Identification of Loop Mediated Protein-Protein Interactions and Design of Cyclic Peptide Locked Loop Inhibitors

A dissertation submitted by

Timothy Ryan Siegert

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

Tufts University

May 2016

Adviser: Dr. Joshua A. Kritzer

Abstract

The landscape of drug discovery has recently begun to shift from traditionally druggable enzymes to protein-protein interactions (PPIs). Enzymes possess well defined active sites, while PPIs often have large and flat surface areas that small molecules can not interrupt. As this landscape shifts, it is essential that the tools used to target these proteins change to meet these demands. One class of molecules particularly adept at binding proteins are peptides. While rational design of substrate transition state mimics as inhibitors has led to successful targeting of enzymes, an analogous approach for rational design PPI inhibitors using peptides is a difficult task. The identification of hot spot residues at protein interfaces has enabled the characterization of essential amino acid residues at protein interfaces amenable to targeting or mimicry using peptides. Much previous work on this subject has used computational approaches to identify epitopes on proteins that facilitate PPIs containing either well-defined secondary structure, such as α-helices and β-sheets or have disregarded any structural information searching for purely extended linear regions. This thesis work describes LoopFinder, a computational tool developed to identify loops and turns at protein interfaces for design and development of synthetic cyclic peptide inhibitors of PPIs. This program was used to identify a hot loop residing on the protein stonin2 at its interface with Eps15, a complex essential for clathrin-mediated endocytosis. The hot loop was used to inspire the development of a cyclic peptide “locked loop library”. Libraries were prepared using a panel of bisbromomethy-aryl linkers for bis-alkylation of cysteine residues introduced to flank the hot loop. Using this approach, we were able to identify sub-micromolar binding peptides to the Eps15-EH2. This peptide was then used to develop a novel biochemical assay for Eps15 binding molecules enabling the search for new inhibitors of endocytosis.

Acknowledgments

This dissertation is a culmination of a long journey through graduate school and could not have been possible without the intellectual and emotional support of many individuals.

First and foremost, as I set off on my scientific endeavors it is has been the greatest blessing of my life to have the unconditional support of my loving wife Bridget. I am forever grateful for the trust, confidence, and patience she bestows upon me on a daily basis. I am lucky to have such a steadfast partner throughout life’s journeys. As much as I have worked hard to accomplish the scientific findings that culminate in this thesis, Bridget has worked tirelessly to find success in her own career and find the patience to continually encourage me towards achieving my potential in chemistry. She is an equal part in the emotional rigor poured into my research and equally deserving in all my accomplishments, for without her support I would definitely not be where I am today.

I am incredibly grateful for my parents Thomas and Kathleen who imparted on me an endless curiosity and have always been supportive of me pursuing my dreams. Always there to pick up the phone in times of frustration, they have constantly encouraged me and helped me believe in myself. My siblings Megan, Callie, Rebecca and Christopher I can never express how much it meant to have your support. My grandparents have always been incredible role models for me and have taught me so many things over the years. I am incredibly grateful for their example, especially Virginia Kearney and Noel Siegert who were able to see me begin this endeavor, but were not able to see me complete this goal. In addition to the family I went into graduate school with, I gained another family during my time at Tufts. The whole Probst family, especially Matt and Cathy, have been an incredible support system over the years and have done so much to encourage me towards my goal. The laughter and happiness all my family consistently bring to my life have been integral to my success and have helped carry me along through this process.

My adviser Joshua Kritzer has been instrumental in my academic and scientific achievement. Thank you for constantly challenging me and pushing me to think with sound reasoning, communicate effectively, and push the limits of my scientific comfort zone. I could not have gotten to where I am today without your guidance and patience. I will be forever grateful to you for helping me pursue my goals. Your excitement for our work is infectious and you have served as a great role model and mentor these past five years. In addition to my adviser, my committee members Professor David Walt and Professor Krishna Kumar were instrumental in helping to guide me through graduate school and have offered valuable advice and criticisms throughout the various checkpoints of my graduate career. Thank you as well to Professor Adrian Whitty from Boston University who graciously accepted my invitation to serve as an outside committee member for this dissertation.

I would also like to thank all the members of the Kritzer Lab, past and present, who have worked along side me in this journey. You have all helped me in so many different ways throughout the process and have constantly motivated and challenged me every day. Intellectual curiosity and success in this field cannot be achieved without a fertile environment for ideas to grow. It has been an amazing and fun experience to grow together in our passion for science. I know looking forward towards our continued friendships that we will share much more knowledge as we all grow in our careers.

Last but not least, the work described in this dissertation could not have been accomplished without the hard work of key individuals. Brad Sheneman wrote the code for the LoopFinder program and implementing the first iteration of the loop search. Dr. Jason Gavenonis was a great teammate helping to interpret the LoopFinder data from this first LoopFinder search discussed in Chapter 1. Michael Bird was essential for the streamline automation of the LoopFinder data extraction and organization, writing the python scripts discussed in Chapter 2. Mathew Eshelman and Dr. Sue Pochapsky (Brandeis University) helped me collect the NMR spectra on my peptide. Dr. Hongtao Yu taught me how to use the NMR computational annealing program and wrote a script to identify structure violations. I am incredibly grateful for all their time and hard work that allowed me to find success.

Table of Contents

List of Tables

List of Figures

List of Equations

1 Chapter 1 : Computational Identification of Loop Mediated Protein-Protein Interactions

1.1 Computational Methods to Identify Peptidic Epitopes from Protein-Protein Interactions for Inhibitor Design

Specific interactions between proteins are responsible for a wide range of signaling processes in the cell. As a result, targeting specific protein-protein interactions (PPIs) is a growing field of drug discovery.[1-5] While the number of small-molecule PPI inhibitors is steadily increasing, overall progress has been slow. One issue is the chemical space covered by most small molecule collections is geared towards maximizing likely oral bioavailability.[6,7] However, these relatively small, hydrophobic, and topologically simple molecules are not ideal for targeting PPIs. Lipinski's Rule of Five provides guidelines for enhancing drug-like properties of small molecules such as possessing no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, a molecular weight of < 500 Daltons, and a solubility of $logP \leq 5$.[8] While this may be sufficient to bind small enzyme active sites, PPIs typically have interaction surfaces spanning 600-1,300 Å^2 or more, and are predominated by polar residues.[9-11] Peptides provide an excellent set of properties allowing them to be promising leads for targeting PPIs. Many canonical examples of successfully targeted PPIs, such as the p53-MDM2[12] and Bcl-xL-BH3[13] interactions, have involved the development of α-helical peptide ligands binding into hydrophobic clefts.[14,15] While campaigns focusing on inhibiting helix-mediated PPIs have been successful, it is less clear how to target the larger subset of non-helix-mediated PPIs as most complexes have broader interfaces made up of non-helical secondary structures.[16]

In order to target PPIs it is important to understand how proteins recognize their binding partners. Despite the large interfaces of most PPIs, numerous studies have shown that residues at the interface do not contribute equally to the binding interaction.[17] “Hot spots” have been defined as individual residues that contribute a significant portion of binding free energy to the overall interaction (often, $\Delta\Delta G \geq 1.5$ kcal/mol).[9] Hot spot residues were first identified by systematic alanine scanning of PPI interfaces, followed by quantifying the change in free energy of binding ($\Delta\Delta G_{res}$) for each mutation.[9,17] It was found that most of the binding interactions could be attributed to relatively few hot spots at the interface.[17] Targeting these hot spot regions, or mimicry of these hot spots using peptides provides a promising technique for targeting PPIs with high affinity and selectivity.[18-22]

Systematic alanine scanning of PPIs is a laborious task and requires time-consuming protein expression, purification and quantification techniques. To further expedite the process of hot spot identification, several computational methods have been developed to perform alanine scanning *in silico*. These include the FOLDEF algorithm, which systematically substitutes residues with alanine and calculates the change in binding energy using FoldX complex energy functions that include terms for desolvation, van der Waals forces, hydrogen bonding, Coulombic interactions, entropy change and dipole interactions [23]. Kortemme and Baker developed a physical model that predicts PPI interface hot spots.[24,25] Using this approach, computational alanine scans of protein-protein complexes have yielded a wide range of information about the different roles specific residues play at PPI interfaces.[26-28] This method uses the Rosetta software package.[24,25] Rosetta's combination of explicit terms for Lennard-Jones energies, solvation energies, protein repacking, solvation and hydrogen bonding are used to calculate changes in free energy when side chains are systematically substituted by alanine and account for protein refolding.[24,25] These computational methods require only the PDB information for each PPI in question, and can accurately identify side chains at the interface that are critical for the binding interaction. Kortemme and Baker's method has been found to be in 79% agreement with experimental techniques.[24]

Some additional approaches to uncovering critical binding sites have been developed that examine protein surfaces rather than performing computational alanine scanning. The Camacho lab developed a tool called ANCHOR that identifies 'anchor' residues at protein interfaces.[29] These anchor residues are identified by measuring changes in the solvent-accessible surface area (ΔSASA) of side chains upon receptor binding. This tool is effective in identifying small hydrophobic pockets on protein surfaces that can later be used to develop small molecule ligands, and is available at the PocketQuery website.[30] Other computational methods introduce small organic probes or molecule fragments such as ammonium cations, carboxyl groups or methane to identify small pockets on protein surfaces.[31,32] These techniques for measuring pockets are useful, and lend themselves best to small molecule binding site discovery.

Beyond identifying potential binding sites for inhibitors, structural information can also be used to inform the rational design of inhibitors. This is a particularly promising opportunity for the development of peptides that inhibit PPIs. Recently, multiple techniques have been developed that identify essential binding epitopes within PPIs. These can be used directly to design peptides that mimic the epitopes and inhibit the native interaction. With the goal of peptide inhibitor development in mind, the Schueler-Furman lab developed a method called PeptiDerive to identify ‘hot segments’, defined as stretches of ten amino acids within a protein that contributes most of the binding energy to complex formation.[33] This program sequentially breaks down protein chains involved in PPIs into all possible ten-amino-acid segments, and then uses Rosetta energy functions via a program called FlexPepDock to calculate the binding energy of each peptide segment to the receptor protein. The total binding energy of each individual peptide segment is then compared to the estimated energy of the entire PPI to identify hot segments. Using a small starting set of 151 PPIs, it was found that about 60% of PPIs possessed a single ten-mer peptide segment that contributed $\geq$ 50% of the interaction energy of the whole complex.[33] This method was used to predict a peptide segment from myeloid differentiation factor 2 (MD2) that should bind and activate MD2's binding partner, Toll-like receptor 4.[34] The linear peptide was synthesized and tested for ability to activate TLR4 signaling as measured by increased production of nitric oxide in macrophages. The linear peptide showed no activation. In order for activity to be observed, further computational modeling was required to introduce and optimize a disulfide cyclization constraint that would better stabilize the native loop structure. This macrocyclic peptide showed some activation of the inflammatory response, inducing nitric oxide production in macrophages.[34] These results reveal a key limitation of ignoring structure when identifying linear segments for the design of PPI inhibitors. Structural stabilization is often required for high-affinity binding by short peptides, due to the inability of the peptide to be predisposed to adopting the optimal binding conformation.

One of the most comprehensive programs that combine hot spot identification with structural analysis is the helix interface in protein-protein interactions (HIPP) program developed by Arora and co-workers.[35,36] This method identifies α-helical regions at protein interfaces that are critical for mediating PPIs. Alpha-helical regions were identified using Rosetta secondary structure calculations of ψ and φ angles. Subsequent computational alanine scanning mutagenesis was used to determine the energy contribution of each residue within the helices. Helices were then analyzed with respect to their binding pockets, with some binding into deeper clefts and others binding over more extended surfaces. Helices were also sorted based on the proportion of hot spots residing on the helix compared to the overall chain.[26] Numerous strategies are available for mimicking or structurally stabilizing α-helices, including side-chain-to-side-chain crosslinks,[37-39] backbone-to-backbone crosslinks,[40] and backbone replacements using unnatural residues such as β-amino acids[41-44], non-peptidic scaffolds such as terphenyls[45] or macrocycles[46] all of which can be applied to the development PPI inhibitors. Recently, Arora and co-workers adapted a similar strategy to β-strands, identifying strands that are responsible for a large proportion of the binding energy of their associated PPIs.[47] Together, these tools identify PPI-mediating epitopes that contain common secondary structures. Existing and novel strategies for mimicking those structures can then be used to translate these epitopes into real-world PPI inhibitors.[48-50] These methods significantly enhance the rational design approach for the development of peptidic inhibitors of PPIs.

While linear segments, α-helices and β-strands are all effective starting points for inhibitor design, we were inspired by the rich diversity of cyclic peptide natural products to ask whether epitopes that do not fall into these structural categories could be promising starting points for peptide PPI inhibitor design. Despite successes identifying and targeting α-helical PPIs, not all protein-protein interactions occur between regions with well-defined secondary structure. According to a previous analysis of protein complexes, 50% of interface residues have non-regular secondary structure, while only 26% possess α-helical secondary structure and 24% possess β-strand secondary structure.[16] Many canonical examples of PPIs, including Fab•antigen and growth hormone•receptor complexes, are mediated largely by loops lacking ordered secondary structure.[51,52] Novel chemical space and structural motifs will be required to target a broader variety of PPIs. Also, many cyclic peptide natural products, such as cyclosporine A and the amanitins, lack common secondary structure elements.[53,54] Instead, these molecules have extensive intramolecular hydrogen bonding patterns that are rare internally in protein structures, and their unique conformations are critical for their high potency and bioavailability.[55,56] We reasoned that, to look for epitopes that could be mimicked by similar classes of cyclic peptides, we should examine loop regions at PPI interfaces. We developed our LoopFinder program with the hypothesis that loops are critical for PPIs and can be translated into natural-product-like cyclic peptides.

We used LoopFinder, a program written by Brad Sheneman, to survey the entire PDB and identified interface loops that mediate PPIs.[57] We analyzed these interface loops using computational alanine scanning mutagenesis to identify those that contribute significantly to binding interactions. In analogy to hot spots and hot segments, we coined the term "hot loops". Taken together, this overall methodology, including the set of interface loops and hot loops, provided an original analysis of molecular recognition involving short peptide loops. These loops present an attractive starting point for the rational design of macrocycles as PPI inhibitors.

1.2 Initial Analysis of PPIs by LoopFinder

In 2014, we published a dataset of loops and hot loops that we identified within the PDB using our LoopFinder program as outlined in Figure 1.1.[57] To prepare this dataset, 19,657 multi-chain structures were acquired from the PDB in August 2013, representing multi-chain structures with ≤ 4Å resolution and < 90% sequence identity. The LoopFinder program first scrubbed the input PDB files to remove headers, then defined each possible binary protein-protein interface within multi-protein complexes. We defined a shared interface using a distance requirement of 6.5 Å ensuring the chains are in close proximity. LoopFinder then searched for peptide loops at protein-protein interfaces subject to several user-defined parameters. First, loops were limited to segments of 4 to 8 consecutive amino acids to conform with molecular mass ranges typical for peptide and macrocycle ligands. Another parameter required at least 80% of the loop to be composed of interfacial residues, which were defined as amino acids having at least one atom within 6.5 Å of the protein's binding partner. Finally, a 6.2 Å cutoff was used between the α-carbons of the loop termini, to ensure a loop-like conformation and to exclude secondary structures such as β-strands and α-helices. This distance was further restricted to a maximum of 4.67 Å for four-amino-acid loops and 5.83 Å for five-amino-acid loops, in order to eliminate non-loop structures. These parameters were selected to identify specific loops that might be amenable to mimicry by small cyclic peptides and other macrocycles. There are many other applications for LoopFinder that can be implemented by changing these parameters.

For the above parameters, LoopFinder identified 121,086 total loops in 9,388 different structures, including numerous redundancies such as overlapping loops, nested loops, and homologous loops on different chains of homo-multimeric proteins (Figure 1.1). These redundant loops were retained for computational alanine scanning because they would not significantly increase the computational burden, and would allow us to select the most critical loops from among them after interaction energies were assigned to each residue.

The complete set of 121,086 loops was then analyzed by computational alanine scanning with PyRosetta v2.012 and Rosetta 3.0, using a modified version of the *ala_scan.py* Python script developed by Gray. This script recapitulates Kortemme's scoring function but without environment-dependent hydrogen-bonding terms.[24,25,58,59] The computational alanine scan produced data that is interpreted as the relative, predicted $\Delta\Delta G_{res}$ for each residue for that particular PPI. The units of this measurement are Rosetta Energy Units (REUs) and roughly comparable but not identical to kcal mol^{-1}. At this point, the loop set was consolidated to eliminate redundant loops, producing a master list of 25,005 interface loops.

These interface loops were then sorted by the presence and relative location of hot spot residues, with hot spots defined as $\Delta\Delta G_{res} \geq 1$ REU. Our initial criteria for identifying "hot loops" was two consecutive hot spots (1,043 loops), loops with at least three hot spots (269 loops), and loops for which the average $\Delta\Delta G_{res}$ was ≥ 1

REU (472 loops) (Figure 1.2). The loops shown within the Venn diagram were chosen to illustrate the diversity within each hot loop parameter. The placement of the hot spots within these loops vary, as well as their overall structure. As can be seen from the ≥ 2 consecutive hot spot parameter however, we hypothesize that the hot loops within this region may be more amenable to inhibition by small molecules, and therefore intended to focus mainly on hot loops containing non-consecutive hot spots. The three hot loop parameters yielded a set of 1,407 hot loops, covering 1,242 multi-chain PDB structures. Notably, this represented only 6.3% of the input structures, highlighting that this process identified those PPIs that were most directly mediated by short peptide loops.

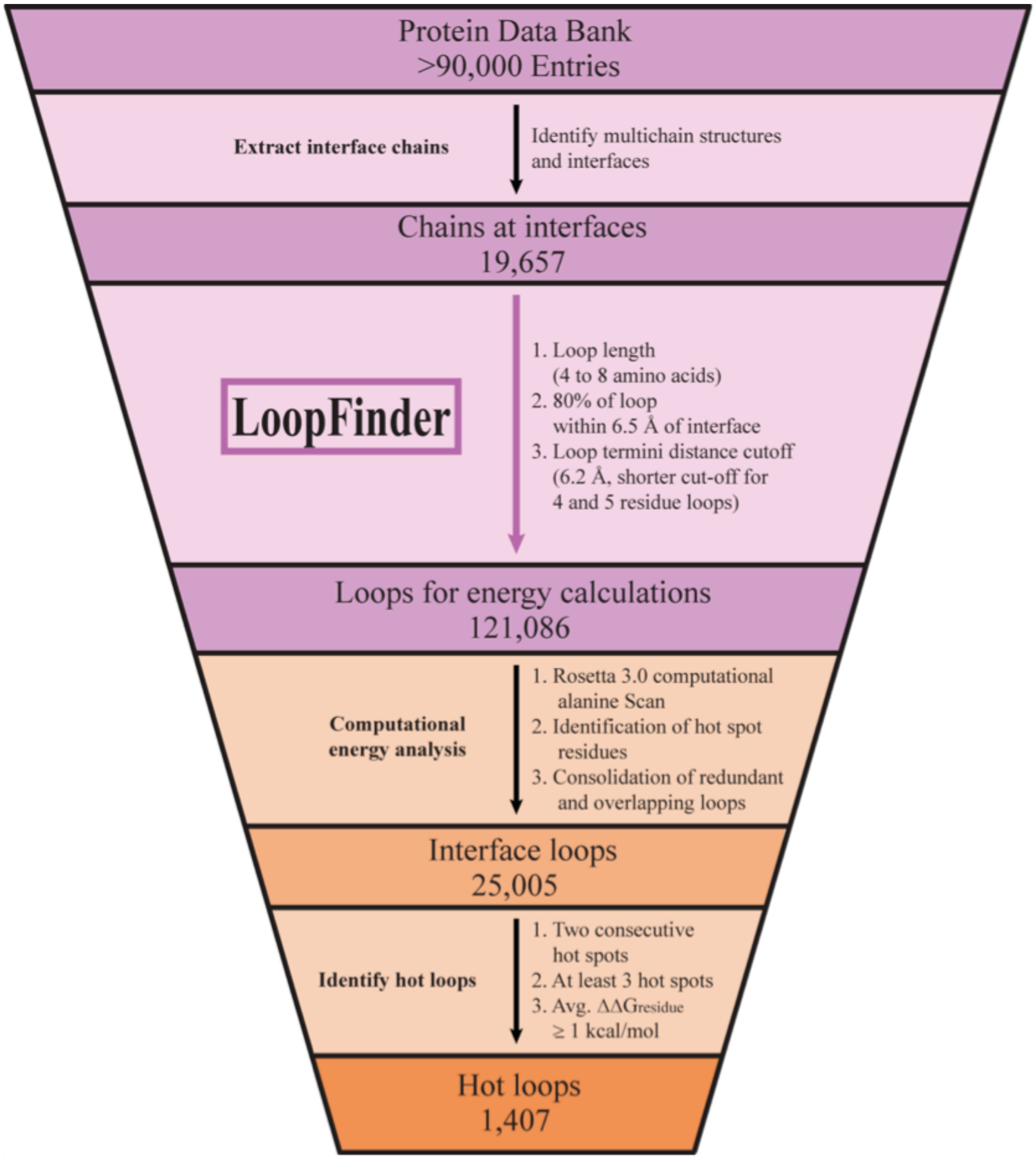


Figure 1.1 Workflow for using LoopFinder. LoopFinder was used to analyze all heterogeneous protein-protein complexes in the PDB, identifying just over 121,000 loops at protein-protein interfaces with the given structural parameters. These loop regions were subjected to computational alanine scan mutagenesis, using Rosetta to calculate the relative contributions of each amino acid to the PPI.[24,25] The resulting data were consolidated into a set of 25,005 interface loops, representing a non-redundant set of all loops that mediate PPIs, each with complete computational alanine scan data. The interface loop set was further sorted by key criteria in order to identify those loops containing large proportions of the overall binding energy of the PPI. These comprise the set of 1407 hot loops.

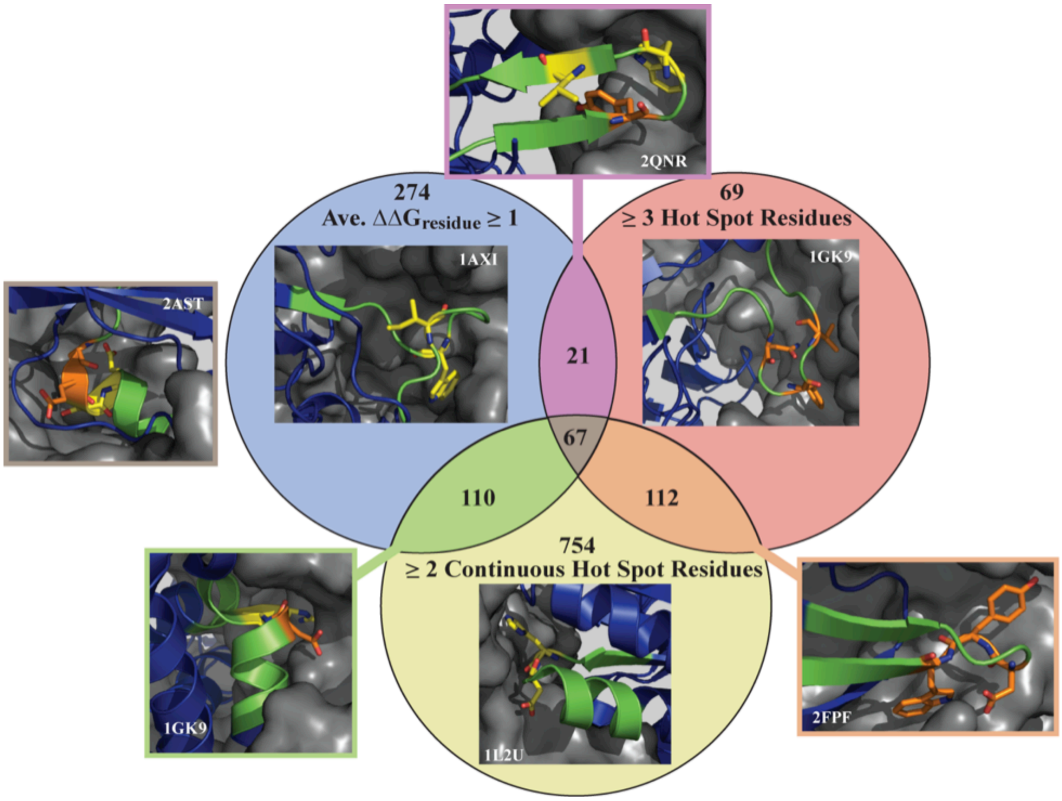


Figure 1.2 Identification of hot loops. Hot loops are identified as those loops that satisfy one or more of three criteria: the average $\Delta\Delta G_{res}$ over the entire loop is greater than 1 REU, the loop has three or more hot spot residues ($\Delta\Delta G_{res} \geq 1$ REU), and the loop has two or more consecutive hot spot residues. Representative loops that satisfy each of these criteria are shown within the blue, red and yellow circles (structures from 1AXI, 1GK9, and 1L2U, respectively).[1,60,61] Some hot loops satisfy two of these criteria, with representative loops from these categories shown in the purple, orange and green boxes (2QNR, 1GK9 and 2FPF, respectively). In addition, 67 hot loops satisfy all three requirements, an example of which is shown in the gray box to the left (2AST).[62] All structures, rendered in Pymol,[63] show the chain at the interface in blue, the binding partner as a gray surface, the hot loop in green, and hot spots in orange ($\Delta\Delta G_{residue} \geq 1$ REU) or yellow ($\Delta\Delta G_{residue} \geq 2$ REU).

1.3 Structural Analysis of Hot Loops from LoopFinder

The original set of 25,005 interface loops that mediate PPIs and 1,407 hot loops identified by computational alanine scanning warranted closer characterization to better understand how proteins use loops to mediate PPIs. Hot loops were analyzed with respect to loop structure by identifying secondary structures flanking the loops and canonical turns within the loops. Canonical turn motifs were identified by submission to PDBeMotif webtool.[64] This analysis revealed structural information about the types of loops categorized as hot loops and a breakdown of overlapping loop structures (Figure 1.3). The hot loops identified by LoopFinder can be grouped into three categories: loops with defined structural motifs (62%), loops with helical turns (11%), and other loops (27%). Thus, roughly 73% of the hot loops possessed specific turn motifs with intramolecular hydrogen bonds.[64] This analysis also showed that the parameters of LoopFinder were defined conservatively enough to exclude most canonical α-helix structures. In addition, we found that specific α-turns identified by LoopFinder had very little overlap with previously identified α-helix-mediated PPIs.[26,35,65,66]

Unsurprisingly, of the loop regions with well-defined structural motifs, β-turns were the most common turns found in hot loops, represented in 48% of all loops. Several specific subcategories of β-turns are prominent in the hot loop set, including αβ-motifs and Schellmann loops. Furthermore, we found that other structural motifs commonly overlapped with β-turn regions within hot loops

(Figure 1.3). The second most common types of structured loops within hot loops utilize serine or threonine to form a side-chain-to-backbone hydrogen bond. There are three classes of such motifs, S/T-turns, S/T-motifs and S/T-staples, which together comprised 20% of all hot loops. Another prominent loop within the loop set were β-hairpins, defined as any loop that connects two antiparallel β-strands regardless of the presence of a β-turn.[67] 14% of the hot loops were β-hairpins, and these types of structures have been thoroughly studied and successfully mimicked using peptides and peptidomimetics and represent an excellent structural space to identify loops for inhibitor design.[68-72] The final significant category of loop structures was similar to S/T-loops, but instead used aspartate or asparagine to form side-chain-to-backbone hydrogen bonds. Asx-turns and motifs constituted an additional 14%. One additional small subset of structured hot loops are β-bulges, which are short breaks within β-strands. Structures of representative hot loops from each of these categories are shown in Figure 1.3, and demonstrate the wide diversity of loop structures identified by LoopFinder. No immediate correlation was observed among the relative 3D orientations of hot spot residues within each structural class, indicating that, while common structural motifs were observed, different molecular scaffolds may need to be developed to target different structural classes or even different loops within each structural class. Another interesting observation from the hot loop set was that “non-canonical” loops and turns (Figure 1.3) possessing unique structures are nonetheless potentially excellent starting points for inhibitor design due to their expected specificity. Some of these loops act as unusual N-terminal or C-terminal caps of α-helices. Representative examples of loops with unusual conformations and hydrogen bonding patterns are shown in Figure 1.4. There are a wide variety of backbone-backbone, side chain-side chain, and side chain-backbone hydrogen bonding interactions that occur within these loops, giving each a unique topology that may promote high selectivity for their respective binding partners. This type of intramolecular hydrogen-bonding pattern is reminiscent of the cyclic peptide natural products that inspired this approach. Cyclization of these loops with flexible or rigidified linkers should produce constrained peptides with uniquely folded structures. These would represent novel three-dimensional space for targeting these and other PPIs.

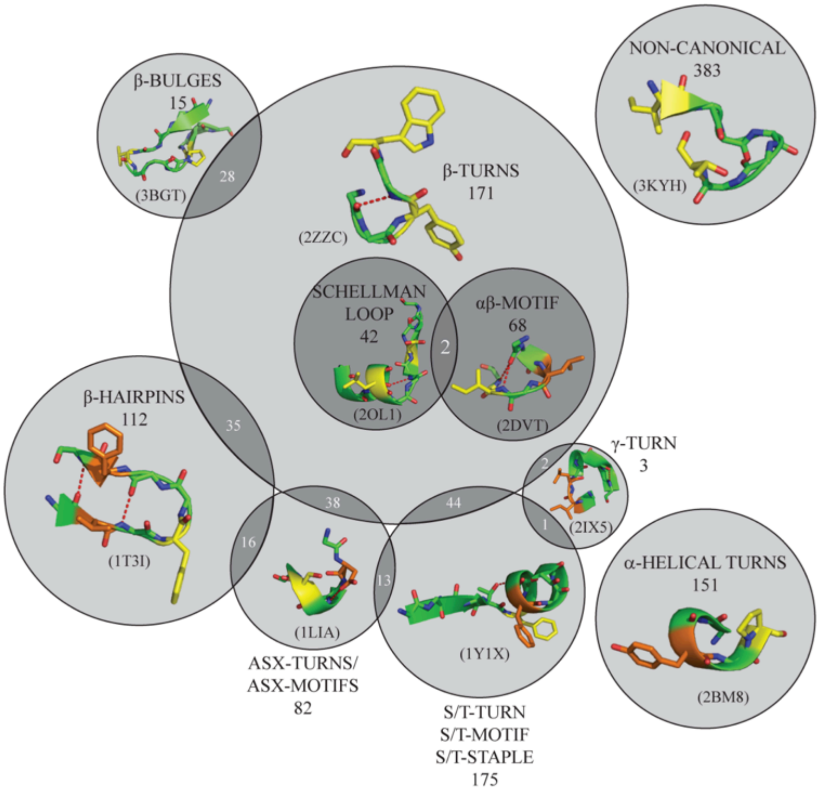


Figure 1.3 Loop structures observed among the hot loops. Representative examples of each type of loop are shown within each circle, including: β-turns (2ZZC),[73] Schellman loops (2OL1),[74] αβ-motifs (2DVT),[75] β-bulges (3GBT),[76] β-hairpins (1T3I),[77] Asx-turns and motifs (1LIA),[78] S/T-turns, motifs and staples (1Y1X),[79] and γ-turns (2IX5).[80] The remaining two categories shown above are α-helical regions identified by their backbone torsional angles (2BM8),[81] and loops lacking canonical structural motifs (3KYH).[82] All structures, rendered in Pymol,[63] show the hot loop in green and hot spots in orange ($\Delta\Delta G_{residue} \geq 1$ REUs) or yellow ($\Delta\Delta G_{residue} \geq 2$ REUs).

