1-oxy** is highly structured in aqueous solution.

The solution structure of **ST1-oxy** reveals successful mimicry of the structure of the stonin2 hot loop (Figure 8c). The side chains of the Phe and Asn residues extend into solvent on a shared face, forming an epitope that matches the critical hot spots on stonin2. This epitope docks well into the binding pocket of Eps15-EH2 (Figure 8d).[49] In models of the complex, the *ortho*-xylene linker sits in the pocket occupied by the hot spot Trp in the stonin2-Eps15 structure, with the Trp side chain playing a secondary role in molecular recognition and structural stabilization. This predicted binding mode is supported by the extensive structure–activity data (SI, Table S11). Future work will reveal additional determinants of robust aqueous structure and target binding, allowing this well-structured scaffold to be applied to other EH domains and other loop-mediated PPIs with similar loop structure.

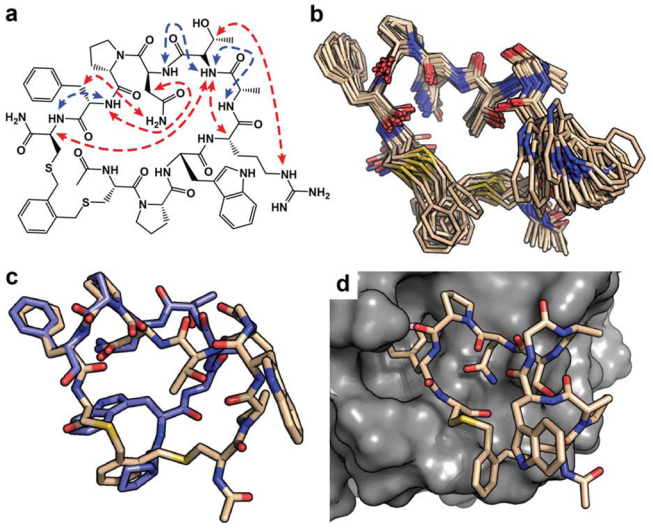


Figure 8. Solution structure of ST1-oxy in water. (a) Diagram showing key NOEs for structure determination of **ST1-oxy**. Key short-range NOEs are indicated by blue arrows, and key medium- and long-range NOEs are indicated by red arrows. A full listing of NOEs and torsional angles used as restraints for simulated annealing are given in SI, Table S13 and Figure S55. (b) The 30 lowest-energy solution structures of **ST1-oxy** resulting from simulated annealing. Backbone RMSD for these structures is 0.37 Å. Structural data and refinement statistics are given in SI, Table S14. (c) Representative NMR structure of **ST1-oxy** (tan) overlaid with structure of stonin2 hot loop (blue). (d) Model for **ST1-oxy** binding to Eps15-EH2 (surface). Images omit the side chain of Arg4 for clarity.

■ CONCLUSIONS

Loop-mediated PPIs are found in a diverse array of biomedically relevant protein complexes but have received little attention compared to PPIs mediated by other secondary structure motifs. LoopFinder is a broadly applicable tool for identifying loop-mediated PPIs, and for the development of those loops into macrocycles as potential PPI inhibitors. In this work, we improved computational procedures and identified common structures and motifs used by nature for loop-mediated PPIs. We found that hot loops typically contain known turn structures, suggesting that well-structured macrocycles can be designed as inhibitors. The 29 structural clusters we identified among the macrocycle hot loops serve as a promising basis for developing useful scaffolds for peptide, protein and small molecule inhibitors.

In this work, we also demonstrated a general strategy for the translation of hot loops into PPI inhibitors, producing submicromolar ligands for the Eps15-EH2 domain. 10-mer cyclic peptides are rarely well-structured in aqueous solution,

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884

especially when they lack canonical α-helical or β-hairpin structure. Several lines of evidence support that **ST1-oxy** is surprisingly well-structured, perhaps due to cooperativity between the cyclization geometry and the observed NPF turn. **ST1-oxy** has only 2-fold weaker affinity than the native binding partner, and has 30-fold selectivity for Eps15-EH2 over related EH domain proteins. Structure−activity relationhips indicated that the key hot spots and overall loop structure identified by LoopFinder were critical for the high degree of affinity and selectivity observed for S**T1-oxy** compared to prior EH domain ligands.[40,50,59,63] **ST1-oxy** is a promising starting point for developing selective inhibitors of Eps15-mediated endocytosis.

Although the specific linker geometry preferred by stonin2-derived peptides may not generalize directly to other loops (the way that one helix-stabilizing "staple" can be generalized to many helical peptides),[24,64] the approach taken here is robust enough to allow rapid parallel testing of many conformations for each individual hot loop. The chemistry is readily extended to many thiol-containing amino acids and many linkers, including naphthyl and biphenyl linkers with various substitution patterns (Figure 6).[54−57] By applying a large set of linkers and thiol-containing amino acids, we are currently adapting this strategy to produce hundreds or thousands of conformationally diverse peptides in arrayed and pooled formats.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05656.