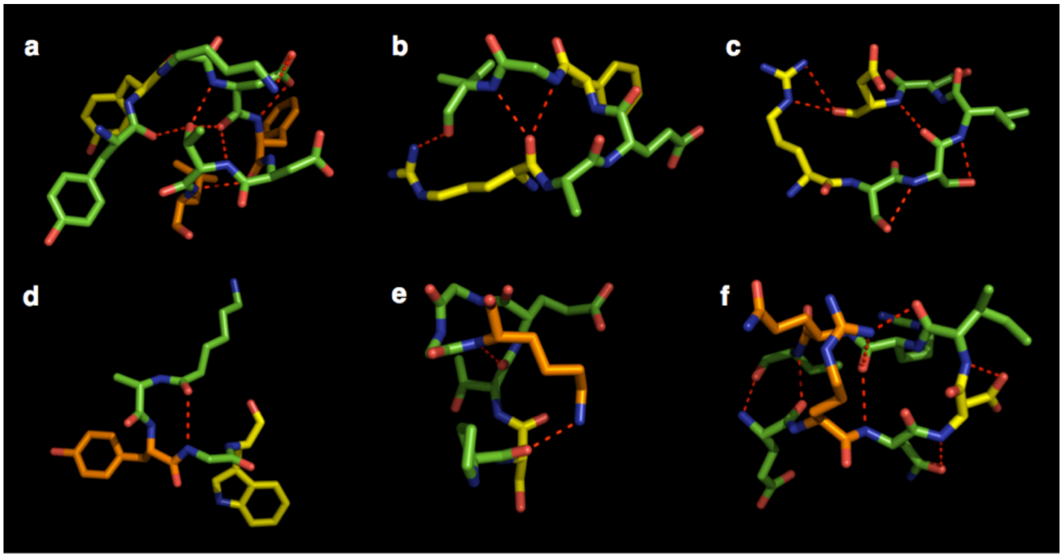


Figure 1.4 Hot loops containing unusual, non-canonical turn motifs. a) An interface loop from *Borrelia burgdorferi* BbCRASP-1 with three non-consecutive hydrophobic hot spots constrained by an (*i*,*i*+1) salt bridge and a central threonine with an (*i*,*i*-3;*i*,*i*-6) ST-staple (1W33).[83] b) An interface loop from the *Mycobacterium tuberculosis* toxin-antitoxin complex RelBE2 with a bidentate (*i*,*i*+4;*i*,*i*+5) main-chain hydrogen bond at the C-terminal end of an alpha helix (3G5O).[84,85] c) An interface loop with a non-canonical β-turn of non-standard backbone torsional angles and two (*i*,*i*+1) side-chain-to-backbone hydrogen bonds (3OA8); d) An interface loop from *Plasmodium falciparium* thioredoxin reductase with a non-canonical β-turn at the C-terminal end of an α-helix (4B1B).[86] e) An interface loop from potassium ion channel Trek2 with a side-chain-to backbone hydrogen bond from lysine at the N-terminal end of an α-helix (4BW5).[87] f) An interface loop from R-spondin-1 with an extended β-hairpin-like structure containing an (*i*,*i*+4) hydrogen bond and side-chain-to-backbone hydrogen bonds among Arg (*i*,*i*+3;*i*,*i*+4), Asn (*i*,*i*+1), and Asp (*i*,*i*+1), with both charged residues also being hot spots (4KNG).[88] All structures, rendered in Pymol,[63] show the hot loop in green, and hot spots in orange ($\Delta\Delta G_{residue} \geq 1$ REUs) or yellow ($\Delta\Delta G_{residue} \geq 2$ REUs).

1.4 Amino acid composition of loops at protein interfaces

We analyzed our dataset to look at how each amino acid side chain is incorporated within loops at PPIs. Previous work in the literature has shown protein surfaces are generally enriched in the charged amino acids (aspartate, glutamate, lysine and arginine), while protein interiors are generally enriched in hydrophobic amino acids (valine, leucine, isoleucine and phenylalanine).[89] Hydrophobic patches on the surface are relatively destabilizing for monomeric proteins, but these effects can be balanced by interactions with binding partners.[90] Thus, hot spots are more likely to be found at hydrophobic patches on a protein surface.[17] In an excellent survey of protein surfaces in both monomeric and multimeric protein structures, Bogan *et al*. found that the most common amino acids that mediate PPIs are tryptophan, arginine and tyrosine, comprising 21%, 13.3% and 12.3% of hot spots respectively.[17] By contrast, leucine, methionine, serine, threonine and valine residues were found very rarely as hot spots.[17] This hot-spot bias has been explained by the fact that tryptophan and tyrosine are large, aromatic but still somewhat polar, allowing some favorable interactions with solvent and maximal interactions with binding partners. This hypothesis is supported by the observation that tryptophan and tyrosine are found on the surface and interior of proteins with identical frequency.[91] Phenylalanine, by contrast, was found to be relatively underrepresented at protein surfaces, and as a result is three times less likely to be a hot spot compared to tyrosine or tryptophan.[17,89,91,92] Other observations of prior studies of PPI hot spots include aspartate being over two times more likely as a hot spot than glutamate, and isoleucine being ten times more likely as a hot spot than leucine.[17] These are surprisingly wide differences considering their similar structures and physicochemical properties, and the differences were attributed to changes in side-chain conformational entropy.[17,91,93] By analyzing the amino acid bias within our sets of non-redundant loops and hot loops, we sought to explore how molecular recognition by hot loops differs from that of surface hot spots in general.

To quantify amino acid bias within loops at PPI interfaces, we compiled the average $\Delta\Delta G_{res}$ for each amino acid for both the interface loop set and the hot loop set (Table 1.1 and Table 1.2). The overall trends for both sets are similar, indicating that the hot loops were similar to all interface loops with respect to amino acid usage. The amino acids that had the highest average $\Delta\Delta G_{res}$ within all interface loops were tryptophan, phenylalanine, histidine, aspartate, tyrosine, leucine, glutamate, isoleucine, and valine. These amino acids span charged, hydrophobic, and aromatic residues, and contain several striking features. Phenylalanine, which was found to be disfavored for protein surface hot-spots in general,[13] had a higher average $\Delta\Delta G_{res}$ contribution than tyrosine, and almost as much as tryptophan. Further, while phenylalanine comprised 4.0% of all loop residues, its occurrence in hot spots was over threefold its occurrence in non-hot-spot positions. For comparison, tryptophan comprises only 1.3% of all interface loop residues, but is 3.4-fold more likely to appear in a hot spot than elsewhere in a loop. Thus, whatever reason phenylalanine is disfavored as a hot spot for PPIs in general does not affect loop-mediated PPIs, making phenylalanine as important for loop-mediated PPIs as tryptophan. Another surprising result was that histidine was a major contributor to the binding energy of hot loops, whereas histidine was not observed at higher proportions as a hot-spot residue for PPIs in general.[17] Finally, leucine and isoleucine had nearly equal average $\Delta\Delta G_{res}$ within the interface loop dataset, and were similarly enriched in hot spots within those loops. Thus, while isoleucine is more well-represented at surface hot spots in general,[17] leucine and isoleucine contribute nearly equally to binding interactions within hot loops.

To gain more insight into how proteins use loops to mediate PPIs, we normalized the percent abundance of each amino acid within the interface loop set to the propensity of each amino acid to reside at the protein surface. Then, we broke down these data into change in percent abundance (relative to surface propensity) for hot spot positions and non-hot spot positions (Figure 1.5).[92] Overall, these values were similar for the interface loop set and the hot loops. Some of the results of this analysis were unexpected but not particularly surprising. For instance, glycine residues were highly enriched in the interface loop set, which was expected in loop regions and in the specific loop architectures identified in Figure 1.3. Despite their prominence in interface loops, glycine was highly under-represented as hot-spot residues within loops, which is also expected considering the alanine-scanning protocol used to determine $\Delta\Delta G_{res}$ and the small interaction surface of glycine. Another expected result was the prominence of large and hydrophobic amino acids, since they commonly mediate PPIs and have high average $\Delta\Delta G_{res}$ values within loops.

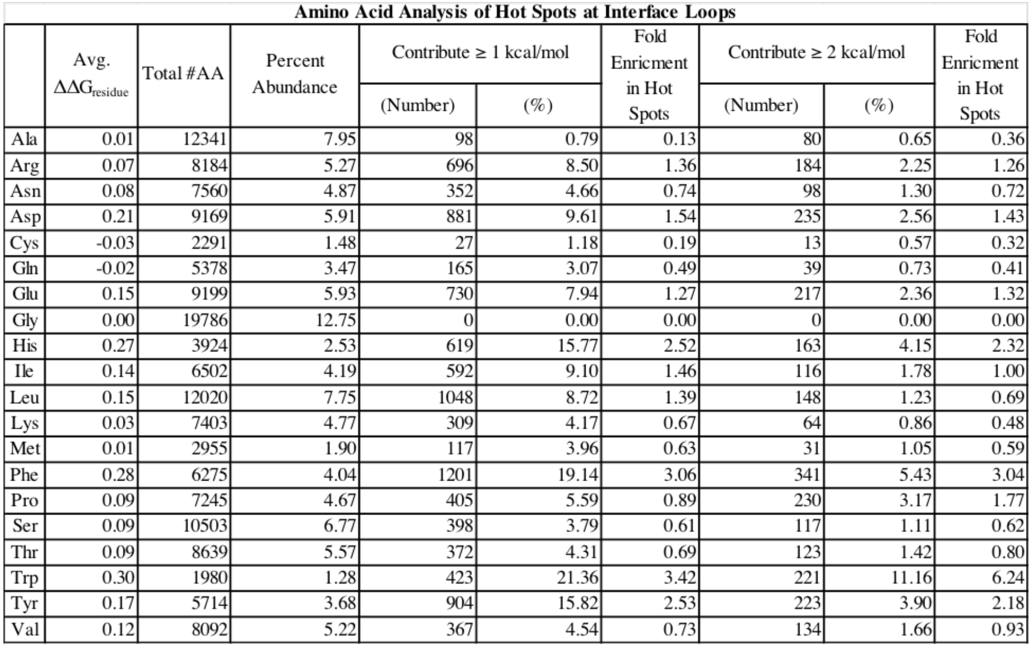
Proline might also be expected to be prominent in loops, but it was not over-represented in interface loops or hot loops. This is despite the fact that, when it was present, it (on average) contributed significantly to the binding energy

(

Table 1.1). Closer examination of the subset of hot loops containing a proline hot spot (179 loops, 13%) further elucidated the roles of "hot prolines" within these loops. Among these loops, the hot proline was often the residue within the loop contributing most to the interaction (125 loops, 70% of proline-containing hot loops), and in a smaller subset, it was the only hot spot found in the loop (19 loops, 11%). In the majority of these loops, the hot proline sat at the boundary between the loop and an α-helix or β-strand. This suggested that proline may not only act as a secondary structure disruption, but can also play a prominent role in intermolecular interactions.

Within the interface loop set, charged amino acids contribute relatively large $\Delta\Delta G_{res}$. However, these are also generally prominent on protein surfaces, and are therefore not over-represented within loops (Figure 1.5). Thus, these data indicate that charged amino acids play similar roles in loop-mediated PPIs as they do in all PPIs, but that loop-mediated PPIs take advantage of arginine, aspartic acid, and glutamic acid equally (without a large over-abundance of arginine). Strikingly, lysine is among the most abundant amino acids at protein surfaces and one of the most abundant amino acids in flexible loop regions,[17,91,94,95] but was greatly under-represented within the non-redundant loop set, both overall (Figure 1.6) and at hot spots (Figure 1.5). This may be because arginine, aspartate, and glutamate can more readily facilitate hydrogen-bond networks with high cooperativity and stability. The underrepresentation of lysine within interface loops, both at hot spots and at non-hot spot positions, may distinguish the interface loops found by LoopFinder from other loops located at the protein surface. This indicates that underrepresentation of lysine may be a method for identifying interface loops solely from primary sequence data and secondary structure predictions.

Finally, histidine was over-represented at hot spots within loops (Figure 1.6), and contributes a very high $\Delta\Delta G_{res}$ on average (Table 1.1). This overall analysis did not distinguish whether histidine was contributing to polar interactions, or to aromatic and/or hydrophobic interactions. Visual inspection of "hot histidine" residues in the hot loop set indicated that histidine acts mainly as a hydrogen bond donor, making specific polar contacts both within the loop and to the binding partner. Hydrophobic, Van der Waals, or π-stacking interactions involving the imizadole appeared to play a less important role for these histidines, which are typically exposed to solvent at the binding interface.



Amino Acid Analysis of Hot Spots at Interface Loops

	Avg. $\Delta\Delta G_{residue}$	Total #AA	Percent Abundance	Contribute ≥ 1 kcal/mol		Fold Enricment in Hot Spots	Contribute ≥ 2 kcal/mol		Fold Enricment in Hot Spots
				(Number)	(%)		(Number)	(%)	
Ala	0.01	12341	7.95	98	0.79	0.13	80	0.65	0.36
Arg	0.07	8184	5.27	696	8.50	1.36	184	2.25	1.26
Asn	0.08	7560	4.87	352	4.66	0.74	98	1.30	0.72
Asp	0.21	9169	5.91	881	9.61	1.54	235	2.56	1.43
Cys	-0.03	2291	1.48	27	1.18	0.19	13	0.57	0.32
Gln	-0.02	5378	3.47	165	3.07	0.49	39	0.73	0.41
Glu	0.15	9199	5.93	730	7.94	1.27	217	2.36	1.32
Gly	0.00	19786	12.75	0	0.00	0.00	0	0.00	0.00
His	0.27	3924	2.53	619	15.77	2.52	163	4.15	2.32
Ile	0.14	6502	4.19	592	9.10	1.46	116	1.78	1.00
Leu	0.15	12020	7.75	1048	8.72	1.39	148	1.23	0.69
Lys	0.03	7403	4.77	309	4.17	0.67	64	0.86	0.48
Met	0.01	2955	1.90	117	3.96	0.63	31	1.05	0.59
Phe	0.28	6275	4.04	1201	19.14	3.06	341	5.43	3.04
Pro	0.09	7245	4.67	405	5.59	0.89	230	3.17	1.77
Ser	0.09	10503	6.77	398	3.79	0.61	117	1.11	0.62
Thr	0.09	8639	5.57	372	4.31	0.69	123	1.42	0.80
Trp	0.30	1980	1.28	423	21.36	3.42	221	11.16	6.24
Tyr	0.17	5714	3.68	904	15.82	2.53	223	3.90	2.18
Val	0.12	8092	5.22	367	4.54	0.73	134	1.66	0.93

Table 1.1 Analysis of hot spots from the total interface loop data set of 25,005 structures. The average $\Delta\Delta G_{res}$ value for each amino acid, in alphabetical order, is shown in the first column. The total number of each amino acid present in the interface loop set is shown in the second column, followed directly by percent abundance of each amino acid in the total loop set. The number that each amino acid appears as a hot spot and the percent abundance of each amino acid in all loops. In order to identify which amino acids are favored as hot spots, we calculate the fold enrichment of each amino acid in hot spots for two hot spot cut-offs of 1 REU and 2 REUs.

Amino Acid Analysis of Hot Spots in Hot Loops

	Avg. $\Delta\Delta G_{residue}$	Total #AA	Percent Abundance	Contribute ≥ 1 kcal/mol		Fold Enricment in Hot Spots	Contribute ≥ 2 kcal/mol		Fold Enricment in Hot Spots
				(Number)	(%)		(Number)	(%)	
Ala	0.30	468	5.39	36	7.69	0.22	32	6.837606838	0.54
Arg	0.83	545	6.28	234	42.94	1.23	67	12.29357798	0.97
Asn	0.56	418	4.82	99	23.68	0.68	40	9.56937799	0.76
Asp	0.97	597	6.88	284	47.57	1.37	80	13.40033501	1.06
Cys	0.37	97	1.12	11	11.34	0.33	7	7.216494845	0.57
Gln	0.53	246	2.84	61	24.80	0.71	21	8.536585366	0.68
Glu	0.86	613	7.06	262	42.74	1.23	87	14.19249592	1.12
Gly	0.00	860	9.91	0	0.00	0.00	0	0	0.00
His	1.07	319	3.68	181	56.74	1.63	60	18.80877743	1.49
Ile	0.95	368	4.24	165	44.84	1.29	49	13.31521739	1.05
Leu	0.83	659	7.59	294	44.61	1.28	65	9.863429439	0.78
Lys	0.50	348	4.01	99	28.45	0.82	27	7.75862069	0.61
Met	0.55	119	1.37	29	24.37	0.70	13	10.92436975	0.86
Phe	1.29	484	5.58	318	65.70	1.89	113	23.34710744	1.85
Pro	1.27	459	5.29	183	39.87	1.15	125	27.23311547	2.16
Ser	0.59	594	6.85	155	26.09	0.75	54	9.090909091	0.72
Thr	0.75	470	5.42	140	29.79	0.86	52	11.06382979	0.88
Trp	1.55	186	2.14	116	62.37	1.79	78	41.93548387	3.32
Tyr	1.11	408	4.70	233	57.11	1.64	70	17.15686275	1.36
Val	0.74	419	4.83	120	28.64	0.82	56	13.36515513	1.06

Table 1.2 Analysis of hot spots in hot loops. Upon comparison to Table 1.1, conclusions can be made as to which amino acid residues are most likely to be contained within a hot loop.

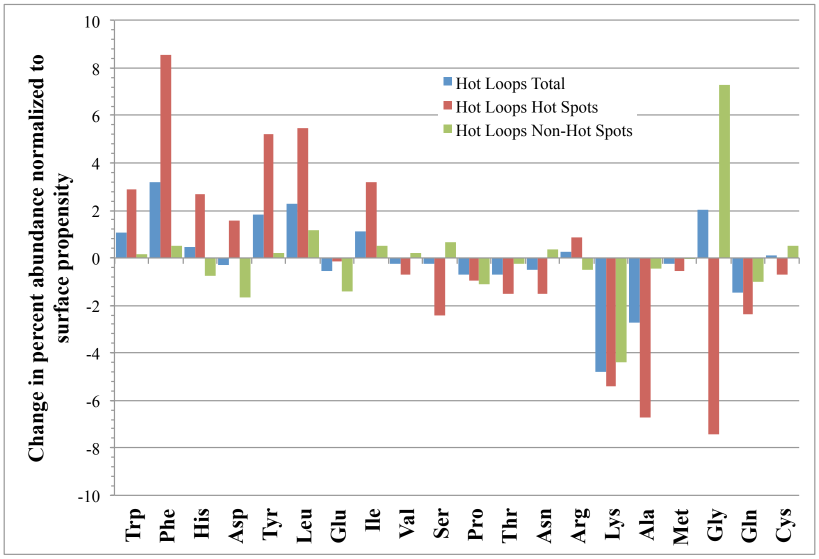


Figure 1.5 The percent abundance of amino acids in hot loops. Amino acid abundance in the hot loop data (blue), hot spot residues (red) and non-hot spot residues (green) were compared directly to the natural surface percent abundance observed previously by Janin *et al.*[92]

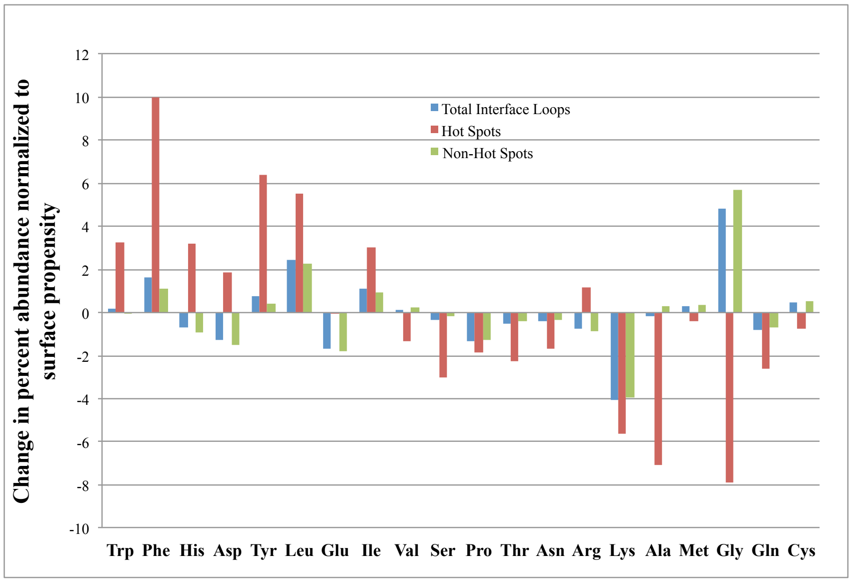


Figure 1.6 The percent abundance of amino acids in all loops. Interface loops use a unique set of amino acids to recognize their binding partners. The percent abundances of each amino acid was normalized relative to expected surface propensity. [92] These normalized values were further broken down into all residues (blue), hot spot residues (red) and non-hot spot residues (green).

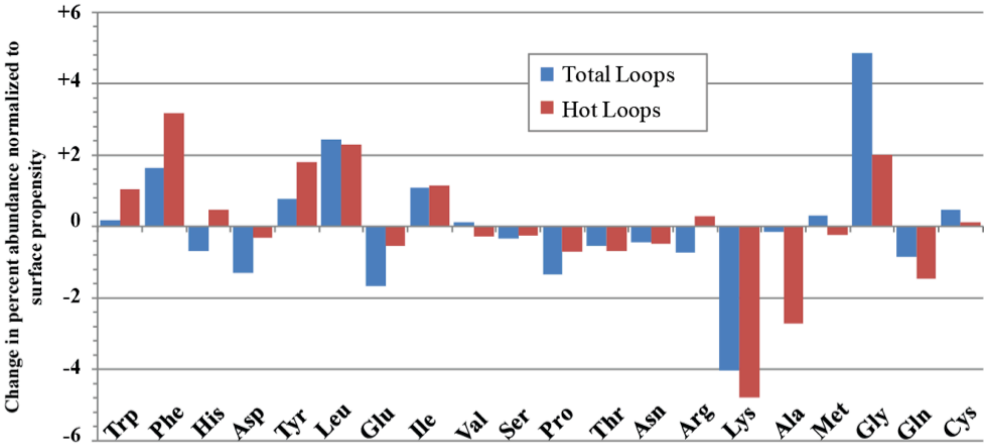


Figure 1.7 Comparison of percent abundance between total interface loos (blue) and hot loops (red) normalized to surface propensity as measured by Janin *et al*. [92]

1.5 Function of proteins present in loop sets

The 25,005 interface loops identified by LoopFinder covered 499 different functional classes of proteins from all kingdoms of life, as well as viral proteins. The 1,407 hot loops, by contrast, covered only 132 functions of proteins. The functional class with the highest number of hot loops was the hydrolases, isomerases, lyases, oxidoreductases, transferases, and ligases; this is not surprising given their prominence in the PDB as a whole. When representation of each enzyme type within the interface loop set and the hot loop set were compared to representation in the input PDB set (Figure 1.8), it is clear that oxidoreductases were over-represented in the interface loop set, and were even more prominent in the hot loop set. Lyases and isomerases showed a similar, but less prominent, trend, and hydrolases showed the opposite trend. We concluded that oxidoreductases form loop-mediated PPIs more often than other types of enzymes. Further analysis may illuminate the reason for this over-representation.

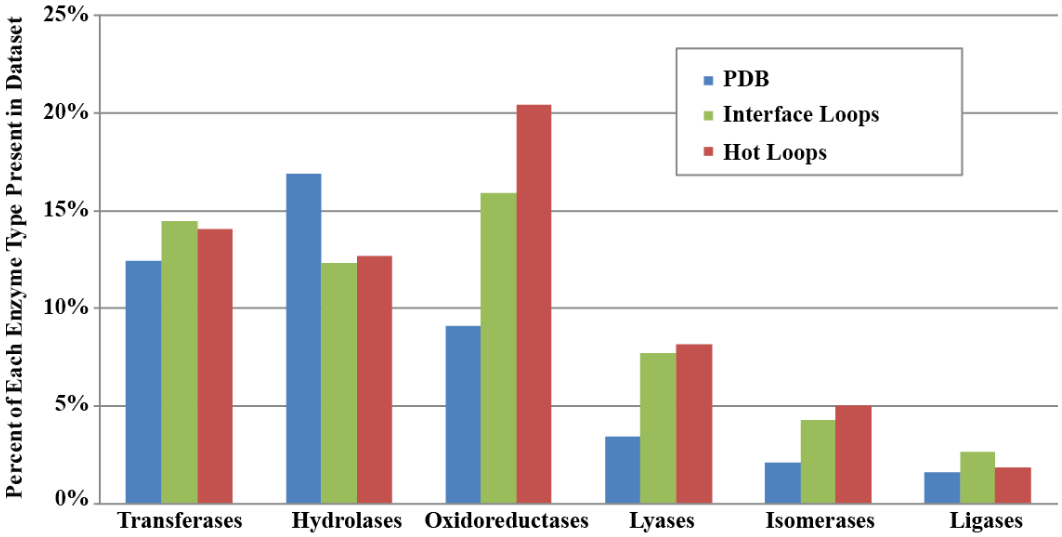


Figure 1.8 Analysis of enzyme categories in the dataset. Percentages of each enzyme type are shown for the total PDB (blue), the 25,005 interface loops identified by LoopFinder (red), and the 1,407 hot loops (green).

2 Chapter 2 : Updating LoopFinder with New Hot Loop Criteria and Total Energy Analysis

2.1 Update of hot loop criteria after the first publication

While our original LoopFinder parameters offered many viable starting points for the design of cyclic peptide inhibitors of PPIs, a comparison between the energy of the loop to the energy of the whole interface was lacking. Our original hot loop parameters were: an average $\Delta\Delta G_{res} \geq 1$ REU, there being ≥ 3 hot spot residues on a loop, and the loop possessing two consecutive hot spot residues (where a hot spot was defined as ≥ 1 REU). The parameters yielded a set of loops with a large overlap between the second and third hot loop parameters. We expect that loops with two adjacent hot spots may be best for mimicking with small molecules, since this type of interaction is concentrated over a small surface area. One of the goals of developing potent cyclic peptides targeting PPIs is to target extensive surfaces and recapitulate much of the binding interaction. As a result we aimed to identify the contribution of the binding energy of the peptide epitope to the overall binding energy of the natural interaction.

As previously observed by computational methods identifying hot segments, many of these hot segments recapitulate at least 50 % of the calculated binding energy of the full protein-protein interaction.[33] We sought to replace the two consecutive hot spot parameter and include a new parameter that accounted for the percent contribution of the loop to the total interface energy. To do so, we needed to rerun the LoopFinder program to retain the total interface energy calculation. Concurrently, we sought to make improvements to the LoopFinder program for running on the Tufts University Computing Cluster for more high-throughput submission of LoopFinder runs and computational alanine scanning mutagenesis.

2.1.1 Unix Compatible LoopFinder

The LoopFinder program was originally written by Brad Sheneman in the C++ language for running on a Microsoft based system. To be compatible for use on a computer cluster, the LoopFinder program was recompiled for a Unix based operating system. This converted the LoopFinder.exe program to the *LoopFinder_Unix* executable file. In addition, to permanently preserve our data on the Tufts Cluster system, we secured a permanent directory for LoopFinder on the Tufts Research Cluster located at */cluster/tufts/loopfinder*.

Within the */cluster/tufts/loopfinder* directory, new folders were made to contain the PDB files organized by number of chains in the biological assembly *(/oligomeric_state_2*, */oligomeric_state_3*, */oligomeric_state_4* etc.). PDB files from the RCSB were downloaded in May 2014 by chain number (2-10), according to the criteria previously mentioned for our initial LoopFinder search (representing all multi-chain structures with ≤ 4Å resolution and < 90% sequence identity) resulting in 27,938 PDB entries with chains at interfaces (Figure 2.1).

Now that the LoopFinder program was compatible with Unix on the Tufts Cluster, we were able to develop a series of simple shell scripts to automate the

LoopFinder process and enable multiple LoopFinder runs to be conducted in a batch format. For the most recent LoopFinder run we set the loop parameters: length of 4 to 8 amino acids, 80 % of the atoms within the loop must reside within 6.5 Å of the partner chain, and a termini distance cutoff of 6.2 Å was used with an *lfactor* to adjust for shorter loops. Running these parameters on the set of 27,938 PDBs resulted in 229,519 loops, located on 40,781 protein chains (Figure 2.1).

The 40,781 interfaces were submitted for computational alanine scanning using PyRosetta as previously discussed. An advantage of running these calculations on the cluster was that now 1,000 interfaces at a time could be submitted for calculation, greatly reducing the total calculation time. In addition, now that all of the information was permanently saved in the */cluster/tufts/loopfinder* directory, once an interface was calculated the data is preserved. As a result, if we decide to change the loop parameters, as long as a loop was previously found at that interface the energies can be culled from the results folder immediately without the need for re-calculation. As more loops are identified on new interface the database will continue to grow, and previously calculated interface energies and hot spot energies can be retrieved without the need for recalculation.

An initial removal of loop redundancies was carried out to remove identical loops from homomultimeric interfaces that persisted due to variations in chain naming and numbering within the PDB files. We maintained overlapping and nested loops within this dataset until hot loop analysis was done. We could later weed out the loops with lower percent energy contributions. This pared the list of loops down to 83,170, which we then analyzed for their potential to be designed into cyclic peptides for PPI inhibition.

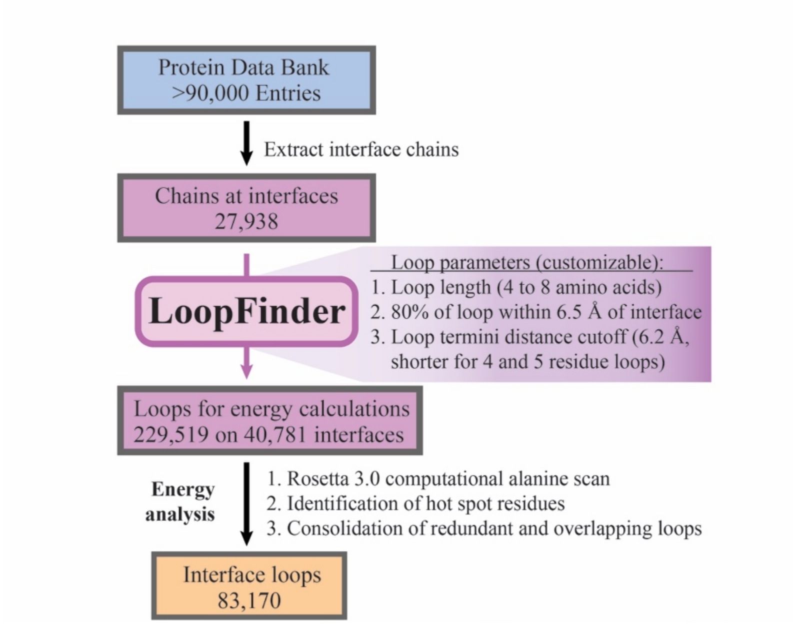


Figure 2.1 LoopFinder re-run on 27,938 PDB files. Using loop parameters similar to the previous iteration of LoopFinder we recalculated interface and loop energies for 229,519 loops on 40,781 interfaces. Removal of identical redundant loops resulted in 83,170 loops for analysis.

2.1.2 Choosing New Hot Loop Criteria

We aimed to better inform our hot loop parameters through an analysis of hot spot values calculated from structures of peptides bound to proteins. We could then use experimental binding data for comparison. PDBs from the RCSB were downloaded for which the description contains the keyword peptides. These PDBs were downloaded and analyzed using LoopFinder. Peptide epitopes were identified by changing the maximum number of residues to 12 amino acids. In addition, the maximum distance from end-to-end was set to 20 Å to allow for the identification of linear peptides. The identified peptide-protein interfaces were then subjected to PyRosetta analysis and subsequently compiled into a data set of loops located at these peptide-protein interfaces. Manual analysis of this loop set identified biomedically relevant structures of known peptide and cyclic peptide binders to protein targets that act as PPI inhibitors. A handful of representative peptides were analyzed in more detail for which there was sufficient experimental hot spot and binding data to better calibrate our expectations of computational hot spot values for loops that we wish to directly translate into peptide inhibitors. This type of analysis is also helpful in providing a comparison between calculated average $\Delta\Delta G_{res}$ values and experimental K_d of the peptide.

We selected five representative interactions to better calibrate the values used for hot loop sorting. The first interaction that came to light was that of the ABC transporter P-glycoprotein, CmABCB1, with a macrocyclic peptide identified through mRNA display techniques, aCAP (Figure 2.2A).[96] The aCAP peptide has an extended loop region of six amino acids that reside at the interface. The valine and tyrosine residues were highlighted experimentally as has having important contacts with the CmABCB1 binding partner (Figure 2.2A).[96] From the calculation of the peptide-protein interaction, the identified loop has a total loop energy of -3.4 REUs, contributes 100 % of the hot-spot energy of the interface, and has an average loop energy of -0.26 REUs. Though this peptide has been shown to be a strong inhibitor of the ABC transporter, it would not have been identified as a hot loop using the previous requirements using the first LoopFinder hot loop criteria. Tyr7 was known to be essential for binding, and was calculated by PyRosetta to have a $\Delta\Delta G_{res}$ of -4.5 REUs, which identifies the need to redefine these highly negative values and convert them to a positive hot spot value. In addition, Ile5 with a $\Delta\Delta G_{res}$ of 0.78 REUs makes an important hydrophobic interaction with the protein partner and should be counted as a hot spot. The analysis of the aCAP-CmABCB1 interaction suggests a need to convert strongly negative $\Delta\Delta G_{res}$ values to hot spots, as well as to reduce the cutoff for a hot spot to a value lower than 1.0 REUs.

The second interaction relevant to our LoopFinder goals was the interaction between a bicyclic peptide UK729 bound to the Urokinase-type plasminogen activator (uPA) serine protease (Figure 2.2B).[97] In this case LoopFinder and PyRosetta identified two amino acids on the loop region of UK729 that interact significantly with the protease. These are the threonine and arginine, which are residues commonly found as hot spots (Table 2.1). What can be learned from this interaction is that the total loop energy of a loop comprising a stretch of seven amino acids from Gln4 to Arg10 is 4.8 REUs and has an average $\Delta\Delta G_{res}$ of 0.67 REUs. This cyclic peptide comprising the aforementioned hot spots has a K_i of 11.56 μM.[97] Using this information, we hypothesize that we may need to change the average $\Delta\Delta G_{res}$ hot loop criteria to at least 0.67 REUs.

A third example is the interaction of SIRT2, a protein responsible for deacetylation of RIP1 (receptor interacting protein 1) that controls many cellular processes, with a macrocyclic peptide S2iL5 that binds with a K_d of 1.0 nM (Figure 2.2C).[98] While this interaction does not directly translate in terms of translating the 1 nM binder to hot spot calculations due to its acetyl-lysine mimic moiety that interacts with the enzyme active side, the flanking specificity amino acids can shed light on LoopFinder hot loop criteria. In this case a loop region from Try4 to Arg9 contains only one hot spot (Table 2.1) at the Tyr4 position, but has a large contribution from Arg9 at 0.71 REUs. This further suggests the need to decrease the requirement of hot spot residues within the loop regions. The total loop energy was 3.3 REUs, which contributes all of the calculated hot spot energies of the peptide. The average energy per residue is 0.5 REUs, which is underinflated due to the lack of PyRosetta's ability to calculate the energy for an unnatural amino acid at the trifluoro-acetyl-lysine residue. This data confirms a trend in which amino acids located on peptide inhibitors of PPIs contain residues that will have calculated hot spot energies of around 0.7 REUs and an average loop energy of at least 0.5 REUs.

The fourth example is that of an ExoS cyclic peptide mimic that binds a 14-3-3 protein. This peptide was developed as a therapeutic strategy for neutralizing the pathogenic bacteria virulence factor ExoS, a common protein in drug resistant strains of bacteria (Figure 2.2D).[99] This peptide uses a hydrophobic cross-link to stabilize a ten-amino-acid α-helical peptide region. The peptide is based on the natural peptide sequence and shows a K_d of 0.25 μM, compared to 1.15 μM for the uncyclized linear sequence. Analyzing the cyclic peptide via LoopFinder and PyRosetta yields a six amino acid stretch making up the core of the peptide, contributing the majority of the energy to the total interaction (Table 2.1). This peptide has a total loop energy of 3.9 REUs and an average $\Delta\Delta G_{res}$ of 0.65 REUs. While not a loop or turn based peptide, this is the exact type of interaction we aim to recreate using hot loops, and again due to low average $\Delta\Delta G_{res}$ value would not have registered as a hot loop. This trend of average energy for five to seven amino acid stretches of amino acids seems to be common among a variety different peptide-protein PDBs. In addition, the other obvious trend is that they seem to have one hot spot and at least one other residue with a calculated $\Delta\Delta G_{res}$ around 0.7 kcal/mol.

One of the most convincing and informative analyses was carried out on SPSB2, an E3 ubiquitin ligase that contains a well-known SPRY protein interaction domain. SPSB2 was co-crystallized with the DINNN motif of the DEAD-box RNA helicase VASA peptide (Figure 2.2E).[100] This linear peptide loop epitope extracted from a larger protein would be a desired result from our LoopFinder analysis of the parent PPI. The linear version of the peptide (Ac-DINNN-NH_2) was shown to have a K_d of 318 nM by SPR. Subsequent efforts to develop a stronger binder led to the synthesis of the natural domain flanked by two cysteine residues. This was used to make a disulfide linked cyclic peptide (Ac-CVDINNNC-NH_2) which has a K_d of 4.4 nM by SPR. The overall process for turning a loop epitope into a cyclic peptide an improving affinity by two orders of magnitude is almost the exact design procedure we would undertake for the development of hot loops into cyclic hot loop inhibitors. Computational alanine scanning of the loop showed only one residue with a $\Delta\Delta G_{res} > 1$ REU (Table 2.1). However, the other characteristics of the loop did look promising and follow the trends seen in previous examples. The total hot loop energy was 3.2 REUs, with an average $\Delta\Delta G_{res}$ of 0.65 REUs. This example strongly suggested that an average $\Delta\Delta G_{res}$ of 0.65 will make an excellent requirement for identifying hot loops.

The above observations shifted out perception of hot loops. If the main goal of LoopFinder is to develop inhibitors and like the ones mentioned above, we can use these interactions as benchmarks to redefine hot loop criteria. One common observation was that these peptides contain amino acids that have calculated $\Delta\Delta G_{res}$ of around 0.7 REUs. This value is lower than our previous $\Delta\Delta G_{res}$ value of 1 REU. It appears that the lower $\Delta\Delta G_{res}$ values calculated by Rosetta may be an underestimation of experimental values and a $\Delta\Delta G_{res}$ of 0.7 REUs is enough to constitute a hot spot in situations of peptide loop. As a result, we decided to change the criterion for a hot spot to 0.6 REUs.

The second issue to be resolved when selecting hot loop criteria was the presence of negative $\Delta\Delta G_{res}$ values as calculated PyRosetta. Manual analysis of loops containing negative values, with an example being the analysis of the aCAP peptide introduced above, revealed that residues with highly negative values ($\Delta\Delta G_{res} \leq -2.0$ REUs) should be included as hot spots. These strongly negative values are often due to a hydrogen bond between the loop residue and the protein binding partner that is slightly longer in distance than the ideal hydrogen bonding distance measurement.[25] This allows for water to satisfy the hydrogen bond in a more optimal fashion when the side chain is mutated to alanine. For the purposes of peptide design however, even a sub-optimal hydrogen bond will still promote the overall binding interaction. Therefore, we subsequently converted all calculated negative $\Delta\Delta G_{res}$ values ≤ -2.0 REUs to 1.0 REUs. Calculated energies -2.0 REUs $< \Delta\Delta G_{res} < 0.0$ REUs were set to equal 0.0 REUs. Manipulation of negative values that result from the PyRosetta alanine scan also solves an additional problem. Negative total energies arose when we had calculated total interface energy as a sum of all $\Delta\Delta G_{res}$ contributions from individual amino acids. Fixing these negative $\Delta\Delta G_{res}$ values to 1.0 REU or 0 REUs allows for the calculation of positive interaction energies for all PPIs in the loop set.

We can then introduce a new criterion for hot loop assignment and included a percent interface energy parameter. In this case, the total hot spot energy of the loop itself is compared to the total energy of all interface residue energies calculated by PyRosetta computational alanine scan. This was a logical requirement as the subsequently designed peptides will have to compete with the natural protein partner for PPI inhibition. This type of analysis was enabled only after conversion of all residue energies to a non-negative value. The inclusion of negative values throughout the PyRosetta computational alanine scan results yielded many total interface energies that were negative values. Now with only positive energy interactions of protein partners in the data set, this percent interface energy can be explored as a hot loop criterion. We reasonably expected peptides designed from hot loops comprising at least 50% of the hot spot energies present at the interface will be able to inhibit the native PPI to at least some extent.

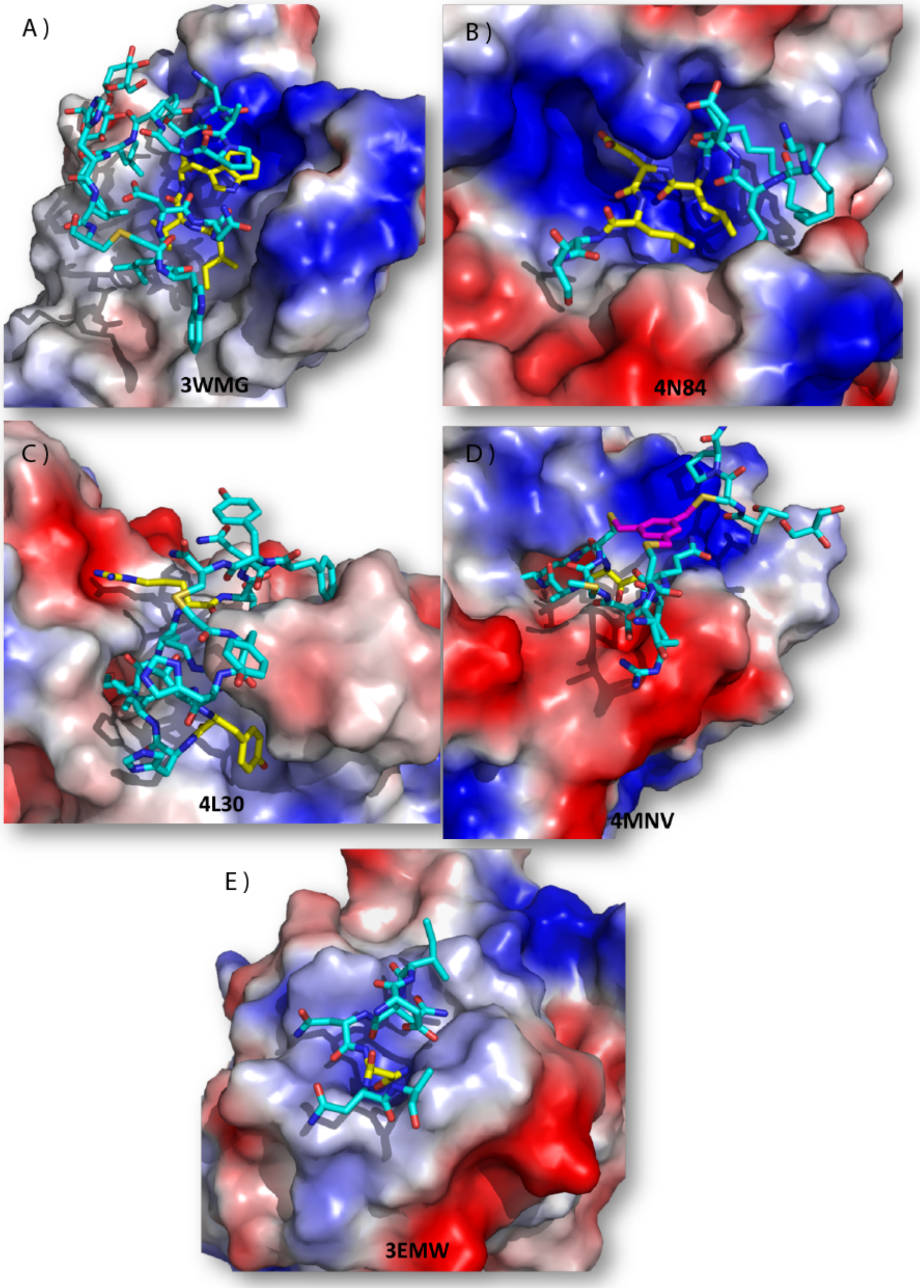
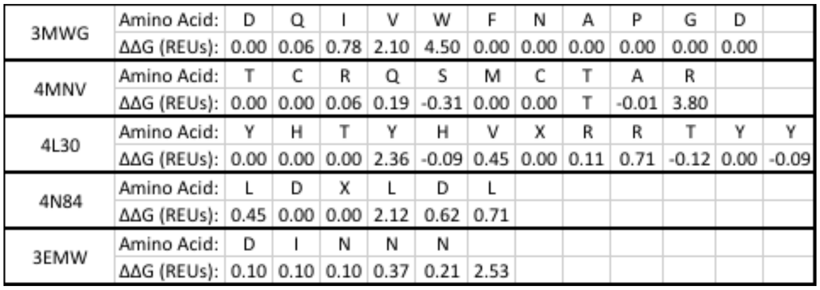


Figure 2.2 Peptides bound to their protein binding partner to inform hot loop criteria. A) aCAP bound to the ABC multidrug transporter (3WMG).[96] B.) A tris-alkylated cyclic peptide bountd to Urokinase-type plasminogen activator uPA (4MNV).[97] C.) A cyclic peptide bound to SIRT2 lysine deacetylase (4L3O).[98] D.) A stapled peptide based on ExoS bound to 14-3-3ζ (4N84).[99] E.) A smaller VASA peptide bound to SPSB2 (3EMW).



3MWG	Amino Acid:	D	Q	I	V	W	F	N	A	P	G	D	
	ΔΔG (REUs):	0.00	0.06	0.78	2.10	4.50	0.00	0.00	0.00	0.00	0.00	0.00	
4MNV	Amino Acid:	T	C	R	Q	S	M	C	T	A	R		
	ΔΔG (REUs):	0.00	0.00	0.06	0.19	-0.31	0.00	0.00	T	-0.01	3.80		
4L30	Amino Acid:	Y	H	T	Y	H	V	X	R	R	T	Y	Y
	ΔΔG (REUs):	0.00	0.00	0.00	2.36	-0.09	0.45	0.00	0.11	0.71	-0.12	0.00	-0.09
4N84	Amino Acid:	L	D	X	L	D	L						
	ΔΔG (REUs):	0.45	0.00	0.00	2.12	0.62	0.71						
3EMW	Amino Acid:	D	I	N	N	N							
	ΔΔG (REUs):	0.10	0.10	0.10	0.37	0.21	2.53						

Table 2.1 The PyRosetta energy calculations for peptide-protein examples. In all cases, the PDBID and amino acid sequence for the peptide loop are in red and the corresponding energy calculations are located below each amino acid. The core loop located on each peptide is highlighted in yellow.

2.2 Results from Latest LoopFinder Run

We developed a new set of hot loop criteria based on our analysis of existing peptide-protein interactions. First, we relaxed the criterion for the strength of a hot spot to $\Delta\Delta G_{res} \geq 0.6$ REUs. This was informed by our previous analysis of peptide-protein interactions from the PDB, in which we observed that many high-affinity peptides did not have many hot spot residues with $\Delta\Delta G_{res} \geq 1.0$ REUs, but did have many with $\Delta\Delta G_{res} \geq 0.6$ REUs. We discarded the consecutive-hot-spot criterion due to the potential for such smaller surface binding pockets to be targeted with small molecules, and replaced it with an estimation of the proportion of the total binding energy attributable to the loop. This was done using a new calculation of total interface energy. Total loop $\Delta\Delta G$ energies were calculated by summing up individual $\Delta\Delta G_{res}$ values and we estimated total interface energy by adding up $\Delta\Delta G_{res}$ values for all residues at the interface. This produced a proportional measure of the percent contribution of the loop to the total interface energy. Using this most recent parameter, we can identify as particularly promising any loops that comprise a large proportion of the total interface energy, and is tunable from 25-100 % contribution. Different users can consider these and other criteria. Ultimately, for the development of cyclic peptides and macrocycles, we chose to identify hot loops as those that have either an average $\Delta\Delta G_{res} \geq 0.6$ REUs, those with three or more hot spots (where a hot spot is defined as ≥ 0.6 REUs), or those that comprise $\geq 50\%$ of the total interface energy. This identified 7,225 hot loops from the total interface loop set of 83,170 (Figure 2.3). Figure 2.4 shows a Venn diagram that illustrates how these criteria identified these loops. As the diagram shows, we recommend using three or more criteria to sort through a large interface loop set, in order to identify those interface loops with multiple desired properties for the proposed application.

For the new set of 83,170 interface loops and 7,225 hot loops, we have quantitated amino acid distribution within the loops, as we had done for the initial LoopFinder run. As seen in Figure 2.5 and Figure 2.6, the new dataset shows similar trends, with a few key exceptions. These differences can mainly be attributed to the modification of amino acids with $\Delta\Delta G_{res} \leq -2.0$ REUs as now being characterized as contributing 1.0 REU. As a result, some amino acids that may have been inaccurately calculated in the previous set are currently better characterized as hot spot residues. The most notable shift can be seen in the increased abundance of valine and proline residues as hot spot residues. In the original hot loop set, we observed that proline was not commonly found as a hot spot, but when it was proline was likely to have a high $\Delta\Delta G_{res}$. We hypothesize that there may have been many proline residues in the previous set that had highly negative values due to inability to properly calculate the contribution of proline to the interaction. The same is seen with valine, where once negative values were adjusted, we see a result consistent with valine being used as a more common hot spot for PPIs. While these are not increased in abundance due to hydrogen-bonding effects, there has been a discussion of inaccuracies for proline and valine calculation due to issues with the interface repacking step within the PyRosetta program.

A list of all hot loops that fit all three hot loop criteria can be located in the Table 5.1. All hot loops identified using this approach, and hot loops from our previous LoopFinder run discussed in Chapter 1, are available on the LoopFinder website (loopfinder.tufts.edu). We hypothesize that these loops will serve as optimal starting points for cyclic peptide design. It was the goal for the remainder of this thesis work to demonstrate the development of a hot loop into a cyclic peptide inhibitor of the associated PPI.

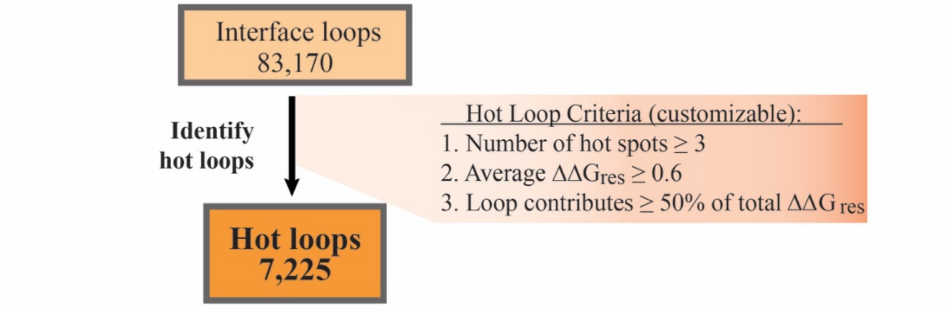


Figure 2.3 New hot loop criteria. We identified 7,225 hot loops that fit least one of the above criteria.

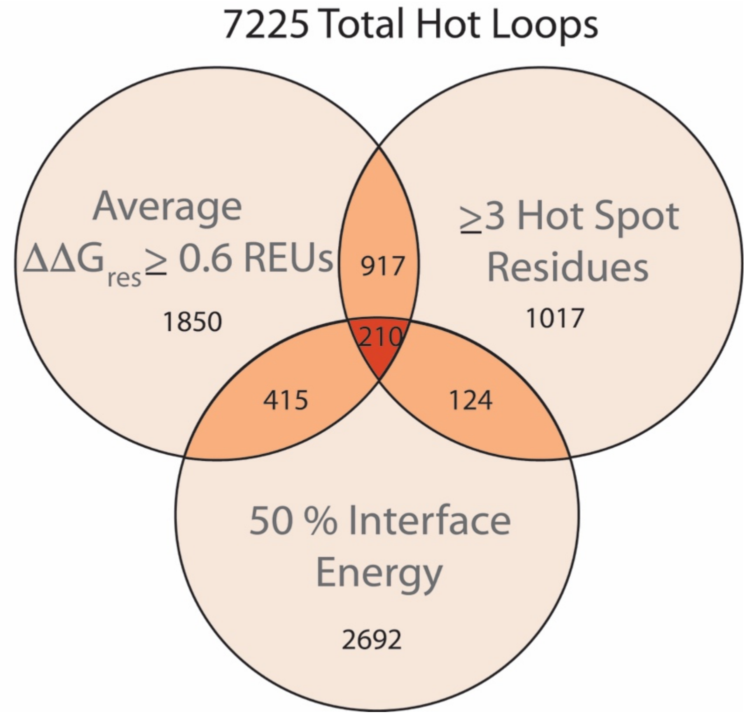


Figure 2.4 Hot loops organized by the criteria they satisfy. The central overlapping region contain loops that satisfy all three criteria and are best suited for cyclic peptide inhibitor design, possessing a high ligand efficiency. These are listed in Table 5.1.

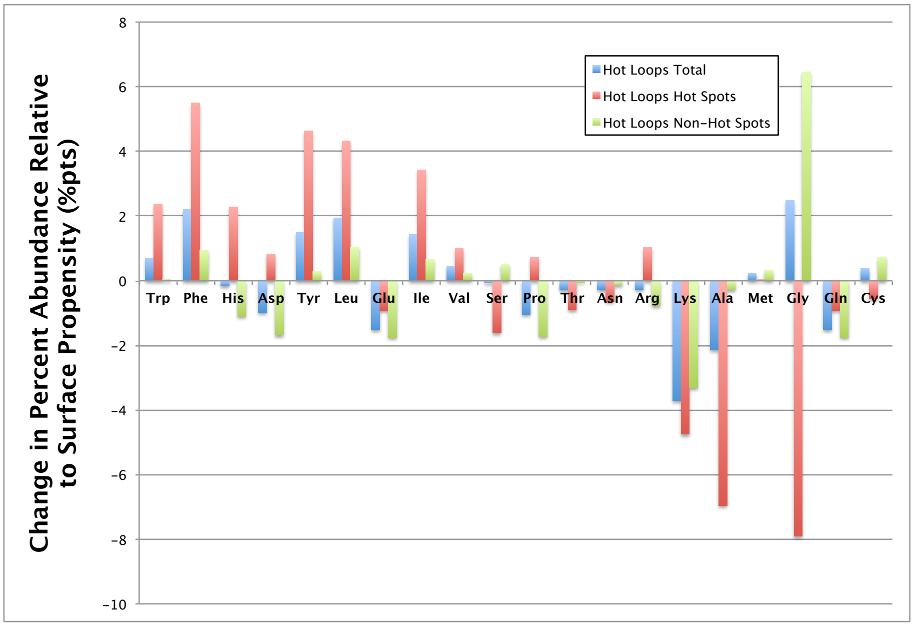


Figure 2.5 An analysis of the change in percent abundance of each amino acid relative to its natural surface propensity is shown for hot loops (blue), just hot spots within hot loops (red) and non-hot spots within hot loops (green)

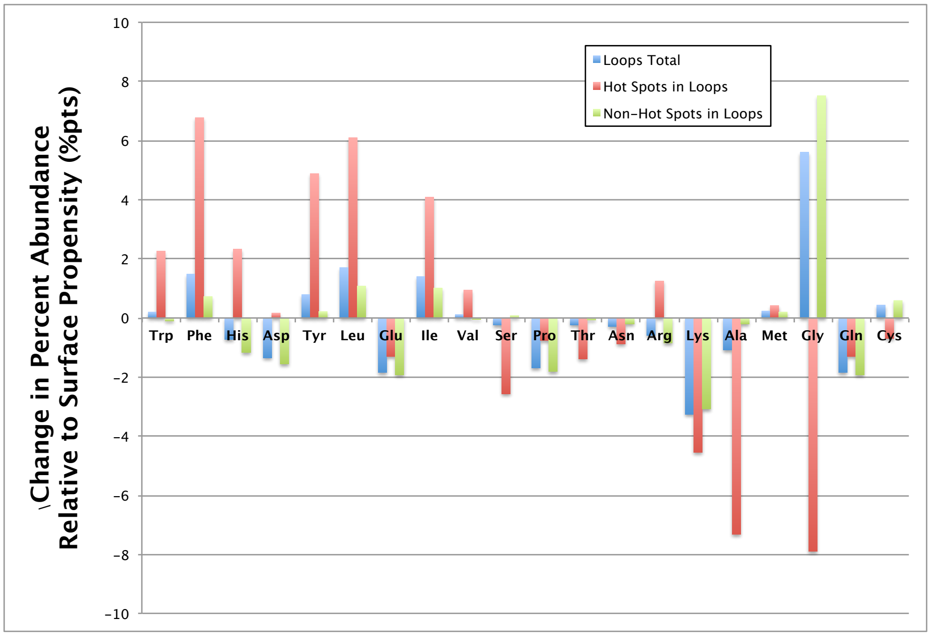


Figure 2.6 An analysis of the change in percent abundance of each amino acid relative to its natural surface propensity is shown for all interface loops (blue), just hot spots within all interface loops (red) and non-hot spots within all interface loops (green).

2.3 Proof of Concept for LoopFinder's Ability to Identify Leads for Cyclic Peptide Design

2.3.1 LoopFinder Identification of a Classic PPI

To judge the potential predictive value of our hot loop set, it was essential that we properly identified known loops that mediate PPIs that had been reported in the literature. One of the most thoroughly-studied PPIs has been the interaction between human growth hormone (hGH) and its soluble extracellular receptor domain, human growth hormone binding protein (hGHbp).[1,9,52,101-104] In classic work establishing the relevance of hot spot residues, mutational analysis revealed that of the whole hGHbp surface, only 31 side chains on the receptor surface are involved in intermolecular contacts. Of these 31 amino acids, eight residues contribute 85% of the binding energy, with the two most essential residues being W104 and W169.[9] In addition, alanine scanning mutagenesis studies of hGH have established that there are three epitopes on hGH that are responsible for receptor binding: the helix 1 region, a loop region that connects helix 1 and helix 2, and the helix 4 region.[52,102-104]

We identified two hGH:hGHbp structures that contain hot loops (Figure 2.7). The first consists of the native hGH and hGHbp in a 1:2 complex (1HWG), and the second is a mutant interface that was re-optimized by phage display (1AXI).[95] LoopFinder identified two hot loops within these structures, P61-E66 of hGH and I165-M170 of hGHbp. The most critical hot spot within the hGH loop, R64, was accurately identified by our computational alanine scan, and contributed to the inclusion of hGH P61-E66 among the hot loops. Likewise, the most critical hot spots within the hGHbp I165-M170 loop, I165 and W169, were also identified by the computational alanine scan and contributed to its inclusion among the hot loops. Overall, the ability to compare our results to such a well-understood PPI speaks to the robustness and predictive power of Rosetta-based computational alanine scans and of LoopFinder, as this interaction was also used to evaluate the accuracy of computational alanine scanning using Rosetta.[24,25]

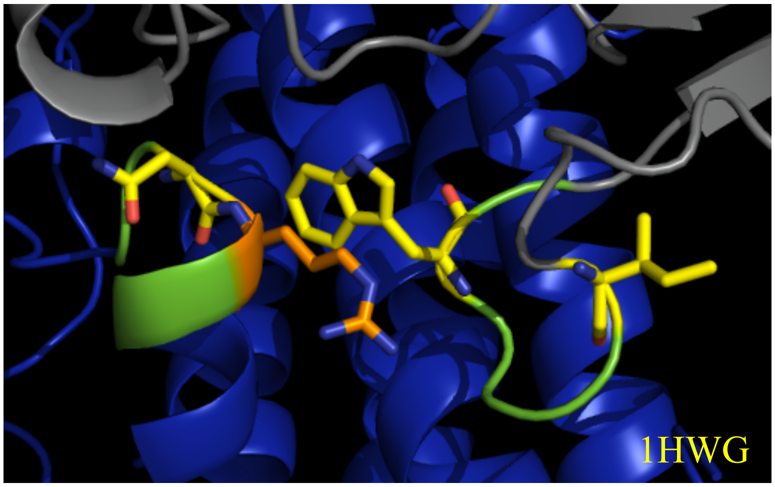


Figure 2.7 Two hot loops within the hot spot region of the hGH:hGHbp interface. LoopFinder identifies a hot loop (green) on the hGHbp (gray) and another hot loop (green) on the hGH receptor (blue). Hot spots are colored in yellow and orange.

2.3.2 Nrf2:Keap1 Interaction

We also identified a loop not only previously identified to be essential for the PPI, but also already developed into a cyclic peptide inhibitor of the interaction. Cellular oxidative stress is associated with many disease states, including inflammation, cardiovascular disease, cancer, and neurodegenerative diseases.[105,106] A coordinated program of protection from oxidative stressors is regulated by the transcription factor Nrf2 (Nuclear factor erythroid-derived-related factor 2). Nrf2 is responsible for transcriptional activation of ARE (antioxidant response element) genes, which increase cellular antioxidants, glutathione transferases, glutathione synthesis, toxin export, and inhibition of cytokine-mediated inflammation.[105] Under normal cellular conditions, Nrf2 remains at low levels through its PPI with the Kelch-like ECH-associated protein 1 (Keap1), which targets Nrf2 for ubiquitin-mediated degradation. Keap1 acts as an environmental sensor. It contains 27 cysteines that, when oxidized, cause a conformational change that disrupts Keap1 recognition of Nrf2.[107] In this manner, increased cellular oxidants lead to increased concentrations of intracellular Nrf2 and activation of ARE genes.

The Nrf2:Keap1 complex's critical role in the cell's antioxidant response makes it a therapeutically relevant target. Targeting the cysteines of Keap1 with electrophilic natural products and cyano-enone Michael acceptors was a therapeutic strategy for treating leukemia and solid tumors that reached Phase I clinical trials.[108-111] However, a lack of specificity associated with this strategy proved to be a major drawback. Inhibition of Nrf2 degradation by blocking the Nrf2:Keap1 PPI has the potential to be a more selective method of Nrf2 activation. LoopFinder identified a six-residue sequence, D77-E82 (DEETGE), from the crystal structure of Keap1 bound to a Nrf2-derived peptide (2FLU) (Figure 2.8).[112] Computational alanine scanning of the hot loop identified E79 and E82 as critical for binding ($\Delta\Delta G_{res}$ of 4.43 and 3.81 REUs respectively), and the loop was categorized as a hot loop. This β-hairpin loop was previously identified as critical for this PPI, and a 16-mer truncation of Nrf2 containing this loop (A69-L85) was shown to bind Keap1 with a K_d of 20 nM, retaining much of the affinity of the full-length protein (K_d of roughly 5 to 9 nM).[112,113] Experimental mutagenesis experiments have also demonstrated the importance of E79 and E82 within the loop, and also showed that truncations down to an optimized 7-mer sequence (Ac-DPETGEL-OH) retains an affinity in the 100 nM range.[114] Finally, while details have not been reported, a recent patent filing discloses a disulfide-bridged cyclic peptide inhibitor of Keap1, with a reported IC_{50} of 14.7 nM, that contains the hot loop identified by LoopFinder.[115] Small molecules have been reported with Keap1 affinities in the μM range, and direct Nrf2•Keap1 inhibitors show longer-lasting effects in ARE-dependent luciferase assays compared to electrophilic Keap1 inhibitors.[114] Taken together, all these data indicate that targeting Keap1 with cyclic peptides or other macrocycles may be an optimal strategy for Nrf2-mediated therapies. Thus, in the case of Nrf2:Keap1, LoopFinder identified a key β-hairpin loop that could be translated directly into cyclic peptide inhibitors with low nanomolar affinity. This directly illustrates the utility of LoopFinder as a resource for identifying hot loops as starting points for developing PPI inhibitors.