Materials, methods, additional data, and discussion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*joshua.kritzer@tufts.edu

Funding

This work was supported in part by NIH DP20D007303 (J.A.K.), NSF 1507456 (J.A.K.), and Department of Education GAANN P200A090303 (T.R.S.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank S. Pochapsky at Brandeis University for assistance acquiring NMR data, J. Baleja at Tufts University School of Medicine for EH domain expression plasmids, H. Yu for assistance with NMR structure refinement, Tufts Technology Services for web assistance, and R. Scheck for helpful comments. Loop databases are provided in full at http://LoopFinder.Tufts.edu.

■ REFERENCES

(1) Makley, L. N.; Gestwicki, J. E. *Chem. Biol. Drug Des.* **2013**, *81* (1), 22.
(2) Arkin, M. R.; Tang, Y.; Wells, J. A. *Chem. Biol.* **2014**, *21* (9), 1102.
(3) Fosgerau, K.; Hoffmann, T. *Drug Discovery Today* **2015**, *20* (1), 122.
(4) Tsomaia, N. *Eur. J. Med. Chem.* **2015**, *94*, 459.
(5) Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. *Chem. Biol. Drug Des.* **2013**, *81* (1), 136.
(6) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5* (7), 502.
(7) Schlippe, Y. V. G.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134* (25), 10469.
(8) Hayashi, Y.; Morimoto, J.; Suga, H. *ACS Chem. Biol.* **2012**, *7* (3), 607.
(9) Kritzer, J. A.; Lear, J. D.; Hodsdon, M. E.; Schepartz, A. *J. Am. Chem. Soc.* **2004**, *126* (31), 9468.
(10) Henchey, L. K.; Jochim, A. L.; Arora, P. S. *Curr. Opin. Chem. Biol.* **2008**, *12* (6), 692.
(11) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. *Nucleic Acids Res.* **2005**, *33*, W382.
(12) Koes, D. R.; Camacho, C. J. *Nucleic Acids Res.* **2012**, *40*, W387.
(13) Tuncbag, N.; Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R. *Briefings Bioinf.* **2009**, *10* (3), 217.
(14) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (22), 14116.
(15) Kortemme, T.; Kim, D. E.; Baker, D. *Sci. Signaling* **2004**, *2004* (219), pl2.
(16) Bullock, B. N.; Jochim, A. L.; Arora, P. S. *J. Am. Chem. Soc.* **2011**, *133* (36), 14220.
(17) Jochim, A. L.; Arora, P. S. *ACS Chem. Biol.* **2010**, *5* (10), 919.
(18) Watkins, A. M.; Wuo, M. G.; Arora, P. S. *J. Am. Chem. Soc.* **2015**, *137* (36), 11622.
(19) Watkins, A. M.; Arora, P. S. *ACS Chem. Biol.* **2014**, *9* (8), 1747.
(20) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. *Nat. Chem. Biol.* **2014**, *10* (9), 716.
(21) Fetrow, J. S. *FASEB J.* **1995**, *9* (9), 708.
(22) Wilson, A. J. *Chem. Soc. Rev.* **2009**, *38* (12), 3289.
(23) Perez de Vega, M. J.; Martin-Martinez, M.; Gonzalez-Muniz, R. *Curr. Top. Med. Chem.* **2007**, *7* (1), 33.
(24) Walensky, L. D.; Bird, G. H. *J. Med. Chem.* **2014**, *57* (15), 6275.
(25) Siegert, T. R.; Bird, M. J.; Kritzer, J. A. *Methods Mol. Biol.* **2016**, accepted.
(26) Yu, H.; Lin, Y.-S. *Phys. Chem. Chem. Phys.* **2015**, *17* (6), 4210.
(27) Geng, H.; Jiang, F.; Wu, Y.-D. *J. Phys. Chem. Lett.* **2016**, *7* (10), 1805.
(28) McHugh, S. M.; Rogers, J. R.; Yu, H.; Lin, Y.-S. *J. Chem. Theory Comput.* **2016**, *12* (5), 2480.
(29) Perchiacca, J. M.; Ladiwala, A. R. A.; Bhattacharya, M.; Tessier, P. M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (1), 84.
(30) Miersch, S.; Sidhu, S. S. *Methods* **2012**, *57* (4), 486.
(31) Pluckthun, A. *Annu. Rev. Pharmacol. Toxicol.* **2015**, *55*, 489.
(32) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285* (5), 2177.
(33) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280* (1), 1.
(34) Arntson, K. E.; Pomerantz, W. C. K. *J. Med. Chem.* **2016**, *59* (11), 5158.
(35) Golovin, A.; Henrick, K. *BMC Bioinf.* **2008**, *9*, 312.
(36) Rodriguez, A.; Laio, A. *Science* **2014**, *344* (6191), 1492.
(37) Macarthur, M.; Thornton, J. *J. Mol. Biol.* **1991**, *218* (2), 397.
(38) Krieger, F.; Moglich, A.; Kiefhaber, T. *J. Am. Chem. Soc.* **2005**, *127* (10), 3346.
(39) Gunasekaran, K.; Nagarajaram, H. A.; Ramakrishnan, C.; Balaram, P. *J. Mol. Biol.* **1998**, *275* (5), 917.
(40) Rumpf, J.; Simon, B.; Jung, N.; Maritzen, T.; Haucke, V.; Sattler, M.; Groemping, Y. *EMBO J.* **2008**, *27* (3), 558.
(41) Benmerah, A.; Gagnon, J.; Begue, B.; Megarbane, B.; Dautry-Varsat, A.; Cerf-Bensussan, N. *J. Cell Biol.* **1995**, *131* (6), 1831.
(42) Benmerah, A.; Poupon, V.; Cerf-Bensussan, N.; Dautry-Varsat, A. *J. Biol. Chem.* **2000**, *275* (5), 3288.
(43) Chi, S.; Cao, H.; Chen, J.; McNiven, M. A. *Mol. Biol. Cell* **2008**, *19* (8), 3564.
(44) Chi, S.; Cao, H.; Wang, Y.; McNiven, M. A. *J. Biol. Chem.* **2011**, *286* (40), 35196.
(45) Doria, M.; Salcini, A. E.; Colombo, E.; Parslow, T. G.; Pelicci, P. G.; Di Fiore, P. P. *J. Cell Biol.* **1999**, *147* (7), 1379.