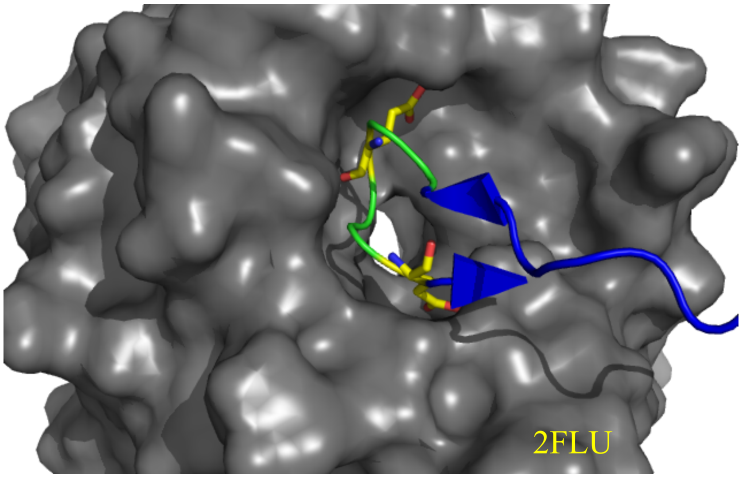


Figure 2.8 The hot loop (green) on the Nrf2 peptide (blue) bound to its Keap1 protein partner (gray). Hot spots are colored in yellow.

2.3.3 uPA:uPAR Interaction

While identifying loop peptides bound to proteins, we used a phage display bicyclic peptide that targeted Urokinase plasminogen activator receptor (uPAR) to help inform our hot loop criteria. We were further able to identify a hot loop residing on the natural protein target. LoopFinder identified the interaction between Urokinase plasminogen activator (uPA) and uPAR. This interaction, a validated drug target, is involved in the proteolytic regulation of the extracellular matrix, and its deregulation plays a major role in tumor progression.[116] One of the major interacting regions of uPA that is responsible for UPAR binding is the Ω-loop, a specific extracellular motif involving a β-hairpin, corresponding to residues S21 to I28.[117] A linear peptide meant to mimic this loop (VSNKYFSNIHW) was reported to have an K_d of 1.3 μM.[118]

LoopFinder was successfully able to identify the Ω-loop on uPA with very strong hot spot characteristics, making it unsurprising that this peptide would be an excellent peptide inhibitor of the PPI (Figure 2.9). LoopFinder identified the residues S21 to I28 (SNKYFSNI) containing four hot spot residues at positions S21 (ΔΔG = 1.12 REUs), Y24 (ΔΔG = 1.77 REUs), F25 (ΔΔG = 3.25 REUs), and I28 (ΔΔG = REUs). This Ω-loop has also already been developed into a disulfide cyclized peptide that was observed to have an IC_{50} of 40 nM for the uPA/uPAR complex.[118] This peptide, WX-360, was shown to prevent tumor growth in mouse models of human ovarian cancer.[118] It is also important to note that a bicyclic analog of this peptide that binds to the identical protein-binding site was used to inform the hot loop criteria. Using those definitions to faithfully identify the native PPI helps verify our approach for identifying hot loops capable of being developed into cyclic peptides.

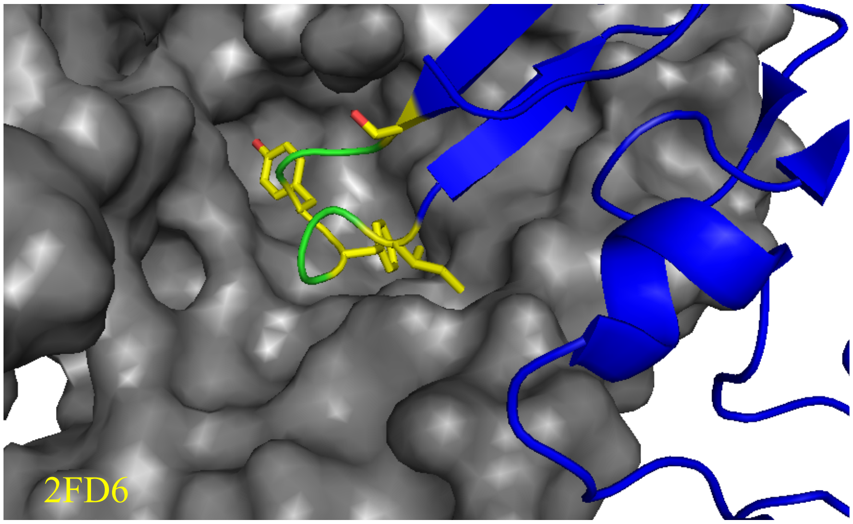


Figure 2.9 A hot loop (green) on the natural uPA protein (blue) was identified by LoopFinder bound to its receptor uPAR.

2.4 Alternative loop databases for antibody loops and DARPins

LoopFinder can also be extended to find larger loops than the original 4-8 amino acids. These may not be ideal for designing natural product-like cyclic peptides, but may be amenable to stabilization within larger protein scaffolds. Two well-studied structural scaffolds are antibodies and designed ankyrin repeat domain proteins (DARPins).[119,120] Antibodies are one of the classic scaffolds for stabilizing epitopes, partly because that is their original use in nature. Within their variable region, antibodies possess complementarity-determining regions (CDRs) or hypervariable loops. These loops are exposed on the antibody surface and are responsible for mediating the interaction between the antibody and the antigen. Within antibodies, there are six CDR loops. Three are located on the heavy chain variable domain (VH) and are designated H1, H2, and H3. These variable loops have specific sizes and lengths of the loops. H1 loops are mostly 10 to 16 amino acids. H2 loops are usually of length 7 or 8. H3 loops are more diverse in length, but are typically 7 and 16 residues. There are also three CDR loops on the light chain variable domain (VL), which possess similar loop lengths to their VH counterparts.[51]

The above loops were used to design loop definitions for a broader LoopFinder search. LoopFinder was run on the previously described set of 27,938 PDBs using the parameters outlined in Table 2.2. Three distinct LoopFinder runs were carried out. The first used parameters that would include H1 and L1 loops. The second used parameters consistent with H2, L2, and L3 loops, while the third search identified loops with parameters that match H3 loops alone. Because these loops are going to be mainly used for grafting, we relaxed the minimum percent of the loop necessary to reside within 6.5 Å of the interface. Distance measurements were taken from an analysis of PDB entries of antibodies.

DARPins presented a second class of scaffold proteins that could be used for grafting of longer loops. Pluckthun *et al.* have developed LoopDARPins as a scaffold for loop presentation and have extensively shown the capabilities of this structured ankyrin repeat scaffold for inhibition of PPIs.[120] The positions for replacement and grafting within these LoopDARPins were used to inform the LoopFinder search for loops of similar length and end-to-end distance (Table 2.2).

Using these LoopFinder parameters, a large number of loops were identified (Table 2.3). These loops are still being processed and analyzed. The relaxed end-to-end distance constraints exponentially increases the occurrence of overlapping and nested loops. The associated energy values for the residues within the loops and the interfaces have also been obtained. Further development is needed to identify the proper hot loop criteria for this loop set. It is unlikely that the same hot loop criteria that is necessary for small loops is also the proper fit for loops of this size.

Table 2.2 Parameters for alternate loops sets of CDR loops and DARPin loops.

Loop Parameters for Newest Loop Set of Antibody Loops and Darpin Loops							
Loop Type	-bdist	-minres	-maxres	-minperc	-maxdist	-lfactor	-nchain
H1	6.5	10	17	50	13	1	x
H2	6.5	7	13	50	6	1	x
H3	6.5	5	16	50	10	1	x
L1	6.5	10	17	50	13	1	x
L2	6.5	7	13	50	6	1	x
L3	6.5	7	13	50	6	1	x
Darpins	6.5	7	20	50	8	1	x

Table 2.3 Loops identified for each class of CDR or Darpin loop.

Loops at Interfaces				
	H2_L3_L2_loops	H1_L1_loops	H3_loops	Darpins
oligomeric_state_2	107784	983042	671848	479626
oligomeric_state_3	45330	368442	102701	196069
oligomeric_state_4	423708	1053388	241718	397239
oligomeric_state_5	13040	112915	15285	57364
oligomeric_state_6	55123	508699	71618	244107
oligomeric_state_7	6362	46285	4006	25459
oligomeric_state_8	30735	290896	38149	137039
oligomeric_state_9	4549	38087	4265	18820
oligomeric_state_10	5779	233051	27134	88320
Loops	692410	3634805	1176724	1644043

2.5 Materials and Methods: Running LoopFinder on the Cluster

As previously discussed, a dedicated LoopFinder directory on the cluster was secured through Tufts UIT to allow for a permanent location of all programs, calculations and data. The name of the directory is */cluster/tufts/loopfinder* and within this directory is located the *LoopFinder_Unix* program. Because LoopFinder needs to run separately on PDB files of different chain number, new directories were added named */cluster/tufts/loopfinder/oligomeric_state_2*, *cluster/tufts/loopfinder/oligomeric_state_3* etc. with a new folder for each oligomeric state. An outline of the overall file organization is shown in Figure 2.10.

The RCSB is a vast database of protein structure data with an advanced search tool that can be easily used to download large sets of PDBs that meet the user-defined parameters of the search.[121] In our LoopFinder run, we were interested in finding loop-mediated protein-protein interactions, and as a result identified a list of PDBIDs that had between 2 and 10 chains in the biological assembly. We excluded structures with > 90% sequence identity in order to avoid redundance. The list of PDBIDs was organized into separate text files based on the number of chains and these text files were deposited into the corresponding oligomeric state folder on the cluster. Bulk download of PDB files from the RCSB was done using the *pdb_download.py* script developed by the Harms Lab at *github.com/harmslab/pdbtools*.

Once the PDB files were downloaded, LoopFinder was executed using an *LF_run.sh* script. Several parameters within this script were edited to customize the loop parameters (Figure 2.11). The first parameter was the distance for categorizing which chains within a multi-chain assembly are in contact, and thus make up a binary interface. In the *LF_run.sh* script, this is the *bdist* parameter. We used a 6.5 Å distance measurement (*bdist 6.5*), whereby if any atom within a chain is ≤ 6.5 Å away from any atom within another chain, the chains are characterized as having an interface. This value was chosen based on the similar distance term used by PyRosetta's *ala_scan.py* program to define interface residues.[24,25]

Next, the three parameters that dictate the size and shape of the loop itself were set. Loop length was specified by minimum and maximum lengths in residues, and was set to *minres 4* and *maxres 8*. The proportion of the loop at the PPI interface, which specifies the percentage of the residues with at least one atom within the *bdist* distance of the binding partner, was set to *minperc 80*. Finally, the distance between the termini of the loop, which specifies a maximum distance in Å from the C_α of the first residue to the C_α of the last residue, was set to *maxdist 6.2*. The 6.2 Å value was estimated as the approximate length of a dipeptide linker which could readily be introduced to cyclize loop epitopes. Importantly, setting *maxdist* to *6.2* also avoids α-helices (α-helices of length 5-8 amino acids have end-to-end distances of 6.3 to 10.8 Å) and β-strands (β-strands of length 4-8 residues have end-to-end distances of 9.8 to 22.7 Å).

The cutoff of 6.2 Å was applied to loops of length 6, 7, and 8 amino acids, but was too long for shorter loops. For loops of length 4 and 5, we empirically adjusted the termini distance parameter to 3.92 Å and 4.9 Å, respectively. This was implemented using a "scaling factor", and in this case the *lfactor* was set to *1*. The *lfactor* is divided by an estimate for the average length of an amino acid in extended conformation, 3.5 Å, to yield a fraction. This fraction was used as an estimated limit for loop length compared to a fully extended structure, which allows for identification of shorter fragments with particularly closed conformations. For example, an *lfactor* of *1* yields a fraction of 0.28. If the maximum linear length of a 5-amino-acid peptide would be estimated at 17.5 Å, the *lfactor* would limit the length to 0.28 times 17.5 Å, or 4.9 Å. An *lfactor* of *1* was used because it limits the loop length for shorter loops to less than one-third of the total length of the segment if it were in an extended conformation. For less stringent end-to-end distances of small loops, the *lfactor* value can be increased.

LoopFinder can also be used to discover larger loops. Such loop sets would be useful for a variety of applications. For instance, sets of loops ranging from 7-20 amino acids were identified with the aim to be used for grafting onto loop DARPins (see Chapter 2.4).[120] Also, antibody hypervariable regions facilitate protein binding with loops of 5-17 amino acids, so this loop range was used to find a set of antibody-like loops.[51] Users can set *minres* and *maxres* to match their ultimate goals for peptide and protein design.

In early development of LoopFinder, we noticed a significant proportion of loops that were not major contributors to the interface, but happened to have one hot spot that was within interaction distance (6.5 Å) of the binding partner. We regarded these as hot spots that happened to be on loops, rather than hot loops suitable for macrocycle design. Thus, to ensure that most loop residues are in close proximity to the binding partner, an interface distance requirement *minperc 80* was introduced as a loop definition parameter. This requires at least 80% of the residues within the loop to have at least one atom residing within 6.5 Å of the binding partner. This parameter can be changed based on the user's goals and expectations. We set the 80% limit empirically, to best identify only those loops that make extended contact with the binding partner. Ultimately, this focused our loop searches on epitopes with high ligand efficiency.[122-124] Reducing *minperc* below *80* also identifies some internal loops that contribute to the hydrophobic core of their parent proteins. Computational alanine scanning of these internally-facing loops would likely produce residues with high $\Delta\Delta G_{res}$ values, but these would be due to destabilization of overall protein structure and not due to direct contacts with the binding partner.

For each individual LoopFinder run, the *nchain* parameter was set to the matching value for oligomeric state. For instance, when LoopFinder was run on

/oligomeric_state_2, the *nchain* parameter was set to *2*. For each oligomeric state folder on which LoopFinder was being run, a new executable script was made (*LF_run2.sh, LF_run3.sh…LF_run10.sh*) whereby the only changes were the PDB directory and *nchain* value. This enabled all LoopFinder runs to be done simultaneously on all PDB directories.

For each set of PDBs, LoopFinder located all interface loops based on the specified parameters, and then output a number of files into a new directory */cluster/tufts/loopfinder/oligomeric_state_#/results*. First, scrubbed versions of each PDB file with the header and remarks removed were produced. Next, LoopFinder saved individual PDB-style files for each binary protein-protein interface. For example the interaction between Nrf2 and Keap1 (PDBID: 2FLU) had two chains denoted P and X.[112] LoopFinder found a loop on chain P that contacts chain X, and made a new file called *2flu_XP_c.pdb*. If an input PDB file (*PDBID.pdb*) contained chains A, B and C and each shared an interface with each other chain, LoopFinder output separate PDB files that containd only chains A and B, only chains B and C, and only chains A and C (*PDBID_AB_c.pdb*, *PDBID_BC_c.pdb* and *PDBID_AC_c.pdb*). This was the necessary format for inputting PDB files into PyRosetta. In addition to the new PDB files, a text file was produced that identified each residue and assigned a binary value based on whether the residue resides at the interface (*PDBID_AB_res.txt*, *PDBID_BC_res.txt*, *PDBID_AC_res.txt*).

The third file produced was a text file called *Loops.txt*, which contained the actual list of loops identified by LoopFinder. The *Loops.txt* file is the primary output of LoopFinder, and contains all interface loops for further analysis. *Loops.txt* included the following data, in the following order: PDBID, the chain on which the loop resides, the number of residues within the loop, termini distance, the total number of residues on the chain, the identity of each residue in the loop, the identity of the partner chain, and the percent of the loop at the interface. The percent of the loop at the interface is the percentage of the atoms within the loop that are within the specified *bdis*t distance to another atom on the partner protein. A representative excerpt from a *Loops.txt* file is shown in Figure 2.12. The binary interface PDB files, the list of residues that indicates which are at the interface, and the *Loops.txt* files were outputted into directories in a manner that was ready for direct input into PyRosetta for computational alanine scanning. LoopFinder also produced a separate file to use as a command line input file for running PyRosetta, called *cmd_line_input.txt*. This file, an excerpt of which is shown in Figure 2.13, contains all information required by PyRosetta including the PDBID, the chains involved in the interaction, the interface cutoff value and the number of trials for calculation.

The PDB files output by LoopFinder were in the proper format for direct submission to computational alanine scanning mutagenesis via PyRosetta v2.012.[58] The *ala_scan.py* script was run and required a specific format of input commands in order to execute the calculation. These commands for running *ala-_scan.py* on a specific PPI are located in individual lines within the *cmd_line_input.txt* file. To execute these commands on a large scan on the cluster, *int_eng_run.sh* was written for submission of each individual alanine scan in batch format on the cluster. This allowed for the submission of up to 1,000 lines for calculation to run concurrently on various nodes on the cluster. Figure 2.14 shows this script and the changes that need to be made in order to run the script properly. Using the slurm cluster, the script was submitted with the command: sbatch --array=1-1000 *int_eng_run.sh* which allows for the submission of the first 1,000 lines of the *cmd_line_input.txt* file to the *int_eng_run.sh* script. The maximum allowed submission limit on the Tufts Slurm cluster is set at 1,000, and as a result the PyRosetta alanine scan jobs were submitted 1,000 interface batches. Due to an issue with the cluster, subsequent submissions instead of being submitted using the sbatch --array=1001-2000 *int_eng_run.sh* were submitted by making a new *cmd_line_input.txt* file that contained only lines 1,001 to 2000, entitled *cmd_line_input_1001_2000.txt* and the *int_eng_run.sh* script was edited to call on the newly created *cmd_line_input_1001_2000.txt* file. Submission to the cluster saved an immense amount of time due to the ability to run calculations simultaneously. Normally, each interface would take 10-15 minutes. As PyRosetta ran, a new directory was created */cluster/tufts/loopfinder/oligomeric_state_#/results/PyRosettaResults*. Within this directory, as PDB interfaces were calculated a new directory was made for each PBDID that contained the resultant energy calculations. The results of alanine scanning for each PPI of interest were output to separate files, *PDBID_ddg_A_B1.txt* (where A and B are the chains forming this binary interface, for example *2flu_ddg_X_P1.txt, 2flu_ddg_X_P2.txt … 2flu_ddg_X_P20.txt*) and averaged results are output into *PDBID_mean_A_B.txt* (for example, *2flu_mean_X_P.txt*). In our *cmd_line_input.txt* file we specified twenty trial calculations for each interface in order to increase the accuracy of the energy calculation. The number of trials can be reduced to save time.

Once all of the energy calculations were completed they were combined with their corresponding list of interface loops contained in the *Loops.txt* file. The *conglomerate.py* script, written by Michael Bird, was used to combine the computational alanine scanning mutagenesis data in *PDBID_mean_A_B.txt* files (for example, *2flu_mean_X_P.txt*) with the lists of interface loops in the *Loops.txt* file. This script compiled energy information for each loop in *Loops.txt*, including $\Delta\Delta G_{res}$ for each residue in the loop and the sum of all $\Delta\Delta G_{res}$ for the loop. Also built into this script was the conversion of negative values and capping of positive values, as discussed in section 2.1.2. The script then output the data into *conglomerated.txt*, which was converted to a CSV file using *CSVconverter.py*, producing *conglomerated.csv*.

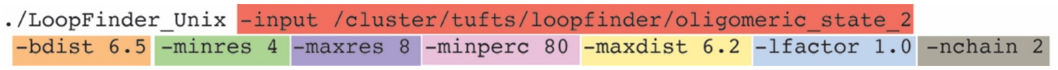
At this point interface loops were processed to check for errors and remove redundancies. *GapExcisor.py*, written by Michael Bird, searched *conglomerated.csv* to remove gaps that arise from non-sequential numbering in some PDB files, and yielded the file *conglomerated_Gapless.csv*. *RemoveRedundantBeforeHot.py*, written by Michael Bird, took the *conglomerated_Gapless.csv* as an input, located identical loops, and retained only the loop with the highest average $\Delta\Delta G_{res}$. This predominantly served to remove redundant loops from homomultimeric structures where multiple, identical protein chains are assembled symmetrically. All together, these redundancy checks winnowed the overall number of interface loops by roughly 60%. The data within the *conglomerated_Gapless_NonRedundant.csv* file was then input for the next step: identifying the loops of greatest interest based on “hot loop” criteria.

The python script *HotLoopAssignment.py*, written by Michael Bird, utilizes the user-defined hot loop criteria of interface energy percentage (%) and hot loop cutoff (where the # is defined as REUs*100) to output a file containing loops that meet those criteria (*conglomerated_Gapless_NonRedundant_Hot_%_#.csv*). Our percent total interface energy of the loop was set for 50 %, and the hot spot cutoff was set at 60 (≥ 0.6 REUs) and the output file *conglomerated_Gapless_NonRedundant_Hot_50_60.csv* was produced that contains all loops that meet this criterion.

After hot loop identification, *LoopCulling.py*, written by Michael Bird, performed a second redundancy check to remove any hot loop that is entirely contained within another and kept only the longest possible version of a given hot loop in the output file *conglomerated_Gapless_NonRedundant_Hot_50_60_NonRedundant.csv*. While both versions of the loop may be useful starting points for peptide design, retaining only the longest loop greatly minimized the numbers of redundant, "nested" loops. *LoopCulling.py* also provided a second output file that removed overlapping loops, named *conglomerated_Gapless_NonRedundant_Hot_50_60_NonRedundant_NonOverlapping.csv*. When two loops overlaped in sequence but one was not completely nested within the other, the locations of hot spot residues were used to decide which to keep. If two overlapping loops had the same hot spots, *LoopCulling.py* kept only the loop with higher average $\Delta\Delta G_{res}$. When one loop contained all the hot spots of the other as well as additional hot spots, the loop with more hot spots was kept. In the case that each loop contained at least one hot spot that is not in the other, they were both kept, since these overlapping loops represented potentially independent epitopes for translation to macrocyclic inhibitors.

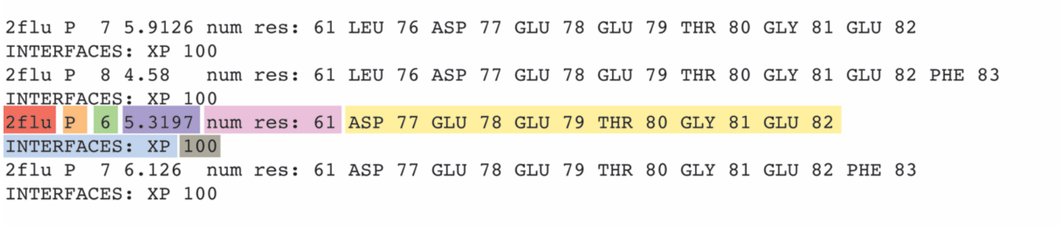
To better organize the final set of 7,225 hot loops, a numbering system was developed to score each loop in terms of what hot loop category it fulfills (Figure 2.15). If the loop had an average $\Delta\Delta G_{res} \geq 0.6$ REUs, the loop received 2 points. If the loop had three or more hot spot residues, it received 3 points. If the loop contributed ≥ 50% of the total interface energy, it received 4 points. As a result, a heat number is generated from 2-9 depending on how many hot loop parameters the loop satisfied and allowed a Venn Diagram of hot loops to be generated (Figure 2.4). The organized set of hot loops is compatible with Microsoft Excel for viewing or SQLite for more efficient database management. It is also available on the LoopFinder website at www.sites.tufts.edu/loopfinder.

Figure 2.10 Organization of the LoopFinder cluster directory. The locations of all relevant programs and scripts necessary for running LoopFinder are shown with directories in blue and scripts in green.



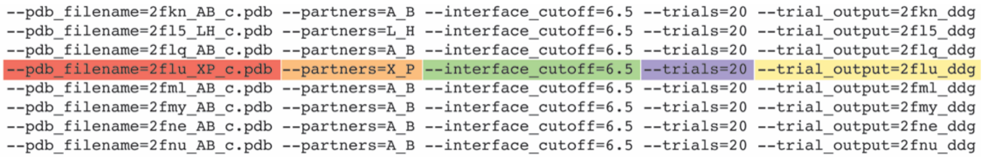
```
./LoopFinder_Unix -input /cluster/tufts/loopfinder/oligomeric_state_2
 -bdist 6.5 -minres 4 -maxres 8 -minperc 80 -maxdist 6.2 -lfactor 1.0 -nchain 2
```

Figure 2.11 *LF_run.sh* script allows for batch submission of multiple LoopFinder runs. The loop parameters used in the search for loop mediated PPIs are shown. Multiple parameters need to manually adjusted for each LoopFinder run. The input directory of PDB files (red) must be set to a single directory of PDBs in which all PDBs in that directory have the same number of chains. This number should also match the *nchain* parameter (gray). The *bdist* defines the interface distance (orange). Minimum and maximum loop length are adjusted with the *minres* and *maxres* parameters (green and purple, respectively). The percent of the loop required to reside at the interface is adjusted with *minperc* (pink). The maximum distance from C_α to C_α of the N- and C-terminal residues of the loop can be set with *maxdist* (yellow). For smaller loops, the *lfactor* was set to 1 (blue).



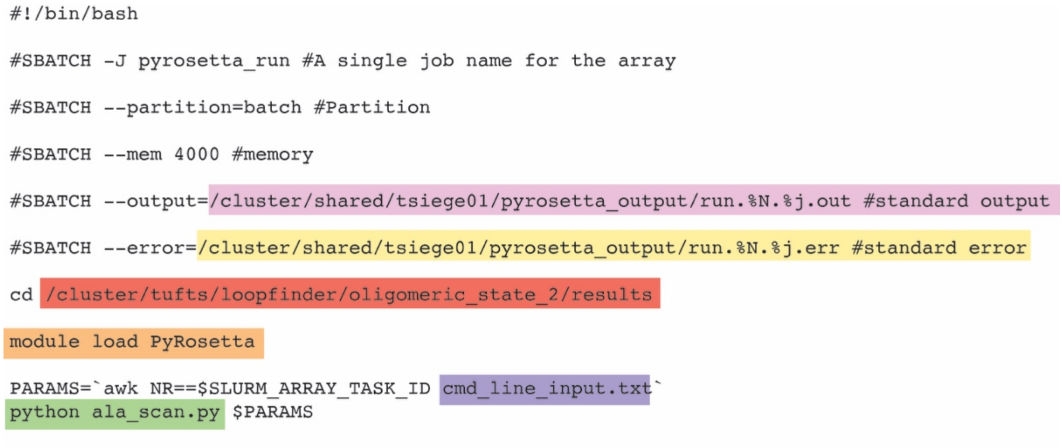
```
2flu P  7 5.9126 num res: 61 LEU 76 ASP 77 GLU 78 GLU 79 THR 80 GLY 81 GLU 82
INTERFACES: XP 100
2flu P  8 4.58   num res: 61 LEU 76 ASP 77 GLU 78 GLU 79 THR 80 GLY 81 GLU 82 PHE 83
INTERFACES: XP 100
2flu P  6 5.3197 num res: 61 ASP 77 GLU 78 GLU 79 THR 80 GLY 81 GLU 82
INTERFACES: XP 100
2flu P  7 6.126  num res: 61 ASP 77 GLU 78 GLU 79 THR 80 GLY 81 GLU 82 PHE 83
INTERFACES: XP 100
```

Figure 2.12 *Loops.txt* provides data on all loops that fit the loop definition parameters. The file has two lines per loop. The first line contains the PDB ID (red), chain on which the loop resides (orange), number of residues in the loop (green), distance from C_α to C_α of N- and C-terminal residues (purple), number of residues at the interface (pink), and each residue of the loop itself (yellow). The second line has the interface (blue) and the percentage of the loop atoms that reside within the given distance parameter of the partner chain (gray).



```
--pdb_filename=2fkn_AB_c.pdb --partners=A_B --interface_cutoff=6.5 --trials=20 --trial_output=2fkn_ddg
--pdb_filename=2f15_LH_c.pdb --partners=L_H --interface_cutoff=6.5 --trials=20 --trial_output=2f15_ddg
--pdb_filename=2flq_AB_c.pdb --partners=A_B --interface_cutoff=6.5 --trials=20 --trial_output=2flq_ddg
--pdb_filename=2flu_XP_c.pdb --partners=X_P --interface_cutoff=6.5 --trials=20 --trial_output=2flu_ddg
--pdb_filename=2fml_AB_c.pdb --partners=A_B --interface_cutoff=6.5 --trials=20 --trial_output=2fml_ddg
--pdb_filename=2fmy_AB_c.pdb --partners=A_B --interface_cutoff=6.5 --trials=20 --trial_output=2fmy_ddg
--pdb_filename=2fne_AB_c.pdb --partners=A_B --interface_cutoff=6.5 --trials=20 --trial_output=2fne_ddg
--pdb_filename=2fnu_AB_c.pdb --partners=A_B --interface_cutoff=6.5 --trials=20 --trial_output=2fnu_ddg
```

Figure 2.13 The LoopFinder program generates a *cmd_line_input.txt* file. This submits jobs for PyRosetta alanine scanning mutagenesis. The information contained within this file is the PDB file name (red), the partner proteins (orange), the interface cutoff distance (green), number of trials to run for each interface (purple), and the name of the file for data output (yellow). The *int_eng_run.sh* script will execute each line of *cmd_line_input.txt* using PyRosetta's *ala_scan.py* program.



```
#!/bin/bash

#SBATCH -J pyrosetta_run #A single job name for the array

#SBATCH --partition=batch #Partition

#SBATCH --mem 4000 #memory

#SBATCH --output=/cluster/shared/tsiege01/pyrosetta_output/run.%N.%j.out #standard output

#SBATCH --error=/cluster/shared/tsiege01/pyrosetta_output/run.%N.%j.err #standard error

cd /cluster/tufts/loopfinder/oligomeric_state_2/results

module load PyRosetta

PARAMS=`awk NR==$SLURM_ARRAY_TASK_ID cmd_line_input.txt`
python ala_scan.py $PARAMS
```

Figure 2.14 Submission of interfaces for energy calculation was carried out using the *int_eng_run.sh* script depicted here. For each run, the PDB directory needs to be changed (red). The script then loads the PyRosetta module (orange) and calls a line from the *cmd_line_input.txt* file located in the directory (red) for submission to *ala_scan.py* (green). Errors and progress data were deposited into other directories (pink and yellow).

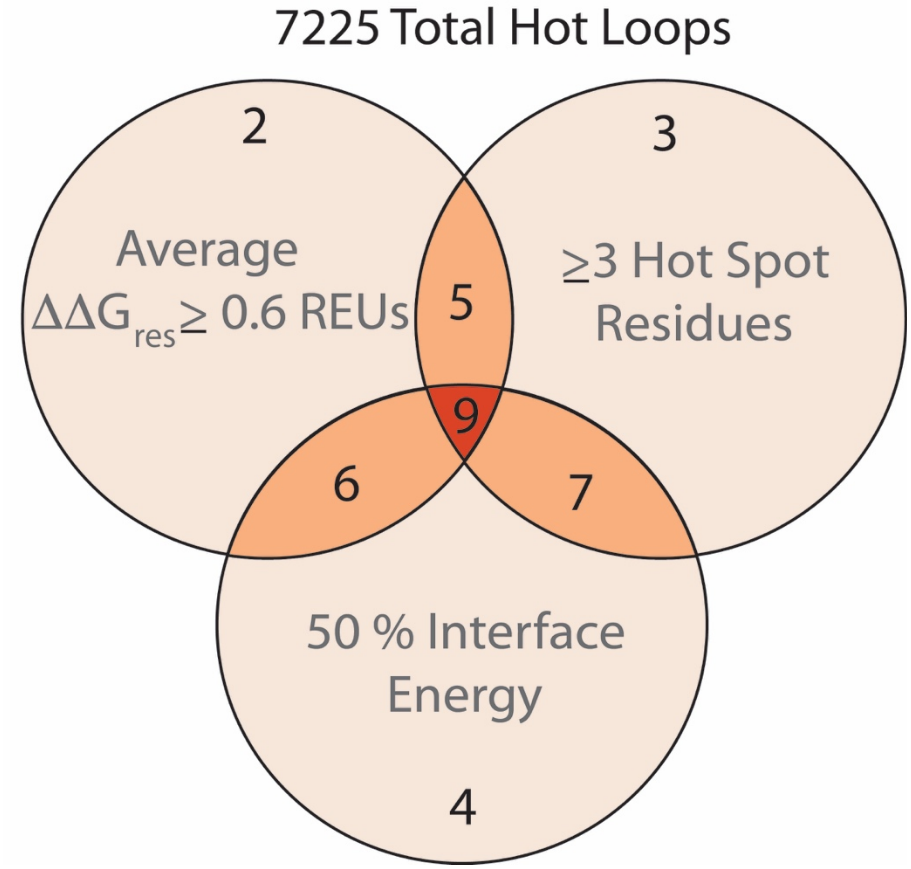


Figure 2.15 Scoring of hot loops ranks the loops for their potential for direct translation into cyclic peptide inhibitors. Each parameter was given a point total, and the loops were scored based on how many hot loop parameters it satisfied.

3 Chapter 3 : Targeting PPIs Identified via LoopFinder

3.1 Suggested Targets for Cyclic Peptide Design

LoopFinder had been carefully developed to best identify essential loops that mediate PPIs. Thus it was logical to search the 7,225 hot loops for biomedically relevant targets. Presented here are a variety of different PPIs that make attractive targets for cyclic peptide design projects. A common theme among the interactions described in this chapter is that the interactions have never been targeted before with in the manner we propose. Some have never been developed into targets for inhibition studies at all. Others may have been targeted with small molecules or peptides with limited success, and the identification of these hot loops for peptide design present a new surface region to target for PPI disruption.

3.1.1 Skp2:Cks1 an E3 ligase complex

Tools for modulating the cell cycle are important for basic cancer research and for emerging therapies. p27^{Kip1} is a G1-checkpoint protein that directly and indirectly regulates many components of the eukaryotic cell cycle.[113] Enhanced degradation of p27^{Kip1} is associated with many common cancers. p27^{Kip1} is targeted for proteolysis by the ubiquitin E3 ligase SCF^{Skp2}, which is a complex of Skp1, Cul1, Rbx1, and Skp2.[113,114] Ubiquitination of p27^{Kip1} also requires the binding of an accessory protein, Cdc kinase subunit 1 (Cks1), to the SCF^{Skp2} complex.[114] Inhibition of this complex restores basal levels of p27^{Kip1}, allowing for proper cell cycle regulation.[115] Ubiquitination of p27^{Kip1} by SCF^{Skp2} has been studied comprehensively *in vitro* and *in vivo*, validating SCF^{Skp2} as an exciting drug target.[116–118] Small molecules have been developed to block the p27^{Kip1} binding site, the interface between Skp1 and Skp2, and the interface between Skp2 and Cks1.[89,90,119] The small molecules that target the Skp2:Cks1 interaction are relatively weak (double digit micromolar) inhibitors of the interaction and their binding pockets are undefined. There is much room in this field for the development of peptide-based inhibitors.

In crystal structures of the SCF^{Skp2} complex (2ASS) and SCF^{Skp2} bound to $p27^{Kip1}$ (2AST),[54] LoopFinder identified a six-residue loop on the surface of Cks1 comprising M38 to W43 (MSESEW) (Figure 3.1). This loop contains four consecutive hot spot residues from S39 to E42 ($\Delta\Delta G_{residue}$ = 1.11 REUs, 1.93 kcal/mol, 2.07 REUs, and 2.94 REUs respectively) located within an α-helical turn. Experimental mutagenesis has shown that S41 is essential SCF^{Skp2} complex activity.[120] Thus, LoopFinder identified a novel starting point for designing SCF^{Skp2} inhibitors that will be accessible using diverse α-turn-mimicking scaffolds.[121–123] The hot loop binds a shallow region on the surface of Skp2 which might be more amenable to peptide-based approaches than small molecule inhibitor design. Efforts towards the identification of cyclic peptide inhibitors of this interaction are currently underway in the Kritzer lab.

3.1.2 Msl1:Msl3 a histone acetyltransferase complex

MOF (males-absent on the first) is a histone acetyltransferase (HAT) that exclusively catalyzes the acetylation of histone 4 lysine 16 (H4K16).[124] Only a handful of HAT inhibitors have been discovered, and all target the catalytic site.[125] Small molecule inhibitors of Tip60, another HAT that catalyzes the acetylation of H4K16, have been shown to sensitize cells to DNA-damaging agents, hinting at the therapeutic potential of HAT inhibitors.[126] Loss of MOF function causes G_2/M cell cycle arrest, nuclear morphological defects, spontaneous chromosomal aberrations, and impaired DNA repair response following ionizing radiation.[92] However, the role of MOF in regulating cell cycle and DNA repair are only just beginning to be understood. A selective inhibitor for this complex would allow for the improved understanding of transcriptional regulation associated with specific histone acetylation sites.

MOF requires complexation of three regulatory proteins (MSL1, MSL2 and MSL3) for activity.[91,92,127–129] LoopFinder identified two hot loops as critical for the formation of the MSL complex (4DNC).[127] One of these is V575 to P580 (VAFGRP) on MSL1 (Figure 3.2), with Phe577 and Arg579 as hot spots ($\Delta\Delta G_{res}$ = 2.74 REU and 4.5 REU respectively). This loop forms a β-hairpin-like structure that binds a shallow pocket on MSL3 (2Y0N).[91] Co-immunoprecipitation experiments showed that variants of MSL1 with mutations in this loop have substantially lower MSL3 affinity, and that a residue in the hot loop binding region on MSL3, F484, is essential for recognition of MSL1.[91] The other hot loop in the MOF-MSL1/2/3 complex is H183 to G188 (HIGNYE) of MOF, which binds MSL1 using hot spot residues Asn186 and Glu188 ($\Delta\Delta G_{res}$ = 1.54 REU and 3.23 REU respectively). To our knowledge, there are no small molecules or peptides that have been developed to inhibit any member of the MSL complex. LoopFinder has thus identified interactions that may lead to the development of completely novel anticancer therapeutics and chemical probes to increase our understanding of this pathway.

3.1.3 A Wnt Signaling Pathway Regulating Interaction

LoopFinder identified two loops at the interface between R-spondin 1 (RSPO1) and its binding partners, leucine-rich repeat-containing G-protein coupled receptor 4 (LGR4) and ring finger protein 43 (RNF43). RSPO1 and LGR4 are components of a ligand-receptor system that plays an integral role in activating the Wnt signaling pathway, affecting stem cell survival and oncogenesis.[125] RNF43 and ZNRF3 are cell-surface transmembrane E3 ubiquitin ligases that inhibit Wnt/β-catenin signaling by inducing the endocytosis and subsequent destruction of the Wnt receptors Frizzled (FZD) and LRF5/6.[126] RSPO1 enhances Wnt signaling by recruiting RNF43 or ZNRF3 to complexation with LGR4, inducing endocytosis of this complex instead of the Wnt receptors.[127,128] The Wnt signaling pathway and its regulation by RSPO1 are still in early stages of understanding, and chemical biology tools that can manipulate these signaling processes will give a greater understanding of the roles these proteins play in stem cell growth, gene expression and cancer development. In addition, Wnt is an exciting and growing field for the design of cancer therapeutics, which mostly target downstream Wnt signaling pathways. Inhibition of this complex could afford a novel and exciting approach to the development of novel cancer therapeutics.

As previously mentioned, LoopFinder was able to identify two loops on RSPO1 that are responsible for drastically different PPIs. One of the loops identified on RSPO1 is a β-hairpin located at its interface with LGR4 composed of residues C105 to F110 (CFSHNF). Hot spots are located at the two phenylalanine residues within the loop (Figure 3.3). This β-hairpin has been identified as being essential for binding and is called the F clamp, responsible for "clamping down" on A190 of LGR proteins. When either one is mutated to glutamic acid or alanine a drastic loss of binding is observed.[125] The affinity of this interaction is 2-5 nM.[125,129] It has been shown previously that rationally designed cyclic peptides based on interacting regions of protein ligands are able to bind leucine rich repeats (LRRs) and modulate other cell signaling processes.[34] Therefore, mimicry of the β-hairpin F-clamp loop on RSPO1 should be able to bind the LRR on LGR4. Inhibition of this PPI would prevent RSPO1 from recruiting RNF43 or ZNRF3 to LGR and enable NRF43 and ZNRF3 to clear the Wnt receptors from the cell surface.

A second hot loop was found by LoopFinder at the interface between RSPO1 and RFN43, providing an alternate approach to inhibiting the Wnt signaling pathway (Figure 3.3). This second hot loop is another β-hairpin, from E65 to V72 (ERNDIRQV) with hot spots at R66, D68 and Q71 having $\Delta\Delta G_{res}$ of 1.66 REU, 2.04 REU, and 1.55 REU respectively. Systematic alanine mutagenesis found that both R66A and Q71A mutations completely abolish all RSPO1-ZNRF3 binding. This complex has a weaker 1-10 μM affinity.[125,129] Blocking this PPI would prevent RNF43/ZNRF43 binding to RSPO1 and allow for the endocytosis and destruction of the Wnt receptors. Mimicry of the RSPO1 loops identified by LoopFinder presents an exciting avenue for inhibition of Wnt signaling with a solid basis in both computation and experimental alanine scanning mutagenesis, showing the importance of these two loops for proper complex formation.

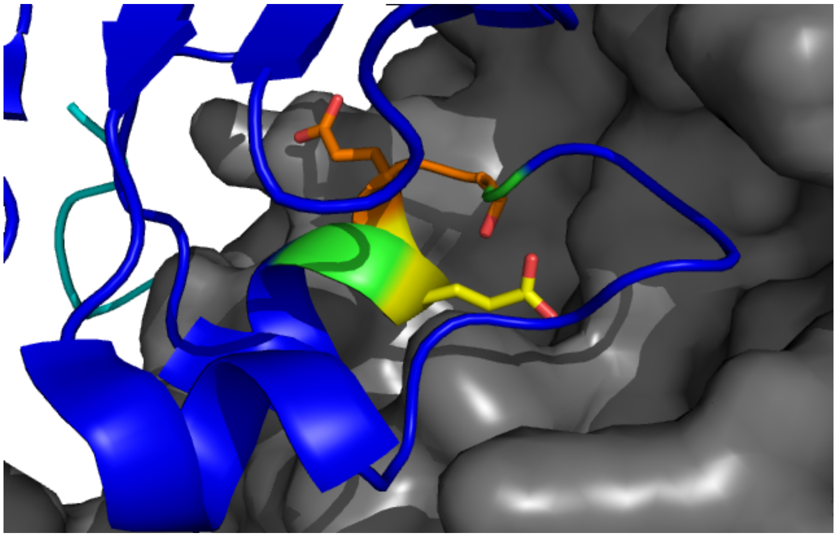


Figure 3.1 A hot loop (green) with hot spot amino acids (orange and yellow) was identified to reside on a helix cap of Cks1 (blue) at the interface with its parner Skp2.

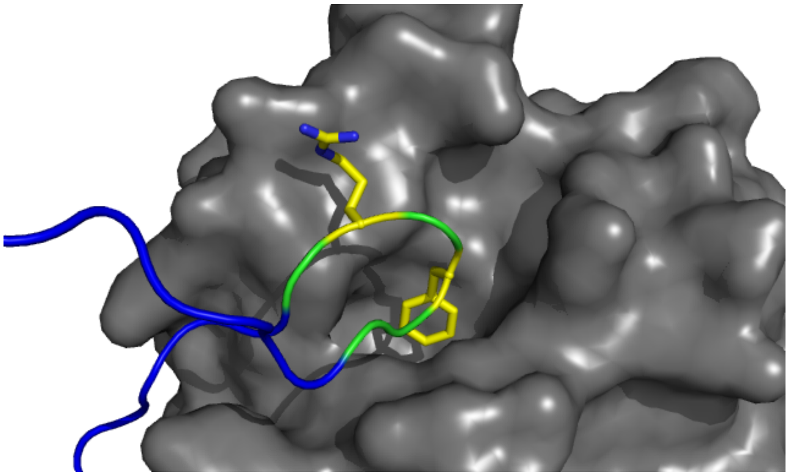


Figure 3.2 A hot loop (green) with hot spots (yellow) on MSL1 (blue) bound to its MSL3 (gray).

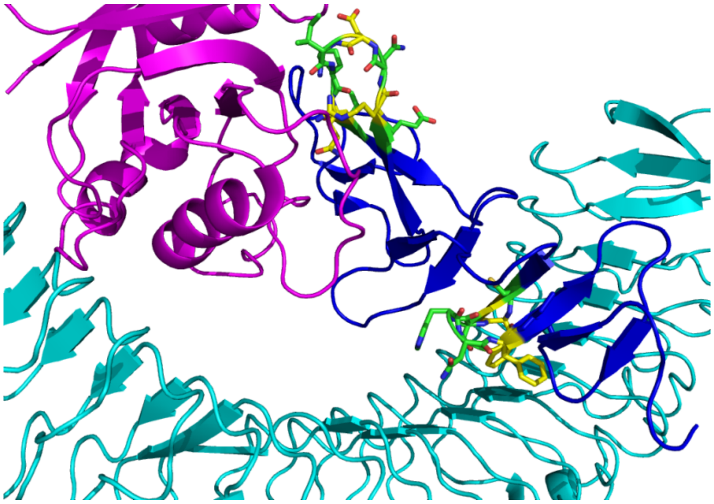


Figure 3.3 The crystal structure of RSPO1 (blue) with LGR4 (cyan) and RNF43 (magenta) (4KNG). LoopFinder was able to identify two loops on RSPO1 that are experimentally shown to be essential for its interaction with LGR and RNF43. The hot loops are shown in green with hot spots in yellow. The first loop is known as the F-clamp, which is responsible for the specific recognition of RSPO1 by LGR. The second is another β-hairpin at the interface with RNF43, which contains a conserved R66 and Q71 that are essential for binding.

3.2 Cyclic peptides targeting the TIMP3-TACE Interaction

From the first generation LoopFinder analysis of the PDB, a protein-protein interaction that was found to be a biomedically relevant target for the design of cyclic peptides was the interaction between tissue inhibitor of metalloproteinases 3 (TIMP-3) and TNF-α converting enzyme (TACE).[130] TACE is a member of the ADAM (a disintegrin and metalloproteinase) family of proteases that are involved in the shedding of transmembrane-bound proteins on cell surfaces, including the release of growth factors and cytokines into the extracellular matrix.[131] TACE, or ADAM-17, is responsible for the processing of proTNF-α into its soluble form, TNF-α (Tumour necrosis factor-α).[132] TNF-α is a cytokine that plays a major role in infection response pathways and inflammation, triggering signaling pathways that result in necrosis or apoptosis.[133,134] TNF-α signaling is a major focus of therapeutic strategies for rheumatoid arthritis and Crohn's disease.[131] TIMP-3 is an extracellular protein that binds TACE with sub-nanomolar affinity, inhibiting proteinase activity.[135] The TIMP-3•TACE interaction has been extensively studied both *in vitro* and *in vivo*, and numerous peptide and small molecule inhibitors (which include broad spectrum MMP inhibitors) have been identified by academic and industrial groups.[136-138] All of these molecules have targeted the catalytic site of TACE, most using a hydroxymate moiety to bind the catalytic zinc.[137] The most successful small molecule dropped out of Phase II clinical trials, most likely due to issues with specificity as there is close homology among proteases of this class.[137]

The TIMP-3•TACE interaction is facilitated by three interface epitopes within TIMP-3 that bind TACE (3CKI).[130] LoopFinder identified one of these epitopes, called the sC-connector loop, as a hot loop for the TIMP-3•TACE interaction (Figure 3.4).[130,134] The hot loop consists of a six-residue stretch from S64 to G69 (SESLCG), in which S64 and L67 are identified as hot spots ($\Delta\Delta G_{res}$ = 3.62 REUs and 1.73 REUs respectively) shown in Figure 3.4. Notably, S64 and L67 are unique residues within TIMP-3 compared to TIMP-1, TIMP-2, and TIMP-4, which show drastically reduced binding to TACE.[130,134] In addition, this region is structurally constrained within TIMP-3 via a disulfide bond from Cys68 to Cys1. The binding pocket for the hot loop has been identified an "alternative pocket" on TACE that could be as important as the actual catalytic zinc active site, but this loop within TIMP-3 has never been used as a potential starting point for TACE inhibitors (Figure 3.5).[130,134] Thus, the S64-G69 hot loop is a starting point for the design of non-zinc-chelating inhibitors of TACE. Such inhibitors would have immense therapeutic potential as highly selective TACE inhibitors that do not bind other ADAMs or MMPs.

To test the ability of designed peptides to inhibit TACE, a TACE inhibitor screening kit was purchased from Abcam. This assay utilizes a FRET peptide probe as a substrate for TACE. by Proteolysis of the peptide releases the fluorescent probe from its proximity to the quencher. Enzyme activity is then measured as a function of emission at 449 nm when excited at 318 nm. A small set of peptides were designed to mimic the hot loop identified by LoopFinder. The first peptide synthesized was a head-to-tail cyclic peptide cyc-YASESLSG as well as a linear control peptide with a similar sequence of Ac-YASESLSG-NH_2. For these peptides, a tyrosine residue was added flanking the hot loop to allow for quantification by UV/Vis. Since N-terminus of this loop is solvent exposed in the crystal structure, we placed the tyrosine at the N-terminus of the loop. Another modification was replacement of the C68 residue for serine, as the cysteine residue is involved in a disulfide bond to other regions of TIMP-3 and would be sensitive to oxidation. Unfortunately, neither of these peptides exhibited any inhibition of TACE activity.

Due to difficulties associated with the synthesis of head-to-tail peptides a bis-alkylation approach that will be discussed in more detail in Chapter 4 was utilized to cyclize a single hot loop epitope flanked by cysteine residues with a variety of different di-bromomethyl-xylene containing linkers. This chemistry was first developed by Peter Timmerman and was later utilized in screening approaches to stabilize peptide conformation.[139] Four other peptides were synthesized based on this approach including Ac-CSESLSC-NH_2, Ac-CECLS-NH_2, Ac-CECLY-NH_2 and Ac-CECLW-NH_2. They were subsequently cyclized using di-bromo-*o*-xylene (***ox***), di-bromo-*m*-xylene (***mx***), di-bromo-*p*-xylene (***px***) to bis-alkylate the cysteine residues.

Unfortunately, none of these peptides were capable of inhibiting TACE. It is important to note that these studies did not look for protein binding, and it is still possible that binding could take place even without observing inhibition. This is plausible because the recognition site of the substrate is distal to the targeted peptide binding region. The peptide may not be competitive with the substrate, allowing the substrate to be bound and cleaved by the enzyme even with the inhibitor peptide present. These inhibitors could be improved by including a hydroxamate-containing amino acid capable of reaching into the enzyme active site and coordinating the active zinc ion. In this case, the hot loop peptide will help introduce specificity that the existing small molecules do not possess. As shown in Figure 3.5, all existing small molecule hydroxamate-containing inhibitors fit into the substrate groove on the surface of TACE. None so far have been developed that bind the hot loop region from TIMP-3 that controls specificity of this protein to TACE over other TIMP proteins such as TIMP-1, TIMP-2 and TIMP-4. With this approach, there is still much chemical space that can be explored and exploited to find novel inhibitors of TACE.

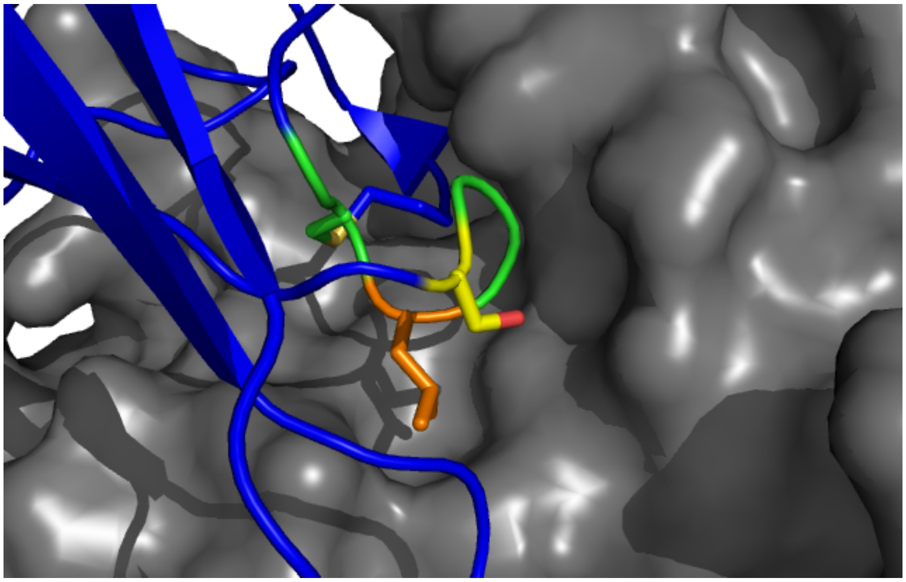


Figure 3.4 The TIMP-3: TACE interaction. The hot loop (green) on TIMP-3 (blue) is shown in this crystal structure bound to its target protein TACE (gray). Hot spot residues of S64 (yellow) and L67 (orange) are shown (PDBID: 3CKI).[140]

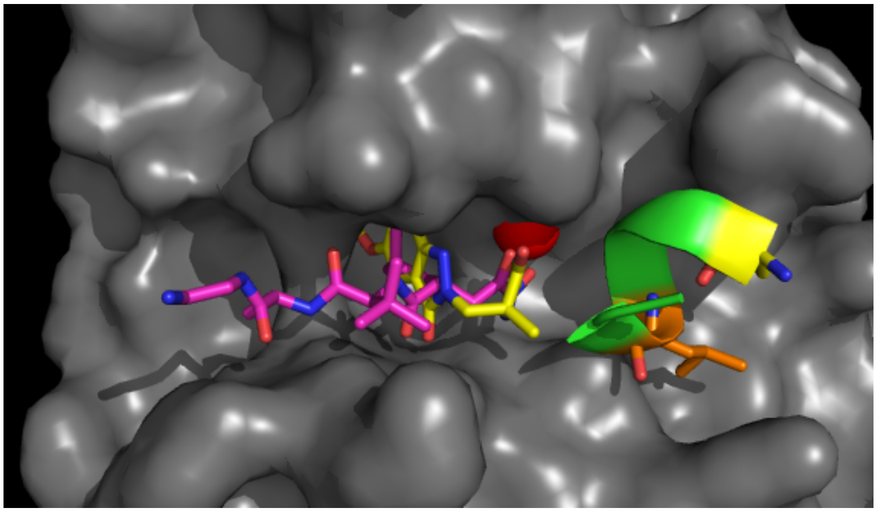


Figure 3.5 Two binding pockets on TACE. TACE (gray) is shown overlaid with existing small molecule hydroxamate-containing inhibitors (pink and yellow) and the TIMP-3 hot loop identified by LoopFinder (green with hot spots in orange and yellow).[141]

3.2.1 Materials and Methods: TIMP3-TACE

Reagents used were as follows. Fmoc-protected amino acids were purchased from either Anaspec, Chem Pep, or Novabiochem. The 1-[Bis(dimethylamino)methylene-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), 1-hydroxy-7-azabenzotriazole (HOAt), 6-Chloro-Benzotriazole-1-yl-oxy-tris-Pyrrolidino-Phosphonium hexafluorophosphate (PyClock) and 6-chloro-1-hydrobenzotriazole dihydrate (Cl-HOBt) were purchased from Chem Pep. Chlorotrityl resin was purchased from Novabiochem. The dimethylformamide (DMF), *N*-methyl-2-pyrrolidone (NMP), *N,N*-diisopropylethylamine (DIEA), acetic anhydride, trifluoroacetic acid (TFA), dichloromethane (DCM), triisopropylsilane (TIS), 2,6-Lutidine, ethanedithiol (EDT) and piperidine were purchased from Sigma Aldrich.

The linear precursor peptide Ac-YASESLSG-NH_2 synthesized using standard Fmoc peptide synthesis protocol on Rink amide resin (substitution = 0.44 mmol/g) on the CEM microwave. Couplings were carried out using 5 equivalents (equiv.) of Fmoc protected amino acid, 4.8 equiv of HOAt, 13 equiv of collidine dissolved in 1.2 mL of 0.4 M HATU in DMF solution and carried out using the coupling protocol on the microwave where it is heated for 3 minutes to 78 ºC. The deprotection was carried out in 5% (vol/vol) piperazine/NMP for 1.5 minutes heated to 78ºC. The resin was washed three times with DMF, three times with DCM and three times with DMF again between each coupling. Upon completion of the peptide sequence assembly on resin and deprotection of the final Fmoc group, the N-terminal amine was acetylated by treatment with 5% (vol/vol) acetic anhydride and 6% (vol/vol) 2,6-lutidine in DMF.

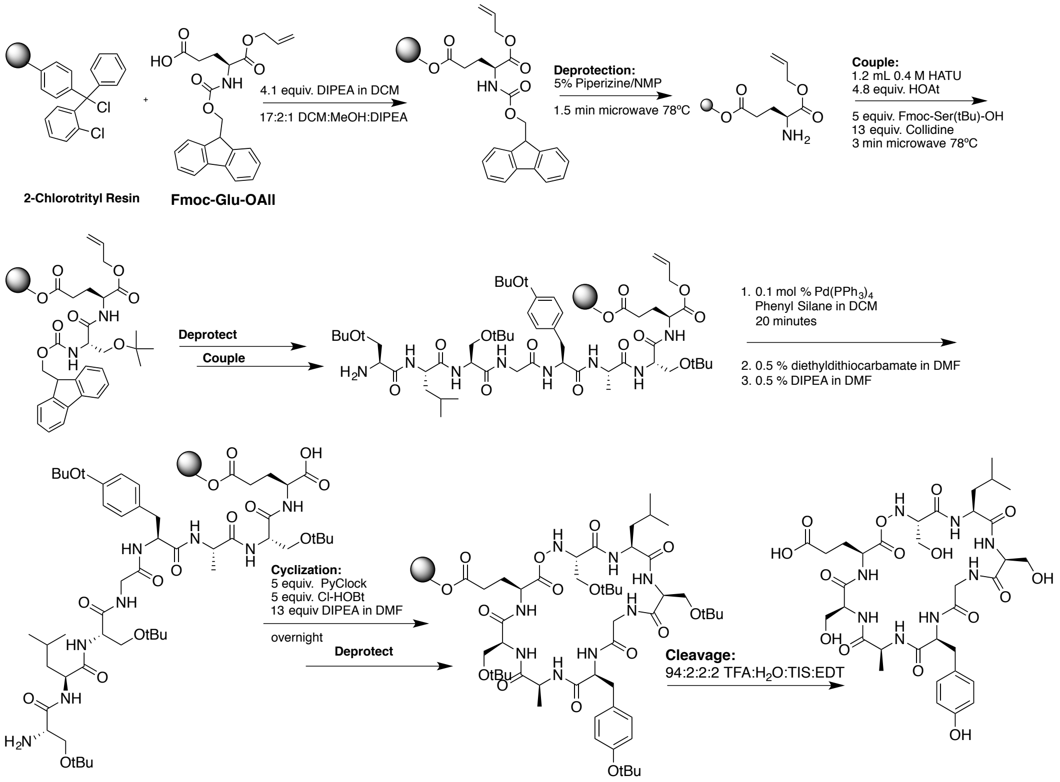
For the head-to-tail cyclic peptide, synthesis was carried out on chlorotrityl resin as outlined in Scheme 3.1. Resin was loaded with backbone-carboxyl-protected glutamic acid (Fmoc-Glu-OAll) with 4.1 equivalents of DIPEA in 8 mL of DCM for 90 minutes with constant shaking. The reaction was then washed with a 17:2:1 solution of DCM:MeOH:DIPEA followed by washing three times with DMF and three times with DCM. The peptide was synthesized using Fmoc solid phase synthesis on the microwave as described above. Upon completion of the final coupling, the C-terminal carboxyl group was deallylated using 0.1 mol % palladium tetrakis (11.5 mg), 123 µL phenylsilane in DCM for 20 minutes with argon bubbled through the solution to keep the reaction dry. The deallylation reaction was then worked up by repeating the following wash steps three times in succession: washing the resin with DMF, then washing the resin with a solution of 0.5 % diethyldithiocarbamate in DMF (100 mg in 20 mL), followed by washing with a solution of 0.5% DIPEA in DMF (100 µL in 20 mL). The final Fmoc deprotection was carried out and to cyclize, the resin was allowed to shake overnight in a solution containing 444 mg PyClock, 139 mg Cl-HOBt and 220 µL DIPEA in DMF.

Cleavage and deprotection from the solid support was carried out using TFA/H_2O/TIS/EDT (94/2/2/2 vol/vol) for 3 hours at room temperature. The resin was then filtered and washed with excess TFA. The isolated TFA solution was then evaporated, and the peptide was precipitated upon addition of cold diethyl ether. The peptide suspension in ether was then centrifuged to pellet the peptide followed by decanting away the ether. The pellet was then washed two additional times with cold diethyl ether to remove as much TFA as possible.

The dithiol-containing peptides were synthesized and cyclized via a method that will be discussed in Section 4.8.1. Purification of the peptides was achieved by high performance liquid chromatography (HPLC, Varian ProStar) on a reversed-phase C_8 column (Agilent; Zorbax C8) to yield pure peptide. Purity was assessed using MALDI in conjunction with analytical scale HPLC on a C_{18} column (Vydac, Protein and Peptide C_{18}).

Peptides were screened for their ability to inhibit TACE using a commercially available kit (Abcam, ab155889). TACE enzyme provided was dissolved in 220 μL of assay buffer (provided) and distributed to a PCR strip into 9x24 μL aliquots and stored in the -80 ºC freezer until needed. Peptides were prepared in 0.5 mL stocks containing 10% DMSO and 90% assay buffer at 4X concentrations. To run the assay, 24 μL of the TACE enzyme aliquot was diluted into 576 μL of assay buffer and 25 μL of this solution was added to a 96 well black well plate with a clear bottom. 12.5 μL of a peptide stock was then added to each well, mixed thoroughly and allowed to incubate for 5 minutes at 37 ºC. A solution of substrate was prepared by adding 24 μL of substrate to 276 μL of assay buffer, and 12.4 μL of this substrate solution was added to each well. The fluorescence was then measured over time with an Ex/Em of 318/449 nm. A reduced fluorescence signal at 449 nm after the 30-minute incubation time compared to the DMSO control was used as a measure of enzyme inhibition.

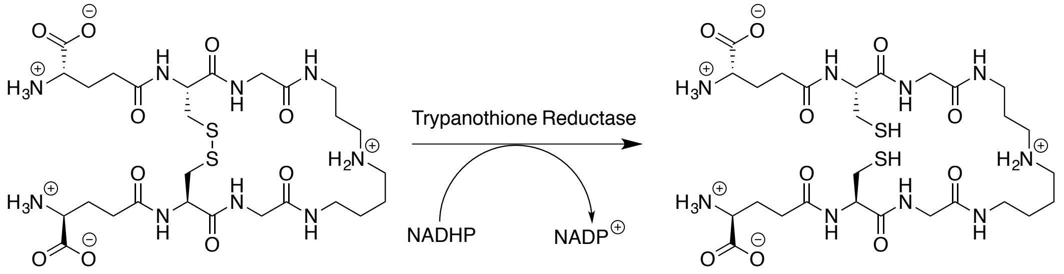
Scheme 3.1 Synthesis of head-to-tail cyclized TIMP-3 mimicking peptide.