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884

(46) Benmerah, A.; Bayrou, M.; Cerf-Bensussan, N.; Dautry-Varsat, A. *J. Cell Sci.* **1999**, *112* (9), 1303.
(47) Carbone, R.; Fre, S.; Iannolo, G.; Belleudi, F.; Mancini, P.; Pelicci, P. G.; Torrisi, M. R.; Di Fiore, P. P. *Cancer Res.* **1997**, *57* (24), 5498.
(48) Bhattacharyya, S.; Warfield, K. L.; Ruthel, G.; Bavari, S.; Aman, M. J.; Hope, T. J. *Virology* **2010**, *401* (1), 18.
(49) de Beer, T.; Carter, R. E.; Lobel-Rice, K. E.; Sorkin, A.; Overduin, M. *Science* **1998**, *281* (5381), 1357.
(50) de Beer, T.; Hoofnagle, A. N.; Enmon, J. L.; Bowers, R. C.; Yamabhai, M.; Kay, B. K.; Overduin, M. *Nat. Struct. Biol.* **2000**, *7* (11), 1018.
(51) Rumpf, J.; Simon, B.; Groemping, Y.; Sattler, M. *Biomol. NMR Assignments* **2008**, *2*, 55.
(52) Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. *ChemBioChem* **2005**, *6* (5), 821.
(53) Chen, S.; Morales-Sanfrutos, J.; Angelini, A.; Cutting, B.; Heinis, C. *ChemBioChem* **2012**, *13* (7), 1032.
(54) Jo, H.; Meinhardt, N.; Wu, Y.; Kulkarni, S.; Hu, X.; Low, K. E.; Davies, P. L.; DeGrado, W. F.; Greenbaum, D. C. *J. Am. Chem. Soc.* **2012**, *134* (42), 17704.
(55) Muppidi, A.; Doi, K.; Edwardraja, S.; Drake, E. J.; Gulick, A. M.; Wang, H.-G.; Lin, Q. *J. Am. Chem. Soc.* **2012**, *134* (36), 14734.
(56) Chua, K.; Fung, E.; Micewicz, E. D.; Ganz, T.; Nemeth, E.; Ruchala, P. *Bioorg. Med. Chem. Lett.* **2015**, *25* (21), 4961.
(57) Micewicz, E. D.; Sharma, S.; Waring, A. J.; Luong, H. T.; McBride, W. H.; Ruchala, P. *Int. J. Pept. Res. Ther.* **2016**, *22* (1), 67.
(58) Peraro, L.; Siegert, T. R.; Kritzer; Joshua, A. *Methods Enzymol.* **2016**, accepted.
(59) Confalonieri, S.; Di Fiore, P. P. *FEBS Lett.* **2002**, *513* (1), 24.
(60) Lehrer, S. *Biochemistry* **1971**, *10* (17), 3254.
(61) Eftink, M.; Ghiron, C. *Anal. Biochem.* **1981**, *114* (2), 199.
(62) Keizer, J. *J. Am. Chem. Soc.* **1983**, *105* (6), 1494.
(63) Kamens, A. J.; Eisert, R. J.; Corlin, T.; Baleja, J. D.; Kritzer, J. A. *Biochemistry* **2014**, *53* (29), 4758.
(64) Harrison, R. S.; Shepherd, N. E.; Hoang, H. N.; Ruiz-Gómez, G.; Hill, T. A.; Driver, R. W.; Desai, V. S.; Young, P. R.; Abbenante, G.; Fairlie, D. P. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (26), 11686.

DOI: 10.1021/jacs.6b05656
J. Am. Chem. Soc. 2016, 138, 12876−12884