3.3 TryR Dimerization Inhibition via Cyclic Peptide Hot Loops

A second enticing PPI target from the initial LoopFinder data set was Trypanothione Reductase. Typanothione Reductase (TryR) is a 54.68 kDa molecular weight oxidoreductase protein that catalyzes the reduction of trypanothione disulfide (T[S]$_2$) to dihydro-trypanothione (T[SH]$_2$) (Scheme 3.2).[142] Trypanothione reductase is an essential regulator protein of redox homeostasis in protozoan parasites of the trypanosomatid family, including *Leishmania* and *Trypanosoma*. *Leishmania* and *Trypanosoma* are protozoa transferred to mammalian hosts through the bite of a sand fly and are responsible for a wide variety of diseases including Leishmaniasis, Chagas Disease and African sleeping sickness.[143] Leishmaniasis is a major health concern in Third World countries and is categorized as a rare or orphan disease. This disease is the most widespread of the trypanosomatid infections and comes in three distinct forms: cutaneous leishmaniasis causing skin sores and skin ulcers, visceral leishmaniasis causing weight loss, swelling of the spleen and liver, anemia and leukopenia, and finally mucosal leishmaniasis causing sores in mucous membranes.[143] Leishmaniasis has an estimated 1.3 million new cases per year with greater than 30,000 deaths occurring annually.[144] These trypanosomatid parasites lack the more common glutathione reductase process of intracellular redox maintenance and rely exclusively on the trypanothione reductase redox system. In addition, this process has little no conservation to the glutathione process in mammalian hosts, making it an excellent target for anti-*Leishmania* and anti-*Trypanosoma* drugs.[145] There are some drugs available to treat these parasites, but most have serious issues such as toxicity due to a lack of target specificity, high cost of treatment or rigorous treatment schedules, all of which make it difficult for treating patients in the Third World nations where these diseases are prominent.[146-149]

Scheme 3.2 Conversion of trypanothione disulfide to dihydro-trypanothione via Trypanothion Reductase.



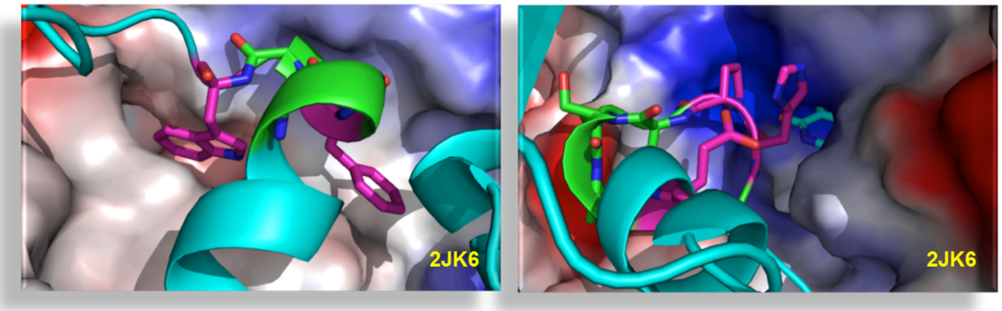
$T[S]_2$
Trypanothione Disulfide

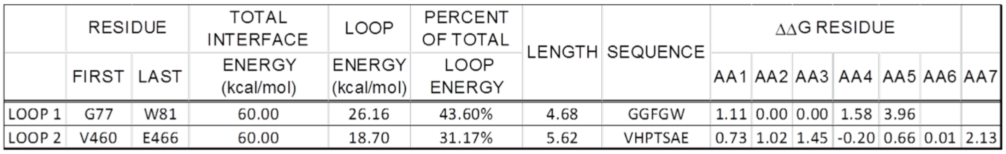
$T[SH]_2$
Dihydro-Trypanothione

3.3.1 LoopFinder Identifies a Hot Loop at the Dimerization Interface of TryR

LoopFinder was able to identify two unique hot loops located at the TryR dimer interface. Dimerization of TryR has been shown to be essential for activity and has also been targeted for inhibition with α-helical peptides based on residues P435 to M447. This strategy yielded a peptide with an IC_{50} of 1.5 μM.[150] In this case, the authors used a computational analysis of solvent-accessible surface area called POPS (Parameter OPtimised Surfaces) in conjunction with a molecular mechanics protein-protein docking program to identify hot spot regions at the dimer interface.[150-152] Using this method, three major hot spots were identified: W81, E436 and Q439. The latter two hot spot residues reside on a single α-helix that was used for peptide design. It is important to note when point mutants were made for each residue in the full protein, only the E436A mutant showed a significant reduction in dimerization of TryR.[150] This approach gives proof-of-principal to the approach of inhibiting TryR with cyclic peptides mimicking interface loops.

As shown in Figure 3.6, LoopFinder was able to identify two loops that reside at the dimer interface of *Leismania infantum* TryR (PDBID: 2JK6).[153] The first loop, defined as hot loop 1 (HL1) is a five-amino-acid loop from G77 to W81 (GGFGW) with hot spots at F79 (1.58 REUs) and W81 (3.96 REUs). The distance from the N-terminus to the C-terminus of the loop is 4.68 Å. Of the total interface energy, calculated to be roughly 60 REUs using PyRosetta computational alanine scanning mutagenesis, HL1 contributes 43.60% of the interface energy (26.16 REUs) (Figure 3.6). This hot loop was attractive for design purposes because it contained two aromatic side chains that have the potential of making strong interactions to the protein target. One of these residues was also previously identified through experimental methods to be a hot spot.[150] The second loop, defined as hot loop 2, is located close to the catalytic site of TryR and is composed of seven amino acids from V460 to E466 (VHPTSAE) with hot spots at H461 (1.02 REUs), P462 (1.45 REUs) and E466 (2.13 REUs) (Figure 3.6). This hot loop contributes slightly less to the total calculated interface energy, making up 31.17% of the total, or 18.70 REUs). These two loops were used to design cyclic peptide inhibitors of TryR activity.





	RESIDUE		TOTAL INTERFACE ENERGY (kcal/mol)	LOOP ENERGY (kcal/mol)	PERCENT OF TOTAL LOOP ENERGY	LENGTH	SEQUENCE	ΔΔG RESIDUE						
	FIRST	LAST						AA1	AA2	AA3	AA4	AA5	AA6	AA7
LOOP 1	G77	W81	60.00	26.16	43.60%	4.68	GGFGW	1.11	0.00	0.00	1.58	3.96		
LOOP 2	V460	E466	60.00	18.70	31.17%	5.62	VHPTSAE	0.73	1.02	1.45	-0.20	0.66	0.01	2.13

Figure 3.6 Hot loops at TryR dimerization interface. The top panel shows Loop 1 and Loop 2, from left to right, identified by LoopFinder to be a hot loop residing at the TryR dimer interface. The table below shows the calculated energy for each residue in the hot loop as well as its contribution to the total interface energy.

3.3.2 Development of Cyclic Peptide Inhibitors of TryR

Leishmania infantum Trypanothione Reductase (TryR) was expressed and purified from bacteria as described in Section 3.3.3. The activity of the enzyme can be measured via a kinetic assay, whereby oxidoreductase activity is measured as a function of increase in measured absorbance at 412 nm corresponding to reduction of Ellman's reagent present as a probe for free thiol.[150] This absorbance corresponds to the reduction of DTNB to TNB$^-$ which absorbs at 412 nM and is coupled to dihydro-trypanothione (T[SH]$_2$) production by TryR (Scheme 3.2). To find the proper amount of TryR for the assay conditions, the amounts of TryR added to constant amounts of trypanothione disulfide (T[S]$_2$, 1 μM), NADPH (150 μM) and Ellman's Reagent (DTNB, 25 μM) were varied. In order for inhibition to be measured accurately, the amount of TryR must be sufficient to maintain a steady initial rate to ensure saturation did not occur too quickly, allowing for several time points to be collected during the initial linear rate period. Increasing amounts of TryR were added to the reaction from 0 nM to 34.4 nM enzyme (Figure 3.7). Concentrations of protein above 20 nM saturated within 400 seconds, which would make large-scale microplate assays difficult because the rate does not stay linear over enough time points. The 10 nM concentration of TryR remained linear over twelve time points (600 s) and was chosen as the working concentration of TryR moving forward. The data from the linear region of the plot (up until 250 s), can be used to fit a linear slope. The slope is the initial reaction rate measured in absorbance units per second. These are shown in Table 3.1.

To further characterize the activity of TryR we measured the initial rate of the conversion of $T[S]_2$ to $T[SH]_2$ using a constant concentration of TryR (10 nM) with varying concentration of substrate. This confirmed that the enzyme was behaving according to single active-site Michaelis-Mentin kinetics. The amount of $T[S]_2$ included in the reaction was varied from 0.25 μM to 10 μM and absorbance at 410 nM was measured as a function of time as discussed before (Figure 3.8A). Data from the linear region of each data set (Figure 3.8B) was fit to a linear treadline, the slope of which corresponds to the initial reaction rate at the various concentrations of substrate. The initial rates were analyzed as a function of substrate ($T[S]_2$) concentration via a saturation curve (Figure 3.9) as well as a Lineweaver-Burk Plot to determine whether the kinetics matched the expected Michaelis-Menten behavior. The Lineweaver-Burk plot was then be used to determine a V_{MAX} of 1.02 mAU/s and a Michaelis constant K_M of 2.02 μM. Characterization of TryR activity was necessary for the optimization of conditions for inhibition assays. From the results above, it was determined that inhibition assay of TryR activity by disruption of dimerization would be done with 10 nM TryR, 1 μM $T[S]_2$, 150 μM NADPH and 25 μM DTNB.

LoopFinder was able to identify two hot loops present at the interface of the TryR dimer. Previous computational techniques corroborated W81 as a hot spot and made hot loop 1 an attractive starting place for cyclic peptide design. As a control, a linear peptide corresponding to the hot loop from G77 to W81 was designed and flanking residues beyond the identified hot spot were included. This linear peptide (HL1) comprises E75 to R83 with the sequence Ac-ESGGFGWQR-NH_2. This linear peptide possesses both the F79 and W81 hot spots. In order to effectively screen a large area of conformational a cysteine bis-alkylation strategy was adapted to produce "locked loop libraries" (Table 3.2).[154] Cysteine residues were introduced into the amino acid sequence of the parent HL1 peptide at varying positions. The C1 peptide series (Table 3.2) introduced the cysteine residues at S70C and Q82C. The C2 series included the cysteine residues at G71C and Q82C, while the C3 series had the cysteines placed at G72C and Q82C. Peptides based on the second loop (HL2) were also made that are intended to mimic residues V460 to E467 (Ac-VHPTSAEE-NH_2). The same locked loop library approach was used to design cyclic variants of this hot loop by introducing three different permutations for cysteiene bis-alkylation yielding the parent peptides P1, P2 and P3 (Table 3.2). Once the linear peptides were synthesized, they were cyclized using either disulfide formation (***ds***) or a variety of di-bromo-containing linkers that bis-alkylate the cysteine residues including: di-bromo-*o*-xylene (***ox***), di-bromo-*m*-xylene (***mx***), di-bromo-*p*-xylene (***px***), 1,8-bis(bromomethyl)naphthalene (***n***). Linear controls were also made using allylbromide (***a***).

The peptides were then tested for their ability to inhibit the reduction of T[S_2] to T[SH_2], and measured as as function of the production of TNB^-. The peptides were initially screened for inhibition at 25 µM concentration according to the methods outlined in 3.3.3. An example of the type of data collected for the peptides ability to inhibit trypanothione disulfide is shown in Figure 3.10. The raw absorbance data was collected over time for 30 minutes, and the inhibition can be measured through either changes in the initial rate of the reaction, by measuring the slope of the linear region, or by measuring the change in final absorbance measurement after 30 minutes. Each family of peptides were analyzed together.

The peptides based on the C1 sequence, C1**ds**, C1**ox**, C1**mx**, C1**px**, C1**n** and C1**a** are compared to controls with no inhibitor, and the HL1 linear peptide (Figure 3.10A). It can be seen from the raw data, that some of the peptides look to be inhibiting the reduction of $T[S_2]$, with the C1m peptide having the largest difference in slope and final absorbance value. When comparing the percent activity using a 30 minute endpoint assay (Figure 3.10D). The C1**mx** peptide reduces activity by about 20 %. In addition, the initial rate for the reaction in the presence of C1**mx** reduces the rate of the reaction from 14 mAU/min to 7 mAU/min (Figure 3.10G). A similar analysis was carried out for peptides in the C2 series (Figure 3.10B). C2**px** appears to have an even larger reduction in activity as measured at the 20 minute endpoint (Figure 3.10E). In addition, the C3 series (Figure 3.10C) all seemed to reduce the activity of TryR (Figure 3.10F and Figure 3.10G). The same analysis was done with the HL2 series of peptides, including the resultant cyclic peptides based on P1, P2 and P3 sequences (Table 3.2) but no measurable changes in initial rate or final absorbance value were observed.

Dose dependences of these peptides were then measured by testing the inhibition of TryR reduction of T[S_2] at increasing concentrations of C1**mx**, C2**px**, C3**ox** and C3**n**. We attempted to determine an IC_{50} for each of these peptides. Unfortunately, while C1**mx**, C2**px**, C3**ox** and C3**n** peptides appeared to be promising, upon further analysis it was determined inhibition was being caused by some type of artifact. When running dose dependence assays we observed a change in the visible appearance of the wells. This change in appearance of the wells was directly associated with the observed increase in inhibition of the peptide at higher concentrations. Upon further we observed these changes even without the addition of substrate. We determined that this was probably caused by aggregation of the peptide. The inability to separate aggregation effects from inhibition led us to not move forward with the identification of cyclic peptide inhibitors of TryR, and refocus our efforts on another target.

Scheme 3.3 Design of TryR activity assay. Ellman’s reagent catalyzes the conversion of dihydro-trypanothione back to trypanothione disulfide and produces 2TNB in the process that can be measured by absorbance at 412 nM. Using this as a measure of catalytic activity of the Trypanothione Reductase, the rate of the reduction of Trypanothione disulfide to dihydro-trypanothione can be quantitated.

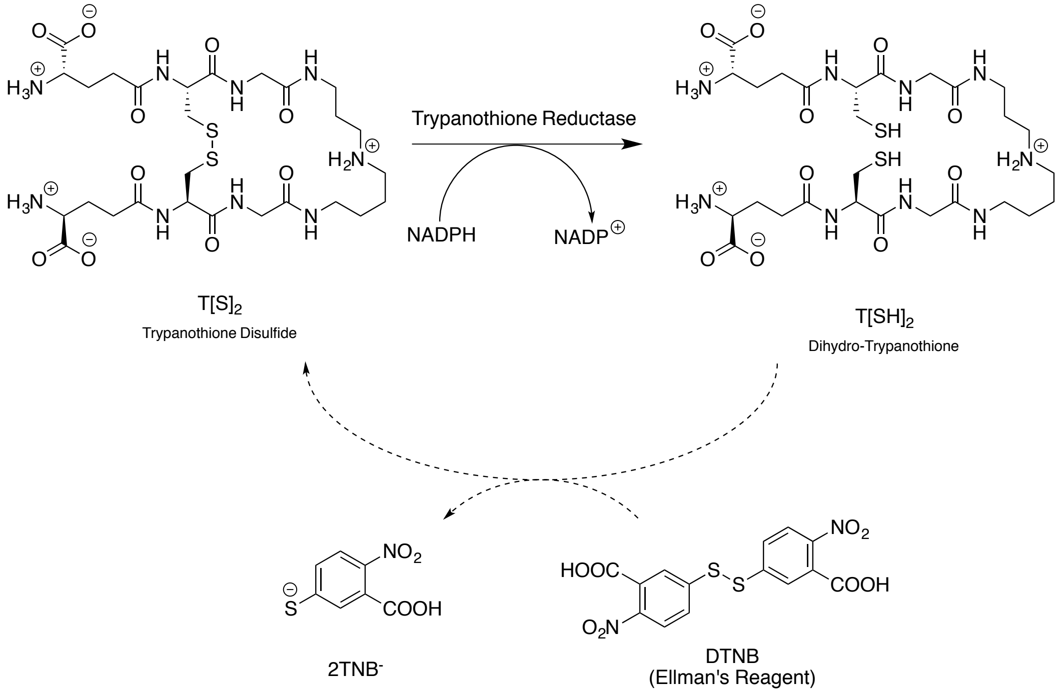


Table 3.1 The initial rate of the reaction or varying concentrations of TryR.

[TryR] µM	V_0 (mAU/s)	V_0 (mAU/min)
0.25	0.12	7.07
0.5	0.18	10.98
0.75	0.24	14.23
1	0.32	19.24
2.5	0.64	38.54
5	0.76	45.76
10	1.48	88.99

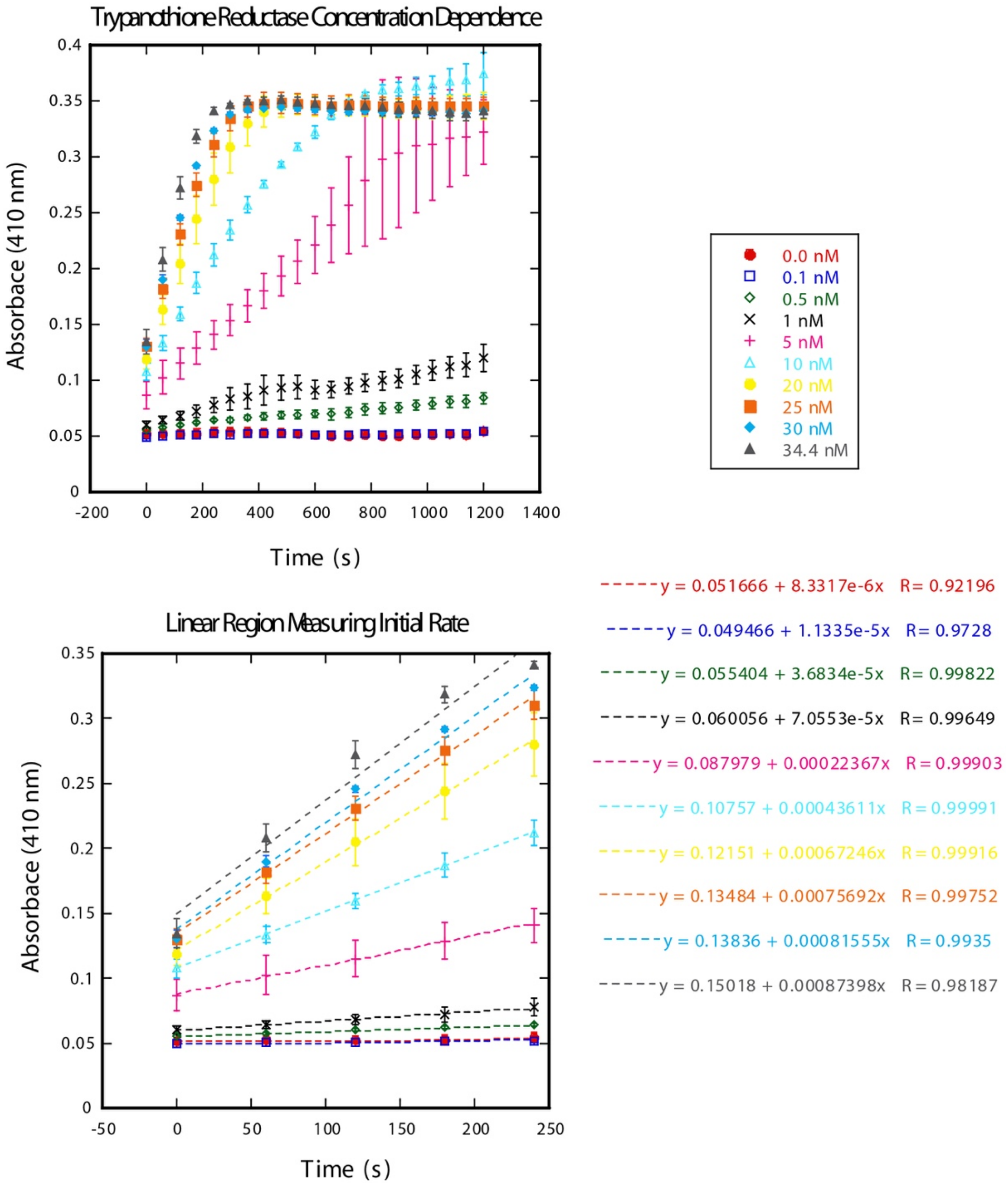


Figure 3.7 Measuring production of TNB$^-$ via absorbance at 410 nm. With increasing amounts of TryR enzyme, the initial reaction rate can be measured at each concentration to identify a TryR concentration that remains linear for enough time points. The graph on the top shows an absorbance of 2TNB$^-$ over time with increasing TryR concentration. Saturation of absorbance occurs quickly for high levels of protein, and remains linear over 600 s for 10 nM TryR. Using the initial rate period, a linear fit can be observed with the slope being the reaction rate.

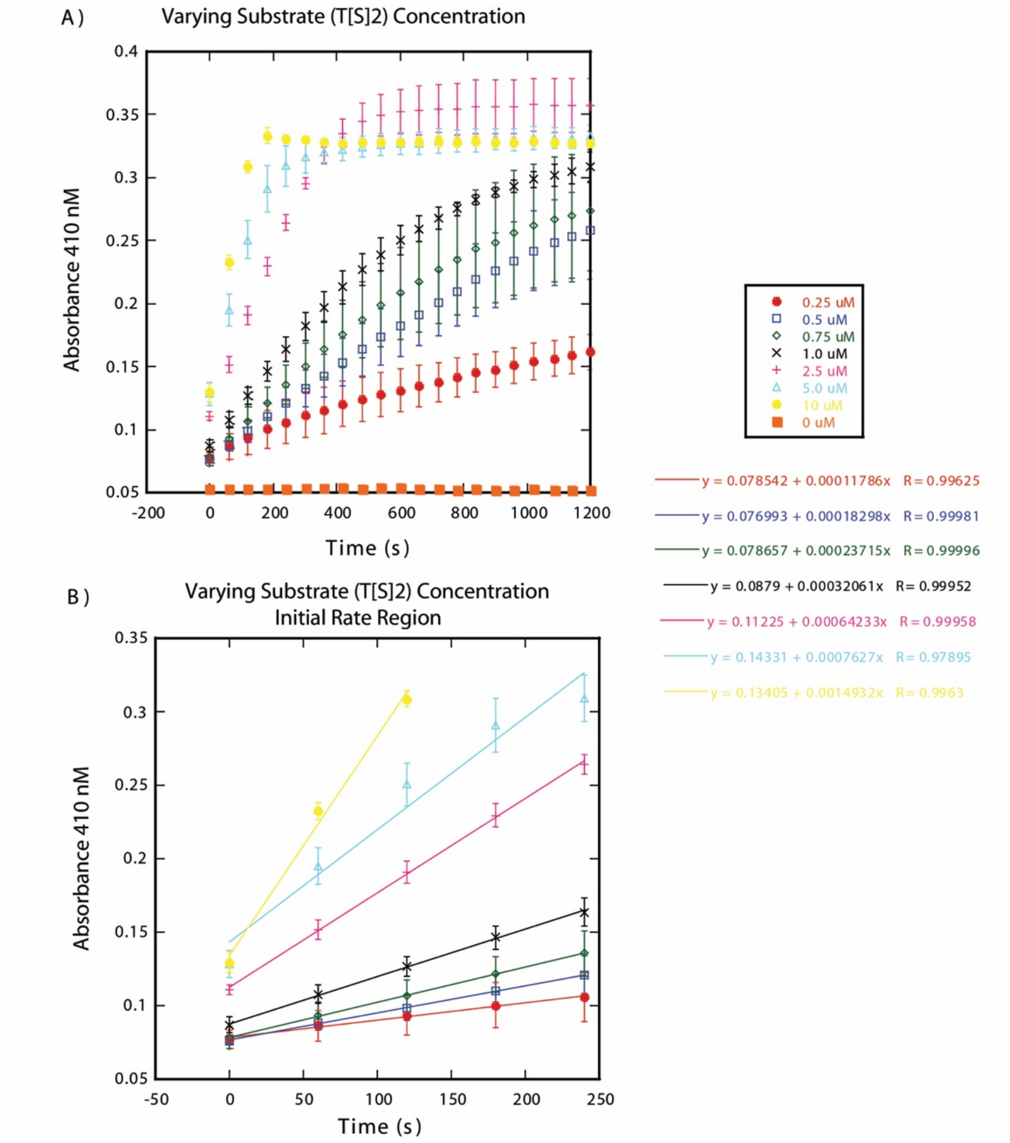


Figure 3.8 Varying substrate concentration in TryR activity assay. Holding the concentration of TryR and varying the amount of trypanothione disulfide ($T[S]_2$) allowed for the calculation of initial rates. A.) Measurement of absorbance at 410 nm for 10 nM TryR with increasing concentrations of $T[S]_2$ substrate as a function of time. B.) Extraction of the initial rate period allowed for calculation of enzyme rates from the slope of the linear treadline. The rate is reported here as absorbance units per second (AU/s).

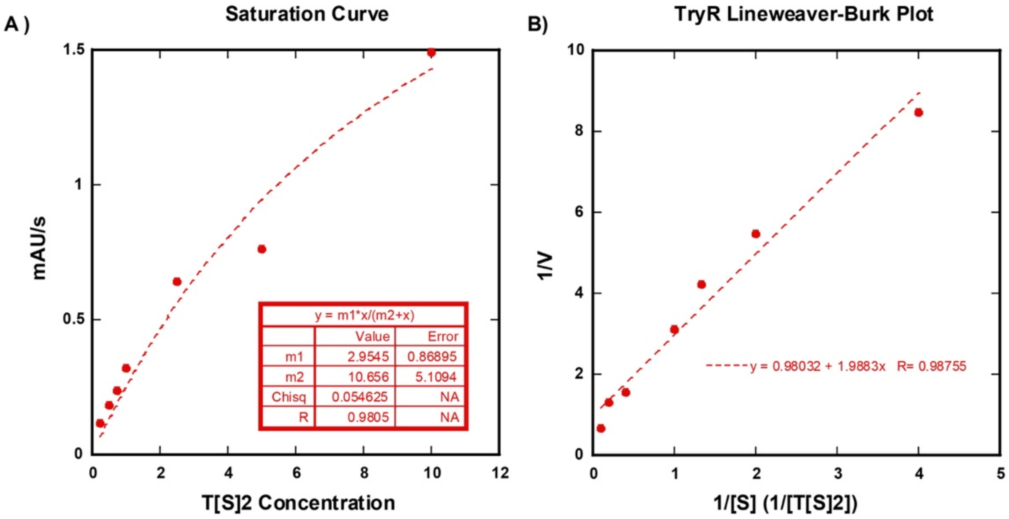


Figure 3.9 The initial rates measured in the enzymatic assay as a function of absorbance units per second can be used to gain kinetic information about TryR using the Michaelis-Menten kinetics. A.) A saturation curve plots the rate (mAU/s) versus substrate concentration and shows the expected non-linear curve fit. B.) The effect of substrate concentration on rate also holds to the expected linearization via a Lineweaver-Burk plot and can be used to calculate the V_{MAX} and K_M.

Table 3.2 Parent peptide sequences for bis-alkylation reactions and development of locked loop libraries to inhibit TryR activity.

Peptide	Sequence											
HL1	Ac-	E	S	G	G	F	G	W	Q	R		-NH2
C1	Ac-	E	C	G	G	F	G	W	C	R		-NH2
C2	Ac-	E	S	C	G	F	G	W	C	R		-NH2
C3	Ac-	E	S	G	G	F	G	W	C	R		-NH2
HL2	Ac-	I	G	V	H	P	T	S	A	E	E	-NH2
P1	Ac-		C	V	H	P	T	S	A	E	C	-NH2
P2	Ac-		C	V	H	P	T	S	A	C	E	-NH2
P3	Ac-	I	C	V	H	P	T	S	A	C	E	-NH2

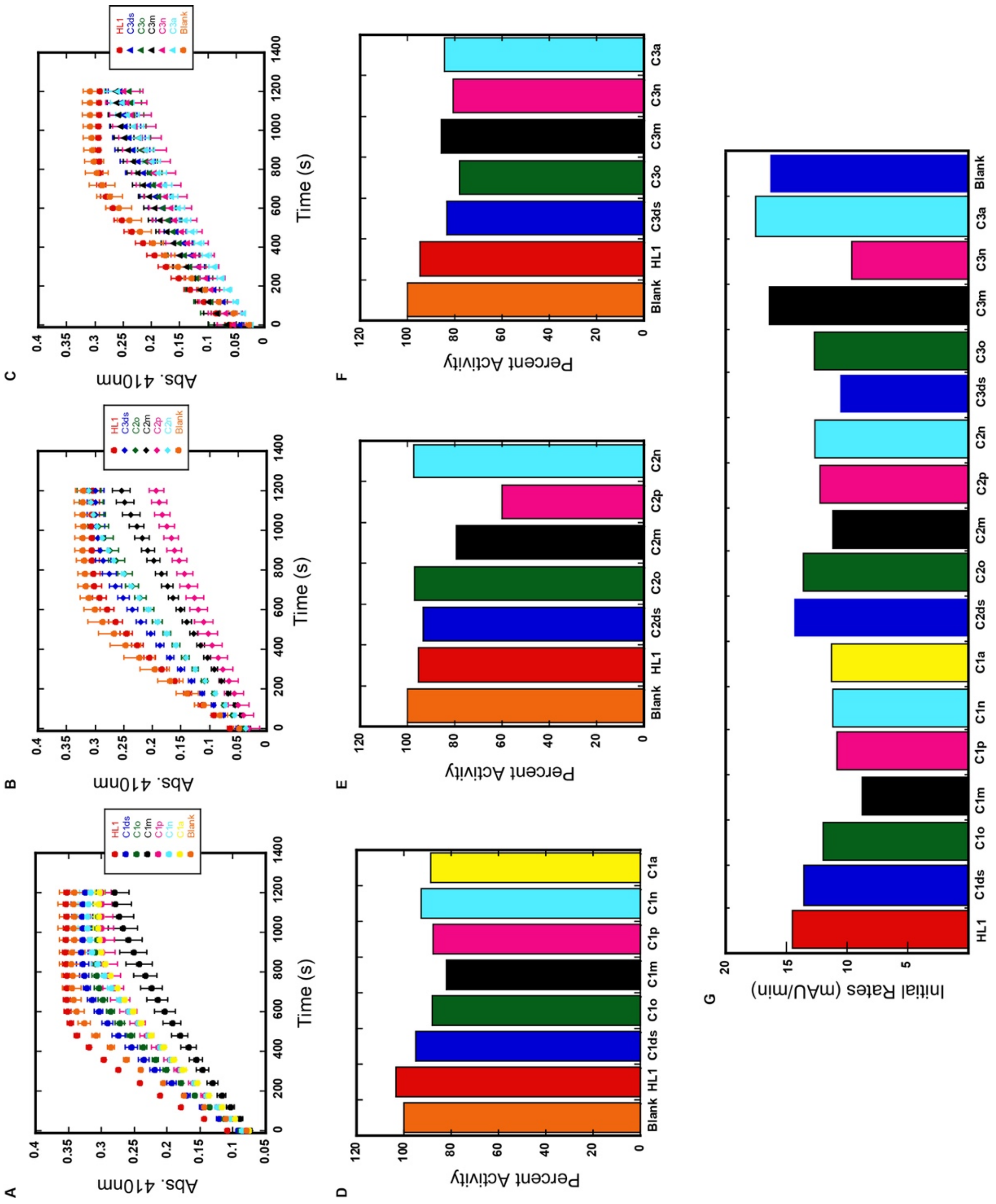


Figure 3.10 Peptides were tested for inhibition of TryR at 25 µM and absorbance at 410 nm was measured over time for the C1 peptide series (A), C2 peptide series (B) and C3 peptide series (C). Data can be interpreted via percent activity as a percentage of the final absorbance measurement for C1 peptides (D), C2 peptides (E) and C3 peptides (F). Another method for analyzing activity of peptides is via the initial rate as measured by the slopes of the curves through the initial linear region (G).

3.3.3 Materials and Methods: TryR Assay Conditions

Trypanothione reductase is easily isolated from bacteria expressing the histidine tagged version of the protein using the methods previously reported by Toro *et al.*[150] Trypanothione reductase was expressed using the pRSET-A-Li-TryR plasmid that was purchased from Life Technologies. The plasmid was then transformed into BL21(DE3) *E. coli* for recombinant expression. Overnights (10 mL in LB-Amp) were prepared and placed in the 37 ˚C incubator on the orbital drum rotor. After 12 hours, the overnight was diluted into 1L of LB-Amp and allowed to grow at 37 ˚C for 4 hours. After 4 hours, the growth cultures were induced with IPTG and allowed to grow overnight at 37 ˚C with shaking at 200 rpms. The cultures were then spun down at 10,000x g for 20 minutes to pellet the bacteria. The cells were then resuspended in 30 mLs of 15 mM Tris, 350 mM NaCl pH 7.5 and sonicated for 12 minutes at 30 second intervals. The suspension was then spun down at 10,000x g for 25 minutes to pellet the cell debris. The lysate was then added to a Ni-NTA column with a bed volume of roughly 5 mL. The initial flow through was collected and then run through the column three times. The column was then washed with 50 mLs of 15 mM Tris, 350 mM NaCl pH 7.5 + 5 mM Imidazole, followed by 50 mLs of 15 mM Tris, 350 mM NaCl pH 7.5 + 25 mM Imidazole, and finally eluted in 10 mL fractions with 15 mM Tris, 350 mM NaCl pH 7.5 + 250 mM Imidazole. Purity was confirmed by SDS-PAGE gel. Fractions containing pure TryR were collected and the protein was separated from the imidazole and buffer exchanged into TRAB- Buffer (40 mM HEPES, 1 mM EDTA, pH 7.5) using a spin concentrator. Concentration was measured using the extinction coefficient for the FAD cofactor (ε_{464} = 11.3 $mM^{-1}cm^{-1}$) yielding a final stock of 15.58 µM TryR. Peptide inhibitors were synthesized according to methods described in 4.8.1.

The Trypanothione Reductase assay was carried out according to the procedure reported by Hamilton *et al.*[142] A 37.2 nM stock of TryR was prepared in TRAB-, and 10 µL was added to each well of a 384 well plate (Greiner Bio-One 781186) to yield a final TryR concentration of 9.3 nM. Peptides were diluted to 133 µM in TRAB- Buffer, and the appropriate volume of peptide solution was added to each well. TRAB- Buffer was then added to bring the total volume of the well to 37 µL. This solution of protein with peptide was incubated for thirty minutes. A 1:1:1 mixture of 40 µM $T[S_2]$, 6 mM NADPH, and 960 µM DTNB was prepared, and 3 µL of this stock was added to each well immediately before reading on the Tecan. This yields final substrate concentrations of 1 µM $[TS_2]$, 150 µM NADPH, and 25 µM DTNB. The absorbance of each well was read at 410 nm on the Tecan with 1-minute increments for 30 minutes.

4 Chapter 4 : Development of a hot loop into a locked loop inhibitor of the Stonin2-Eps15 Interaction

4.1 The Stonin2-Eps15 interaction

In this chapter, we describe a synthetic approach for developing cyclic peptide binders of the epidermal growth factor pathway substrate 15 (Eps15) EH2 (Eps15 Homology 2) domain. The success of this strategy validates the selection of hot loops as starting points for inhibitor design. Eps15 is an important member of the endocytosis machinery being the first protein involved in clathrin-mediated vesicle formation.[155-158] Eps15 is and is responsible for the recruitment of a variety of adaptor proteins to the cell membrane. Eps15 interacts with accessory proteins such as epsin, synaptojanin-p170 and stonin2 to help recruit clathrin to the cell surface via its three EH domains.[159-165] The interactions are mediated by NPF motifs on the binding partners. Upon binding to Eps15 through the second of its EH domains (EH2), stonin2 recruits the clathrin adaptor complex (AP-2), which in turn recruits clathrin to begin formation coated pits and subsequently induce endocytosis.[163,166] Eps15 is an intriguing target for inhibition due to its diverse roles in receptor recycling and controlling nutrient and viral uptake into cells.[156,167]

There are limited specific tools for studying distinct steps within the overall endocytosis pathway and the field could greatly benefit by new inhibitors that are highly specific.[168] A recurring problem in the development of endocytosis inhibitors to freeze cellular uptake via either clathrin-dependent endocytosis

(CDE) or clathrin-independent endocytosis (CIE) is the lack of specificity of the protein targets involved. For example, a key target for blocking endocytosis is Dynamin-2, a GTPase protein that catalyzes membrane fission mostly in CDE.[169] Dynasore, a small molecule inhibitor of this protein, blocks the GTPase activity of dynamin and halts endoycytosis.[170] Unfortunately, dynamin is involved in a variety of other pathways including vesicle trafficking, microtubule stability, and actin cytoskeleton dynamics.[168] The non-specific role of Dynamin prevents the selective inhibition of CDE, and is a common trend among other endocytosis inhibitors such as chlorpromazine and Pitstop-2.[168] The same holds for specific inhibitors developed against CIE such as filipin.[168] Another issue associated with methods of blocking endocytosis via potassium depletion, methyl-β-cyclodetrin, chlorpromazine and genistein is that they show drastically different activities based on cell line. They often have significant cytotoxicity.[171] Finding a target-specific inhibitor for a specific step within the CDE pathway would help better identify modes of cellular uptake of different molecules and pathogens that require CDE for entry.

Genetic methods have helped identify the role of CDE under various conditions. A dominant negative mutant form of clathrin has been used to study the role of CDE and formation of clathrin-coated vesicles on the trafficking of MHC-II molecules.[172] Premature uncoating of clathrin-coated vesicles has also been genetically induced using the expression of an auxilin mutant or via overexpression of an accessory protein (AP180), both of which cause mislocalization of clathrin.[173] A genetic method that gives hope to the utility of specific Eps15-EH2 inhibitors was the inhibition of CDE through the expression of a dominant-negative mutant of Eps15 lacking its three EH domains.[159,166] This mutant, when expressed in HeLa cells, shows inhibition of transferrin uptake.[159] It was also shown that this same mutation greatly reduces the infectivity of *Ebola virus*.[167] We aim to reproduce this type of phenotype through chemical inhibition of an Eps15 EH domain.

When aiming to specifically target the second EH domain of Eps15, it is necessary to understand the function and recognition sequences of other EH domain containing proteins. The EH domain (EHD) is a modular domain that recognizes variants of the NPF-motif. These proteins are often found within key regulators of endocytic transport.[174,175] While Eps15 is involved in the first step of CDE, other EHD proteins are involved in later stages of endocytic recycling.[176] The interaction between EHD and the NPF motif has been the subject of many structural studies, resulting in the characterization of a variety of peptides bound to various EHD target proteins (Table 4.1). Many of the peptides previously developed to bind Eps15-EH2 are weak μM binders and were based on the NPF motif from Rab/Hrb.[177-179] A disulfide cyclized cyclic peptide targeting Eps15-EH2 identified via phage display was mentioned to have a K_d of 12 μM, though no experimental evidence of this binding interaction was ever published in the literature.[178] Peptides have been developed that target other EHD proteins including Reps1 EH and EHD1.[180,181] The Kritzer lab has been developing cyclic peptides to target EHD1, a protein involved in long loop recycling of integrin receptors, identifying a head-to-tail cyclic peptide containing the NPF motif sequence.[182] It is important to note that the flanking sequences of the NPF motif vary depending on the partner EH domain due to differences in surface charges, despite the high structural homology of the domains.[183,184]

LoopFinder identified a hot loop epitope from the stonin2:Eps15-EH2 interaction. We aimed to use this information to develop a cyclic peptide loop mimetic to act as an inhibitor. As the previous literature showed binding partner specificity mediated by the NPF neighboring residues, we predicted that we could use the whole hot loop to impart binding specificity to target Eps15 over other EHD proteins.

Table 4.1 Known peptide binders to EH domain containing proteins form the literature.

EH Domian	**Peptide**	**Source**	**Avg. K$_d$**	**Ref.**
Eps15 EH1	SSST**NPF**L	Rab/Hrb	590 µM	[185]
Eps15 EH2	PTGSSST**NPF**L	Rab/Hrb	560 µM	[178,179]
Eps15 EH2	PTGSSST**NPF**R	Rab/Hrb	560 µM	[178,179]
Eps15 EH3	PTGSSST**NPF**L	Rab/Hrb	560 µM	[178,179]
Eps15 EH3	RTAAPG**NPF**RVQ	Synaptojanin	High µM	[178,179]
Reps1 EH	YEST**NPF**TAKF	Rab11-FIP2	46 µM	[186,187]
Reps1 EH	PRKK**NPF**EESS	Rab11-FIP2	423 µM	[187]
Reps1 EH	IPDS**NPF**DATA	Rab11-FIP2	7 mM	[187]
Eps15 EH2	SSDC^{C}T**NPF**RSC^{C}WRS	Phage Display	12 µM	No data
Reps1	EYEC^{C}T**NPF**TAKC^{C}	Phage Display	65 µM	[186]
EHD1	**NPF**EEEEED	MICAL-L1	57 µM	[183]
EHD1	Cyclo-Y**NPF**EEGG-	MICAL-L1	9.9 µM	[182]

4.2 Hot loop Identification of the NPF-motif Region

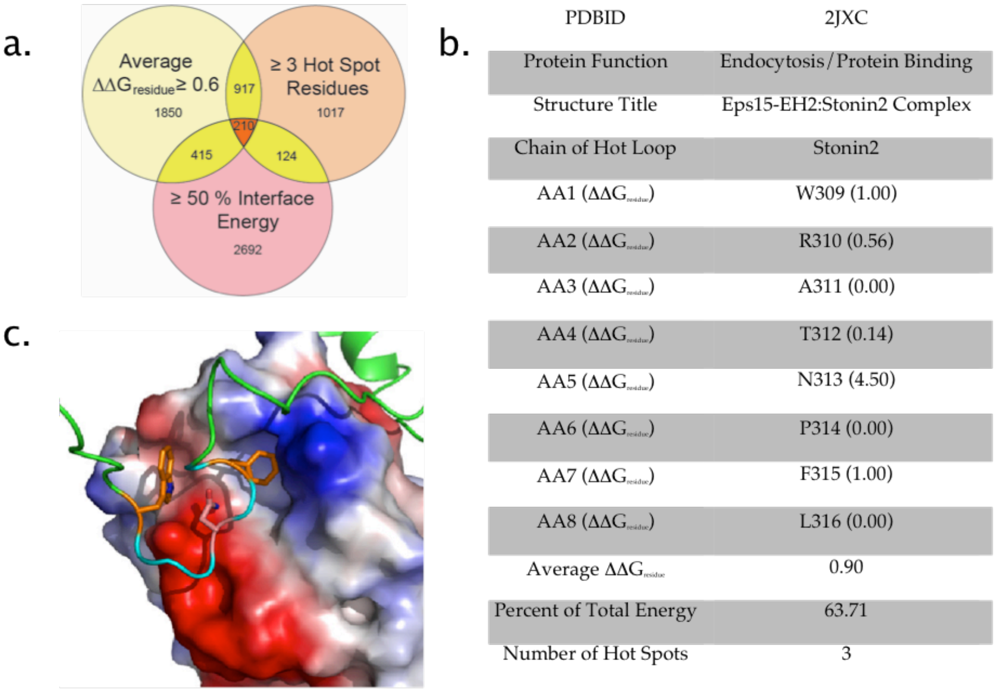
LoopFinder identified many hot loops that satisfied all three hot spot criteria, having an average $\Delta\Delta G_{res} \geq 0.6$ REUs, more than three hot spot residues ($\Delta\Delta G_{res} \geq 0.6$ REUs), and comprising ≥ 50% of the total interface energy (Figure 4.1A). We hypothesize that these hottest of the hot loops are the best set of loops for the direct design of cyclic peptides that inhibit PPIs, differentiating this hot loop selection from the inconclusive attempts discussed in Chapter 3. LoopFinder identified an eight-amino-acid hot loop on stonin2 at the interface with Eps15-EH2 within the crystal structure 2JXC.[163] The eight-amino-acid loop is composed of the NPF motif as well as N-termial residues responsible for specificity to the Eps15-EH_2 target (Figure 4.1B). The hot loop from Trp309 to Leu316 has three predicted hot spots at Trp309, Asn313, and Phe315 (Figure 4.1C). This loop has an average $\Delta\Delta G_{res}$ of 0.9 REUs, comprises about 64% of the total calculated energy for the interface, and has a termini distance of 5.74 Å. LoopFinder identification of the Stonin2 hot loop as responsible for most of the binding interaction with the Eps15-EH2 domain makes it a prime target for mimicry with cyclic peptides.

To test the ability of the hot loop alone to bind the target, a ten-amino-acid peptide from S308 to L316 (**lST2**) was synthesized according to Scheme 4.1 and tested for binding the Eps15-EH2 domain via ITC. The linear hot loop peptide had a K_d of 18.2 ± 3.4 μM (Figure 5.1). It is important to note that the stonin2 39 amino acid peptide composed of residues 301 to 340 had a reported K_d of 0.15 ± 0.02 μM for Eps15-EH2.[163] This is the most accurate representation of the full strength of the PPI that occurs between Eps15-EH2 and stonin2, and as a result a 0.15 μM affinity may the limit for which peptides can achieve. The peptide sequence PTGSSTNPFL, designed from the RAB protein, had a reported K_d of 560 ± 10 μM measured by surface plasmon resonance (SPR).[177] A disulfide-cyclized peptide with sequence SSDC^{C}TNPFRSC^{C}WRS was identified by phage display against intersectin EH domains, which was mentioned to have a K_d of 12 ± 2 μM to Eps15-EH2, though no actual results or data have been published that support this claim.[178,188] Our goal was to improve the binding affinity of the hot loop peptide to an affinity on par with the 39 amino acid peptide through cyclization.

We then asked was how to introduce a structural constraint to stabilize the loop conformation. Many of the advances in structure-based peptide design have focused on the mimicry of protein secondary structure elements. β-strands are readily mimicked by a variety of small molecule and peptide scaffolds.[189] A wide range of techniques are also used to stabilize or mimic α-helices such as side-chain-to-side-chain crosslinks[190], olefin stapling,[38] hydrogen bond surrogate backbone-to-backbone crosslinks[40], and introduction of unnatural β-amino acids[43], or terphenyls[45]. While these strategies can stabilize secondary structure. There are few methods for stabilizing loops except for very well structured β-hairpins and isolated β-turns.[191,192] It is also difficult to predict cyclic peptide structure using computational models, making rational design difficult.

We aimed to develop a method to quickly screen a range of peptide structures by introducing a late stage structural diversity element into our synthetic approach (Figure 4.2). We adapted bis-alkylation peptide chemistry pioneered by Timmerman due to the robustness of the reaction, which would allow for more rapid production of dozens of cyclic peptides with different 3D structures (locked loop libraries).[139] Our design of locked loop libraries was also inspired by phage display bicyclic peptides that used tris-alkylation linkers, developed by Heinis.[193] Ribosome display and mRNA display techniques have also used bis-alkylation chemistry to screen cyclic peptides.[98,194,195] Bis-alkylation linkers were also previously screened for their ability to stabilize α-helical structure. In that work, cysteine residues were introduced at *i* and *i+4* positions within a peptide, revealing di-bromo-*m*-xylene as the optimal linker for α-helix stabilization.[196] The authors let peptide conformation guide linker choice. To stabilize hot loops, we needed to develop an alternate approach by which we allow the linker identity to guide different peptide conformations. By holding the hot loop epitope constant and allowing the linker to impose varations in global peptide conformation, we can screen for binding and identify the peptide that attains the best conformation for binding.

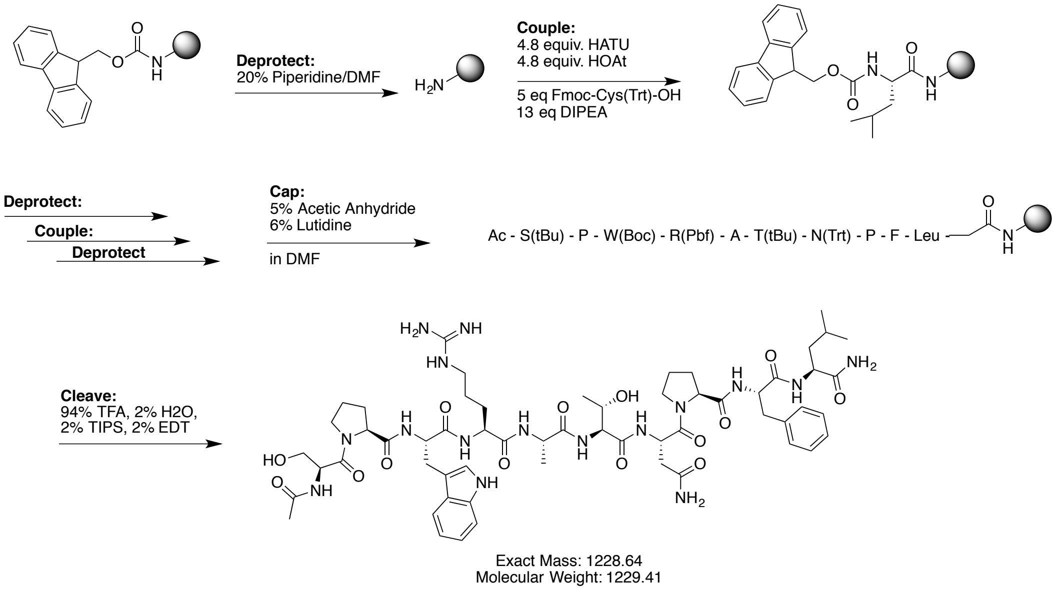
For the design of locked loop libraries based on hot loops, we synthesize a parent peptide containing the hot loop epitope flanked by thiol-containing amino acids which could include cysteine, homocysteine, or penicillamine (Figure 4.2). Using a panel of linker molecules containing two bromomethyl groups, we can then bis-alkylate the peptides to cyclize them. In this way, we arrive at a panel of new peptides containing various cyclization linkers from one parent sequence. The introduction of this late-stage diversity element into the peptides allows for the efficient synthesis of a panel of peptides with different conformations. Testing of these peptides in binding assays will elucidate the peptide that achieves the optimal conformation.



PDBID	2JXC
Protein Function	Endocytosis/Protein Binding
Structure Title	Eps15-EH2:Stonin2 Complex
Chain of Hot Loop	Stonin2
AA1 ($\Delta\Delta G_{residue}$)	W309 (1.00)
AA2 ($\Delta\Delta G_{residue}$)	R310 (0.56)
AA3 ($\Delta\Delta G_{residue}$)	A311 (0.00)
AA4 ($\Delta\Delta G_{residue}$)	T312 (0.14)
AA5 ($\Delta\Delta G_{residue}$)	N313 (4.50)
AA6 ($\Delta\Delta G_{residue}$)	P314 (0.00)
AA7 ($\Delta\Delta G_{residue}$)	F315 (1.00)
AA8 ($\Delta\Delta G_{residue}$)	L316 (0.00)
Average $\Delta\Delta G_{residue}$	0.90
Percent of Total Energy	63.71
Number of Hot Spots	3

Figure 4.1 (a) Hot loops are organized according to their fulfillment of the desired hot loop criteria. As portrayed in the Venn diagram, 210 loops satisfy all three hot loop criteria. We hypothesize that these are ideal starting point for the design of cyclic peptides that bind a protein target. (b) An example of a PPI that resides in the central overlapping region of the Venn diagram is the interaction between stonin2 and Eps15-EH2 which is mediated by a hot loop from W309 to L316 on stonin2. (not final figure quality yet). (c) The overall structure of the peptide shoes the positioning of the three hot spots at the surface of Eps15-EH2 (PDBID: 2JXC).

Scheme 4.1 Synthesis of **IST2** peptide on solid phase.



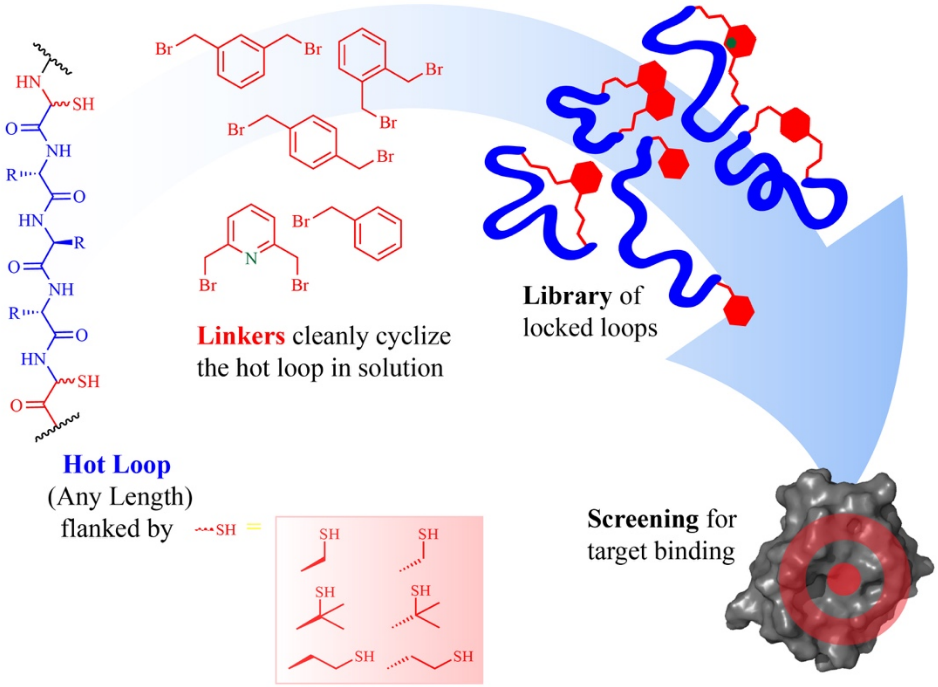


Figure 4.2 Design of locked loop libraries. Flanking the hot loop with two thiol-containing amino acids allows for subsequent bis-alkylation with dibromine containing aryl linkers to produce a cyclic peptide possessing various conformations. This panel of peptides can then be used to screen for target binding.

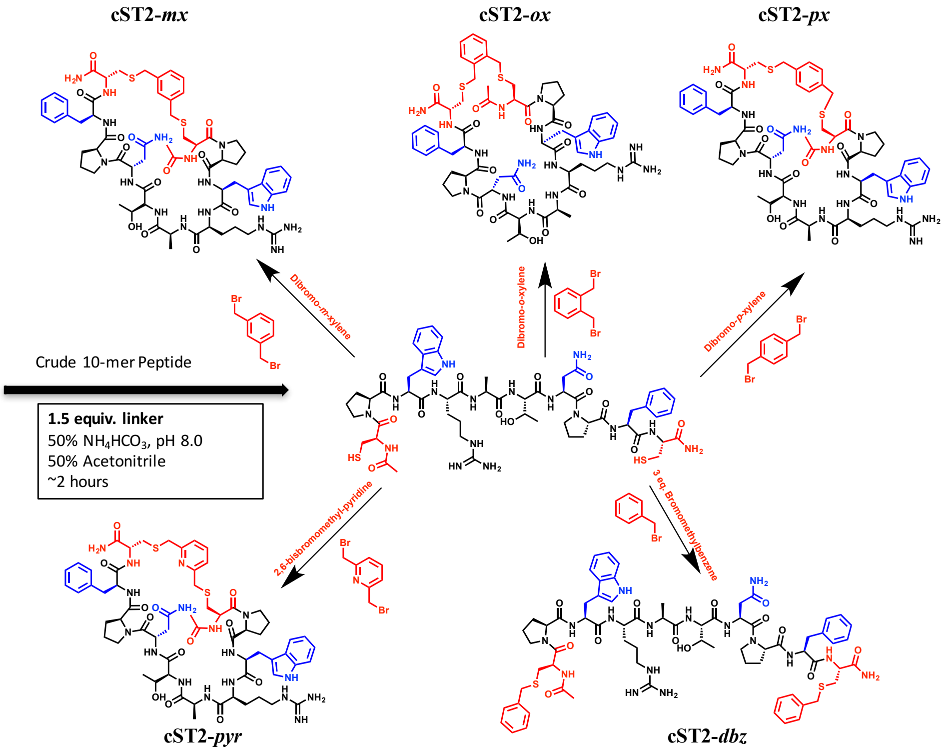
4.3 Assessing peptide binding to Eps15 target by ITC

We applied our locked loop library development approach to enhance binding of the stonin2 hot loop peptide described earlier. Replacement of S308 and L316 with cysteine residues yields a parent peptide with the sequence Ac-CPWRATNPFC-NH_2, synthesized via solid phase peptide synthesis. The crude peptide was divided and subjected to cyclization via the bis-alkylation reaction using a small panel of linkers: di-bromo-*o*-xylene (***ox***), di-bromo-*m*-xylene (***mx***), di-bromo-*p*-xylene (***px***), and 2,6-di(bromomethyl)pyridine (***pyr***). A linear control peptide was also synthesized using benzylbromide (***dbz***) (Scheme 4.2). Upon successful synthesis and purification, the locked loop library was then analyzed for binding to Eps15-EH2.

Results from the ITC binding experiments revealed an interesting trend (Table 4.2). Peptides **cST2-*ox*** and **cST2-*mx*** had similar binding affinity with K_d of 0.326 ± 0.011 μM and 0.373 ± 0.024 μM, respectively. The *para*-substituted peptide **cST2-*px*** exhibits a ten-fold loss in binding affinity with a K_d of 3.67 ± 0.17 μM. We hypothesize that the loss of binding affinity to the target is due to a conformational restriction that disfavors the optimal conformation adapted in **cST2-*ox*** and **cST2-*mx***. Using this approach, we can not only vary linker substitution, but also linker electronics by replacing meta-xylene with di(bromomethyl)-pyridine. This yields the cyclic peptide **cST2-*pyr***. This peptide contains the same linker substitution as **cST2-*mx***, but with different electronics imparted by the nitrogen contained in the pyridine ring. **cST2-*pyr*** loses five-fold affinity compared to the similar **cST2-*mx*** peptide, having a K_d of 1.53 ± 0.05 μM. Additionally, as a control to ensure the addition of aromatic groups within the peptide sequence is not responsible for the increased binding affinity of the cyclic peptides over the linear hot loop, the di-benzyl alkylated peptide **cST2-*dbz*** was tested and no binding was observed against Eps15-EH2 via ITC within the solubility range of the peptide. The only changes made within this panel of peptides was the linker, and these result suggests the most likely cause for changes in affinity is conformational restriction.

Alanine mutants of the **cST2-*ox*** peptide were synthesized to further characterize its mode of binding. These mutants also serve to validate the computational alanine scan results utilized by LoopFinder to detect hot loops. Three such peptides were made that individually substitute each hot spot residue to alanine. Substitution of the tryptophan residue resulted in a >5-fold loss in binding affinity (K_d = 1.70 ± 0.28 μM, Table 4.2). Using the ITC comparison between **cST2-*ox*** and **cST2-W3A-*ox*** we can estimate the actual $\Delta\Delta G_{res}$ of the tryptophan to be 0.98 kcal/mol. It is important to keep in mind that the energy value LoopFinder assigned to this tryptophan was < -2 REU, which we substituted with 1.0 as described in Chapter 2.1.2. The binding results indicated that, in this case, this was a reasonable modification. Mutation of either of the asparagine (**cST2-N7A-*ox***) or phenylalanine (**cST2-F9A-*ox*)** hot spots that reside within the canonical NPF motif results in a complete loss of binding Table 4.2. These results support our hypothesis these peptides bind region on the surface of Eps15-EH2 to which the natural stonin2 protein binds.

Scheme 4.2 Synthesis of cyclic peptides via bis-alkylation. The hot spots are in blue and the linker region including cysteine residues are in red.



Peptides	Sequence	N (Sites)	K$_d$ (μM)^a	ΔH (kcal/mol)	TΔS (kcal/mol)
cST2-*ox*	Ac-CPWRATNPFC-NH$_2$	1.11 ± 0.07	0.33 ± 0.01	-13.8 ± 0.4	-4.9 ± 0.4
cST2-*mx*	Ac-CPWRATNPFC-NH$_2$	0.95 ± 0.11	0.37 ± 0.02	-15.1 ± 0.1	-6.2 ± 0.1
cST2-*px*	Ac-CPWRATNPFC-NH$_2$	0.98 ± 0.08	3.67 ± 0.17	-12.5 ± 0.4	-5.1 ± 0.4
cST2-*pyr*	Ac-CPWRATNPFC-NH$_2$	1.01 ± 0.13	1.53 ± 0.05	-11.7 ± 0.4	-3.7 ± 0.4
lST2	Ac-SPWRATNPFL-NH$_2$	1.14 ± 0.07	18.2 ± 3.38	-12.2 ± 0.8	-5.8 ± 0.9
cST2-W3A-*ox*	Ac-CPARATNPFC-NH$_2$	1.39 ± 0.01	1.70 ± 0.28	-10.2 ± 1.3	-2.3 ± 1.2
cST2-N7A-*ox*	Ac-CPWRATAPFC-NH$_2$	no binding			
cST2-F9A-*ox*	Ac-CPWRATNPAC-NH$_2$	no binding			

Table 4.2 Dissociation constants for peptides binding to their Eps15-EH2 partner were calculated using ITC. *a* ITC experiments were run in 25 mM MOPS, 150 mM NaCl and 1 mM $CaCl_2$ at pH 7.0

4.4 Fluorescence Competition Assay for Eps15 Binding

ITC was initially used to identify peptide binders because there have been no other reported assays for the measurement of ligand binding to Eps15-EH2. One of the main disadvantages of ITC is the requirement for high concentrations of both target protein and ligand. ITC is also time-consuming, as only one sample can be run at a time and each run takes about an hour and a half to complete. We sought to utilize our sub-μM binding peptide, **cST2-*ox***, to develop the first assay for identifying Eps15-EH2 ligands via a more high-throughput method. We converted the **cST2-*ox*** peptide into a fluorescent probe, **Flu-cST2-*ox*** (Figure 4.4), for use in fluorescence polarization (FP) assays. Two β-alanine amino acids were introduced at the N-terminus of **cST2-*ox*** as a spacer, followed by a fluorescein tag. A direct binding analysis can be performed by measuring the anisotropy of the FP probe after incubation with increasing concentrations of protein. Using this technique, we measured a K_d of 0.511 ± 0.065 μM (Figure 4.4). The dissociation constant measured by FP is two times higher than that measured by ITC.

Eps15-EH2 is only one example of a whole class of EH-domain-containing proteins, all of which have the ability to recognize NPF motifs.[155] To assess the specificity of **Flu-cST2-*ox*** towards the Eps15-EH2 domain over other EH domains, EHD1-EH and REPS1-EH domains were expressed and purified from bacteria. Direct binding of **Flu-cST2-*ox*** to these proteins were measured by FP.

The K_d of **Flu-cST2-*ox*** to EHD1 and REPS1-EH were 15.8 ± 1.1 μM and 14.8 ± 1.7 μM respectively (Figure 4.4). We hypothesize that the observed 30-fold selectivity of **Flu-cST2-*ox*** to Eps15-EH2 over these other EH-domain containing proteins arises from the amino acids N-terminal to the NPF motif most notably, the tryptophan hot spot residue that does not exist on the consensus sequence of binding partners of EHD1-EH and Reps1-EH proteins. In addition, as observed in the binding modes of NPF-containing peptides with their EH-domain-containing partners via NMR studies, the surface electrostatics of Eps15 and EHD1 EH domains are very different, and hence would accommodate different amino acids at positions flanking the NPF motif (Figure 4.3). In cell-based experiments to measure effects on vesicle trafficking, we can be more confident that we will be minimizing off-target effects.

The binding of **Flu-cST2-*ox*** to Eps15-EH2 allows for the design of a competition assay to quickly and efficiently identify Eps15-EH2 binding molecules in a medium-throughput manner. By measuring displacement of the probe upon introduction of other molecules, we can now measure binding of a larger panel of peptides with less ligand and less protein than is required for ITC experiments. Each peptide was tested in triplicate and IC_{50} values were calculated based on the fraction of probe displaced across an increasing concentration of peptide in the system (Figure 4.5). **cST2-*ox*** was measured to have an IC_{50} value of 2.15 ± 0.26 μM. Within error, **cST2-*mx*** was measured to have a similar IC_{50} value of 2.33 ± 0.22 μM. These results are consistent with their similar K_d values measured by

ITC. The *para*-subsitituted peptide **cST2-*px*** had an IC_{50} value of 41.3 μM, a slightly larger drop in affinity than was observed by ITC. **cST2-*pyr*** exhibited a IC_{50} of 9.09 ± 3.26 μM, a similar five-fold drop in affinity from the ***mx*** substituted variant. This was similar to the K_d difference observed by ITC. One of the issues with ITC was that peptides needed to be prepared at high concentrations in buffered solution, making it difficult to test peptides that were poorly soluble at high concentrations without DMSO. FP experiments can be run in 1% DMSO and with much less peptide in the 396 well plate format. As a result, we were able to measure an IC_{50} value of 27.5 ± 8.3 μM for **cST2-*dbz***. This is an interesting result considering it binds better than the **cST2-*px*** peptide in this FP experiment. It is possible that there may be some π-π interactions that occur between the benzyl thioether caps, allowing the peptide to access a cyclic conformation. Results for peptides with hot spots mutated to alanine were consistent to what was observed by ITC. We observed a 5-fold loss in IC_{50} upon substitution of the tryptophan, and no binding was observed for either the asparagine mutant or the phenylalanine mutant (Figure 4.6).

We aimed to assess the optimal macrocycle size by adjusting the position of the cysteine residues flanking the hot loop. To do this, we tested a different series of peptides **cST1** series in which the proline residue at the second position within the peptide was removed. The parent peptide sequence of Ac-CWRATNPFC-NH_2 was bis-alkylated with the ***ox***, ***mx***, ***px***, and ***pyr*** linkers. Peptides were tested in the FP competition assay. These peptides showed much weaker affinity to the target

(Figure 4.7). Though weaker in binding, this series shows the efficiency of the approach whereby two initial parent peptide sequences can be synthesized and addition of the five linkers results in ten peptides probing various regions of conformational space.

An iterative design approach was taken to improve peptide affinity for Eps15-EH2. The first changes made were to the termini of the peptides. A pentynyl cap was introduced at the N-terminus to replace the acetyl cap in the original peptide (**p-cST2-*ox***). **p-cST2-*ox*** had a measured IC_{50} of 1.06 ± 0.10 μM, which is about a 2-fold improvement over **cST2-*ox***. With the N-terminus improved, a peptide lacking a C-terminus was synthesized (**p-cST2b-*ox***) using cysteamine resin for solid phase peptide synthesis of the peptide. This iteration showed no clear improvement over **cST2-*ox*** with an IC_{50} of 2.16 ± 0.18 μM (Figure 4.8). Despite the minimal gain in binding affinity, slight changes in number of hydrogen bond donors or acceptors within the peptide may improve cell penetration, and the introduction of the alkyne on **p-cST2-*ox*** can be used as a handle for the attachment of other probes or warheads in cell-based studies.

Further iterations included replacing W3 (**cST4-*ox/mx***) or F9 (**cST5-*ox/mx***), or both (**cST6-*ox/mx***) with 1-naphthylalanine (Figure 4.9). The naphthyl side-chain is accommodated at the tryptophan position, as the IC_{50} values for **cST4-*ox*** and **cST4-*mx*** are 1.62 ± 0.70 μM and 1.07 ± 0.18 μM, respectively. This is a ~2-fold improvement, on par with the best binders yet synthesized. By contrast, the IC_{50}

values for **cST5-*ox*** and **cST5-*mx***, where F9 is substituted with 1-naphthylalanine, were reduced to ~25 μM, a 10-fold reduction in affinity. This suggests the NPF recognition pocket on Eps15-EH2 does not accommodate an extra aromatic ring at the phenylalanine position. When both point mutations were combined (**cST6-*ox*** and **cST6-*mx***), binding could not be measured below 25 μM. This series identified our strongest binder of Eps15 to date. Interestingly, it omits the possibility of polar contact using the indole ring of the tryptophan residue.

As discussed earlier, shape diversity can be imposed not only by the di-bromomethyl linkers, but can also be imparted by selection of thiol containing amino acid that flank the hot loop sequence. In this case, penicillamine was used to make three different parent peptides for bis-alkylation. In the **cST7-*ox/mx*** series, penicillamine is included at both terminal positions. The **cST8-*ox/mx*** series includes the penicillamine residue at the C-terminal position, while **cST9-*ox/mx*** introduces penicillamine at the N-terminus of the peptide (Figure 4.10). Penicillamine introduces two extra methyl groups at the β-position compared to cysteine. We hypothesize that this will affect the conformation of the peptide by limiting the rotation of the linker. The **cST8-*ox*** and **cST8-*mx*** peptides show a minor 2-fold reduction in binding affinity at 4.34 ± 0.64 μM and 5.99 ± 0.85 μM, respectively. The **cST7** and **cST9** series all have IC_{50} values > 25 μM. It appears as though introduction of penicillamine at the N-terminus of the peptide is severely restrictive in the peptides ability to achieve the proper binding conformation. While no major improvement in Eps15-EH2 binding was observed for this series of peptides, added hydrocarbons in the linker may improve the cell penetration of the peptides.

As an important note for the calculation of IC_{50} values via FP competition, we aimed to be transparent in our reporting of the actual experimental error associated with these measurements. As observed in Figure 4.11 IC_{50} values can be calculated in two ways depending on how the data across the three experimental trials is averaged. If we calculate an average value and standard deviation for each data point we can generate a globally averaged curve. Using Kaleidagraph, we can fit the curve and calculate an IC_{50} and the error. This is often reported as the error, but it is not the actual experimental error (Figure 4.11A). A more accurate representation of the error was to average the three IC_{50} values calculated from three isolated experiments. This type of error is consistent with how error is calculated for ITC experiments and is a more realistic representation of experimental error. For all inhibition curves previously discussed, only one representative curve was shown for each peptide. The total set of data used to calculate the average IC_{50} value and associated standard deviation can be found in the Appendix.

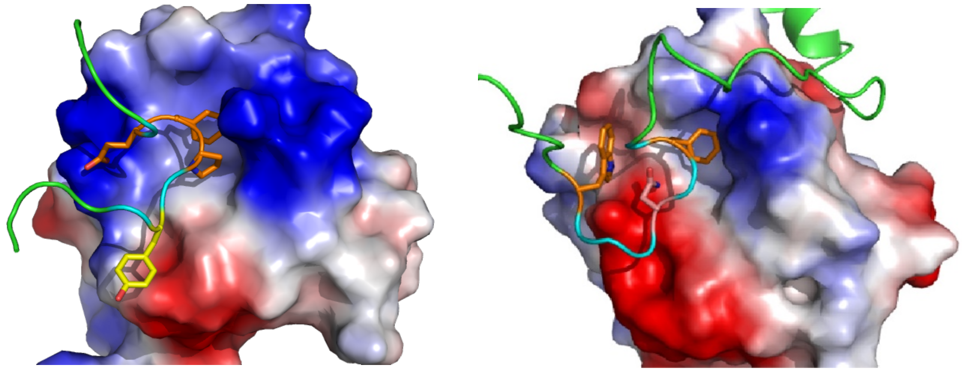
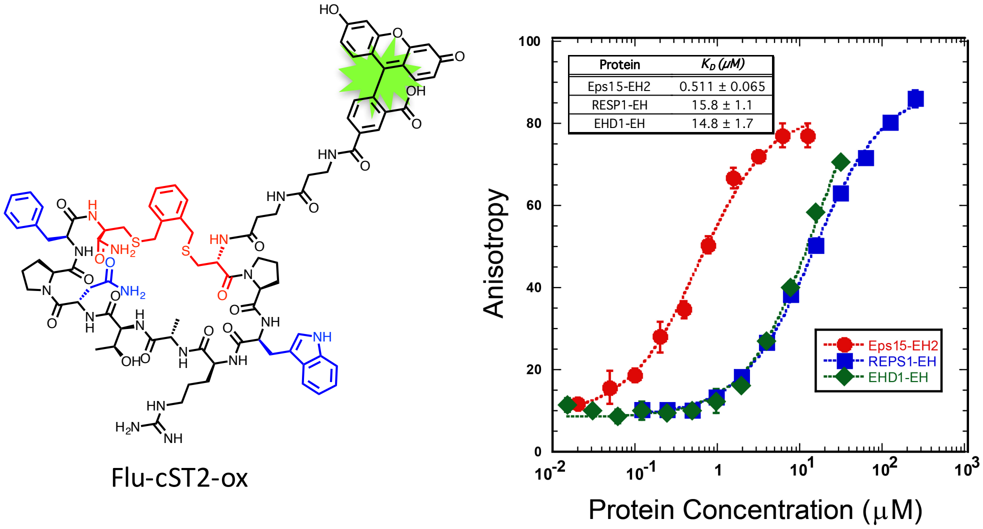
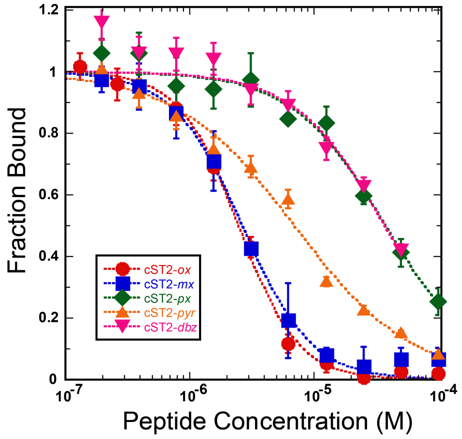


Figure 4.3 (*left*) The surface of EHD1-EH (2KSP[183]) bound to a YNPFEE ligand where four hot spots were identified and colored in yellow and orange. (*right*) The surface of Eps15-EH2 bound to Stonin2 (2JXC[163]).



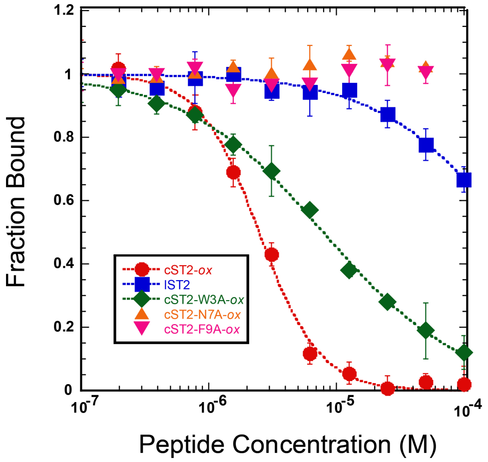
Protein	K_D (μM)
Eps15-EH2	0.511 ± 0.065
RESP1-EH	15.8 ± 1.1
EHD1-EH	14.8 ± 1.7

Figure 4.4 The fluorophore-conjugated peptide **Flu-cST2-*ox*** (*left*) was used as a probe to measure direct binding to EH domains via fluorescence polarization. K_d values were calculated for binding to Eps15-EH2, REPS1-EH and EHD1-EH.



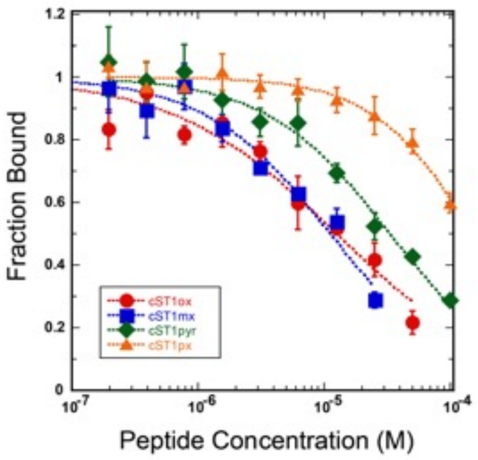
Peptides	Sequence	IC$_{50}$ (μM)
cST2-ox	Ac-CPWRATNPFC-NH$_2$	2.15 ±0.26
cST2-mx	Ac-CPWRATNPFC-NH$_2$	2.33 ±0.22
cST2-px	Ac-CPWRATNPFC-NH$_2$	41.3 ±7.9
cST2-pyr	Ac-CPWRATNPFC-NH$_2$	9.09 ±3.26
cST2-dbz	Ac-CPWRATNPFC-NH$_2$	27.5 ±8.3

Figure 4.5 Inhibition data for the **cST2** series of locked loop inhibitors. Each curve is one representative inhibition curve for each peptide. The average IC$_{50}$ values and associated error are shown in the bottom table.



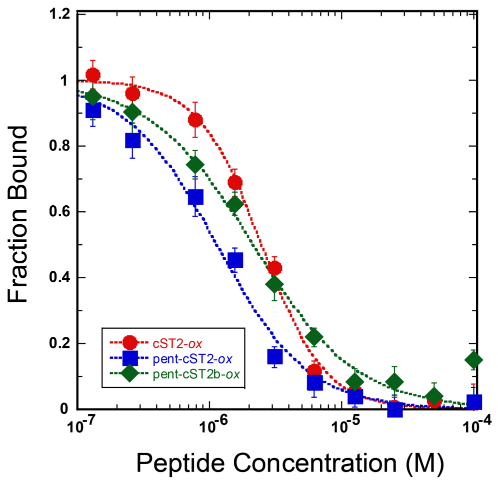
Peptides	Sequence	IC$_{50}$ (μM)
lST2	Ac-SPWRATNPFL-NH$_2$	>100
cST2-ox	Ac-CPWRATNPFC-NH$_2$	2.15 ±0.26
cST2-W3A-ox	Ac-CPARATNPFC-NH$_2$	8.34 ±0.67
cST2-N7A-ox	Ac-CPWRATAPFC-NH$_2$	> 100
cST2-F9A-ox	Ac-CPWRATNPAC-NH$_2$	> 100

Figure 4.6 Inhibition data for the **cST2** control peptides. Each curve is one representative inhibition curve for each peptide. The average IC$_{50}$ values and associated error are shown in the bottom table.



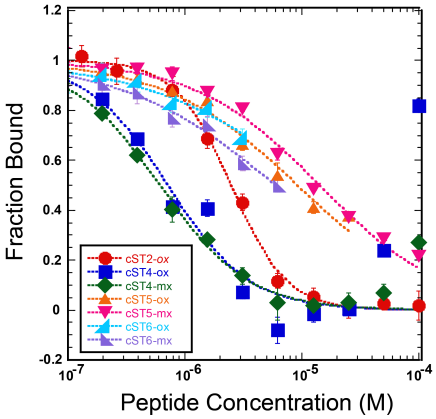
Peptides	Sequence	IC$_{50}$ (μM)
cST2-ox	Ac-CPWRATNPFC-NH$_2$	2.15 ± 0.26
cST1-ox	Ac-CWRATNPFC-NH$_2$	12.5 ± 7.1
cST1-mx	Ac-CWRATNPFC-NH$_2$	11.8 ± 0.5
cST1-pyr	Ac-CWRATNPFC-NH$_2$	41.6 ± 8.5
cST1-px	Ac-CWRATNPFC-NH$_2$	> 100

Figure 4.7 Inhibition data for the **cST1** series. Each curve is one representative inhibition curve for each peptide. The average IC$_{50}$ values and associated error are shown in the bottom table.



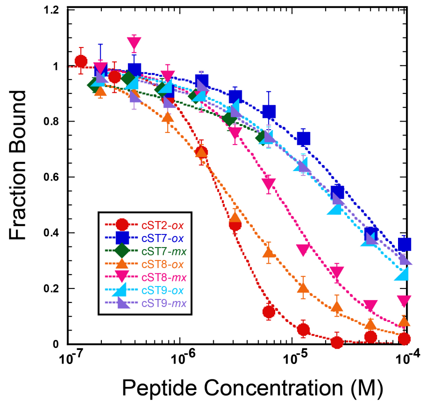
Peptides	Sequence	IC$_{50}$ (μM)
cST2-ox	Ac-CPWRATNPFC-NH$_2$	2.15 ± 0.26
pent-cST2-ox	Pentynyl-CPWRATNPFC-NH$_2$	1.06 ± 0.10
pent-cST2b-ox	Pentynyl-CPWRATNPFX	2.16 ± 0.18
X = Cystamine		

Figure 4.8 Inhibition data for the **cST2** termini modification series. Each curve is one representative inhibition curve for each peptide. The average IC$_{50}$ values and associated error are shown in the bottom table.



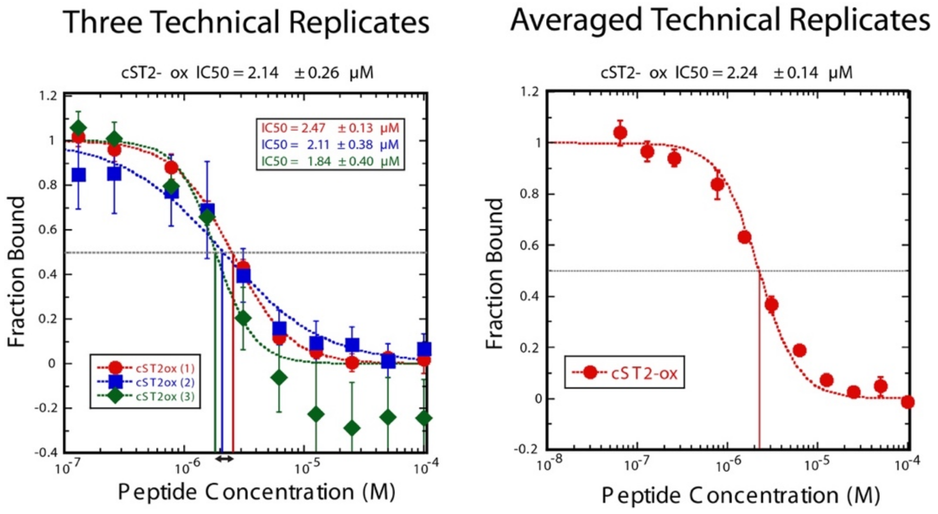
Peptides	Sequence	IC$_{50}$ (μM)
cST2-ox	Ac-CPWRATNPFC-NH$_2$	2.15 ± 0.26
cST4-ox	Ac-CPΦRATNPFC-NH$_2$	1.69 ± 0.70
cST4-mx	Ac-CPΦRATNPFC-NH$_2$	1.33 ± 0.62
cST5-ox	Ac-CPWRATNPΦC-NH$_2$	24.4 ± 12.7
cST5-mx	Ac-CPWRATNPΦC-NH$_2$	30.1 ± 11.3
cST6-ox	Ac-CPΦRATNPΦC-NH$_2$	> 25
cST6-mx	Ac-CPΦRATNPΦC-NH$_2$	> 25
Φ = 1-Naphthylalanine		

Figure 4.9 Inhibition data for the 1-naphthylalanine substituted peptide series. Each curve is one representative inhibition curve for each peptide. The average IC$_{50}$ values and associated error are shown in the bottom table.



Peptides	Sequence	IC$_{50}$ (µM)
cST2-ox	Ac-CPWRATNPFC-NH$_2$	2.15 ± 0.26
cST7-ox	Ac-γPWRATNPFγ-NH$_2$	68.1 ± 24.3
cST7-mx	Ac-γPWRATNPFγ-NH$_2$	> 30
cST8-ox	Ac-CPWRATNPFγ-NH$_2$	4.83 ± 1.53
cST8-mx	Ac-CPWRATNPFγ-NH$_2$	6.82 ± 2.24
cST9-ox	Ac-γPWRATNPFC-NH$_2$	41.3 ± 11.3
cST9-mx	Ac-γPWRATNPFC-NH$_2$	27.0 ± 5.7
γ = Pennicillamine		

Figure 4.10 Inhibition data for the penicillamine-substituted peptide series. Each curve is one representative inhibition curve for each peptide. The average IC$_{50}$ values and associated error are shown in the bottom table.


Three Technical Replicates

Averaged Technical Replicates

Figure 4.11 Calculation of the IC$_{50}$ value and associated error displayed two ways. On the *left* is the value when calculating the average IC$_{50}$ and standard deviation from three individual experiments. On the *right* is the IC$_{50}$ when averaging each data-point across the three experiments.

4.5 Assessing activity in HeLa cell endocytosis assay

Using our locked loop strategy, we identified peptides that bind Eps15-EH2 selectively in biochemical assays. With the goal of developing useful chemical biology tools, we attempted to measure the biological activity of these peptides in a cell-based system. The biomedical relevance of inhibiting the function of the Eps15 EH domain region was previously described.[159] Specifically, genetic deletion of the EH domains resulted in decreased infectivity of human cells with the Ebola virus and decreased internalization of fluorescently-labeled transferrin. [159,167]

To measure the ability of our peptides to inhibit the endocytosis of human transferrin, we adapted a transferrin uptake assay for use with flow cytometry. Figure 4.12 shows the scheme for running the transferrin uptake assay. A known inhibitor of endocytosis, chlorpromazine, was used to test the development of this assay alongside DMSO (vehicle) controls. In the first step, HeLa cells were treated with either DMSO or 10 µM chlorpromazine for one hour in serum-free media at 37 ºC. After the hour, the media was removed and either DMSO-containing media, DMSO and Alexa-488 labelled transferrin in media, or 10 µM chlorpromazine and Alexa-488 labelled transferrin in media was added. Cells were then incubated for one hour at 37 ºC to allow endocytosis of the fluorescently labelled transferrin to occur. After the hour, all cells were treated with unlabeled transferrin and the cells were washed and fixed. Fluorescence intensity of the cells was measured by flow cytometry. The DMSO control measured the background fluorescence of the HeLa cells. The Alexa-488 transferrin-treated cells allowed us to measure the maximum endocytosis of the HeLa cells. Treatment with chlorpromazine inhibits uptake, and a drop in fluorescence was observed. Inhibition of endocytosis can be measured as a drop in fluorescence that is quantified by either a histogram (Figure 4.13B) or by measuring mean fluorescence (Figure 4.13C). Treatment of the HeLa cells with chlorpromazine results in an observed drop of fluorescence from 250 fluorescence units to 150 fluorescence units. This corresponds to a 40% reduction in endocytosis (Figure 4.13), and serves as a benchmark for comparison with the cyclic peptide binders of Eps15.

With a control in hand, we can compare the existing endocytosis inhibitor to the activity of our two best peptides, **cST2-*mx*** and **cST2-*ox***, in this transferrin uptake inhibition assay (Figure 4.14). HeLa cells were treated with increasing concentrations of **cST2-*mx***. As the amount of concentration increases, a decrease in fluorescence intensity is observed (Figure 4.14B). Quantification of the mean fluorescence of the HeLa cells within the plot is shown in Figure 4.14C. Normalization of this data to full uptake observed in the DMSO control cells yields an analysis of the decrease in percent transferrin taken up into the HeLa cells (Figure 4.14D). Unfortunately, the data shows a small drop to around 85% transferrin uptake with concentrations of **cST2-*mx*** from 0.5 μM to 50 μM, then a dose response with increasing inhibition at 100 μM, 150 μM and 200 μM. The cause of the drop in transferrin uptake at lower concentrations has not been fully characterized, and maybe an on-target effect or due to some background toxicity. An identical trend was observed for the treatment of HeLa cells with **cST2-*ox*** (Figure 4.15A,B). To characterize the increased inhibition effect of **cST2-*ox*** at higher concentrations of 100 µM, 150 µM and 200 µM, the flow cytometry raw data was plotted as a scatter plot (Figure 4.15C). It was clear that the observed inhibition at higher concentrations was due to a toxicity effect, because number of cells that fall into the live gated region for forward and side scatter drop significantly at higher peptide concentrations. The effect of **cST2-*ox/mx*** at lower concentrations than 0.5 µM is currently underway to determine whether the mild effect observed at 0.5 µM is dose-dependent. If this is the case, a 20% reduction in transferrin uptake may be the maximum attainable with these peptides, possibly due to the blocking of only one out of the three Eps15-EH domains.

One potential problem with these peptides is that they are not sufficiently penetrating the HeLa cells and gaining access to the cytosol. The difficulties associated with getting peptides across cell membranes is a well-known limitation of using peptides to inhibit intracellular PPIs. Enhancing methodology for understanding how to make peptides more cell penetrant through N-methylation or the addition of cell-penetrating sequences is the subject of intense research.[55,197] Peptide fusions of **cST2-*ox*** with the TAT-polyarginine sequence were synthesized and tested.[198] Unfortunately, the TAT-fused peptides were observed to be highly toxic at concentrations >2.5 µM and disrupted the cell membranes. At concentrations < 2.5 µM, no effect on transferrin uptake was observed. Another issue is that many peptides that can penetrate cells do so via endocytosis.[199] As a result, these **cST2** peptides may be self-inhibitory in that, while inhibiting endocytosis, they also inhibit the penetration of more peptide at higher concentrations. This could account for a plateau in the activity of the peptides, as there could be a maximum concentration achieved within the cells. Further investigation into the ability of these peptides to inhibit endocytosis is needed to fully characterize the observed inhibition at 0.5 μM concentrations.

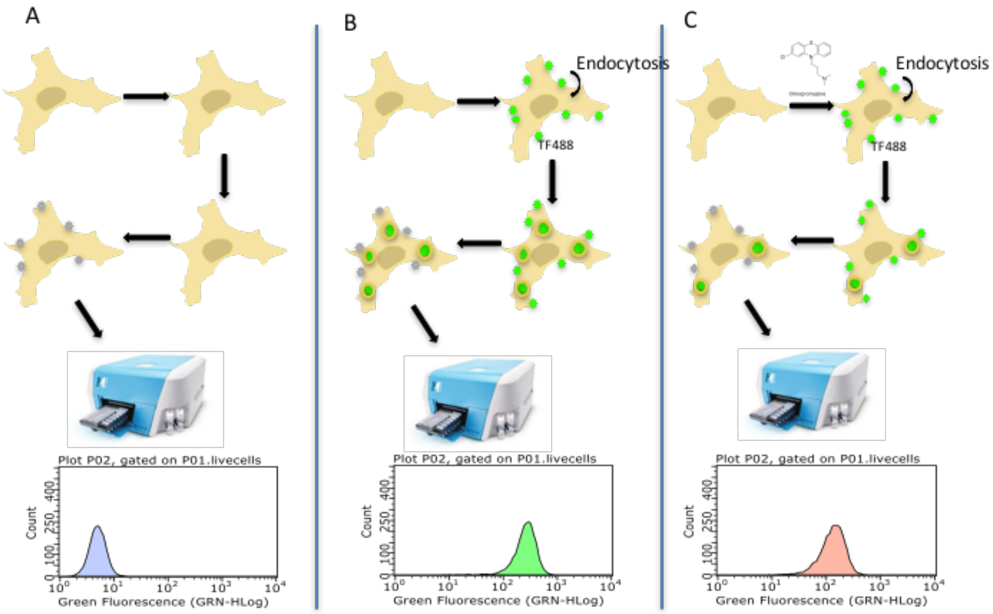


Figure 4.12 A.) Treatment of HeLa cells with DMSO alone allows for measurement of background fluorescence. B.) Uptake of Alexa-488 Transferrin measures full endocytosis via fluorescence. C.) Treatment of HeLa cells with chlorpromazine inhibits endocytosis and results in lower fluorescence of cells by flow cytometry.

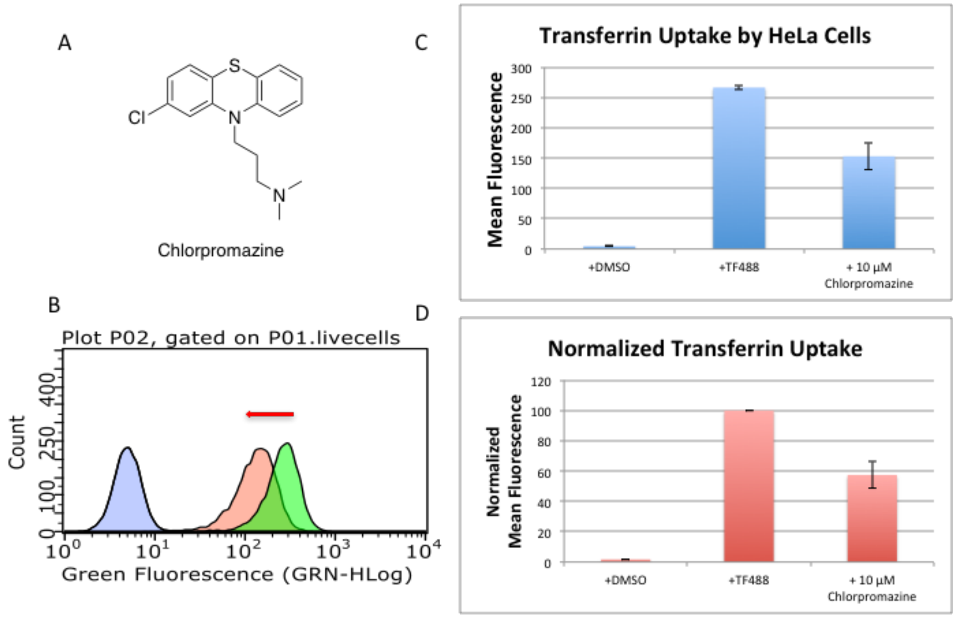


Figure 4.13 A.) Chlorpromazine is an established endocytosis inhibitor. B.) Overlay of histograms for DMSO control (blue), full endocytosis (green) and inhibition of endocytosis by 10 μM chlorpromazine (red). C.) Results quantified by mean fluorescence of cells for three independent trials. D.) Fluorescence can also be normalized to fluorescence observed following TF488 uptake.

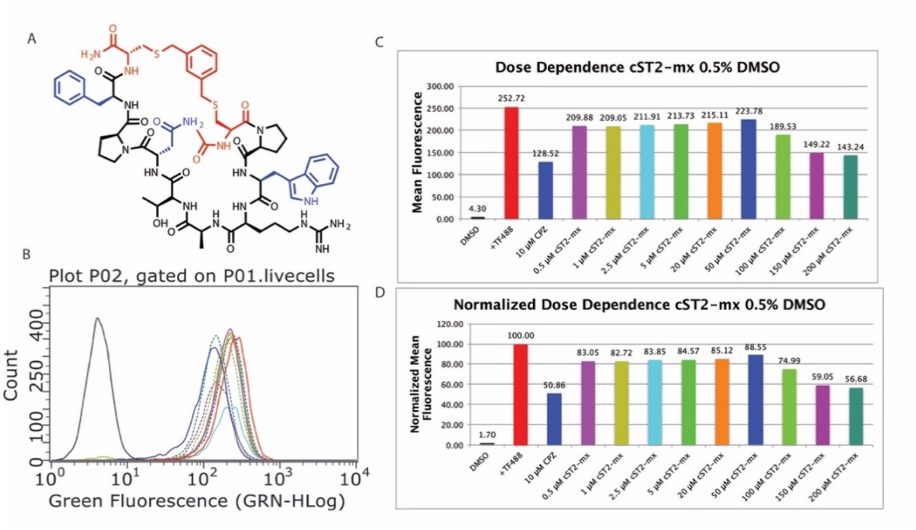


Figure 4.14 A.) Structure of **cST2-*mx***. B.) Histogram of HeLa cells in transferrin uptake experiment when HeLa cells are treated with the indicated concentrations of **cST2-*mx***. C.) Measurement of mean fluorescence of cells with various concentrations of peptide, colors correspond to the histogram plot in part B. D.) Normalized fluorescence compared to observed value following TF488 treatment alone.

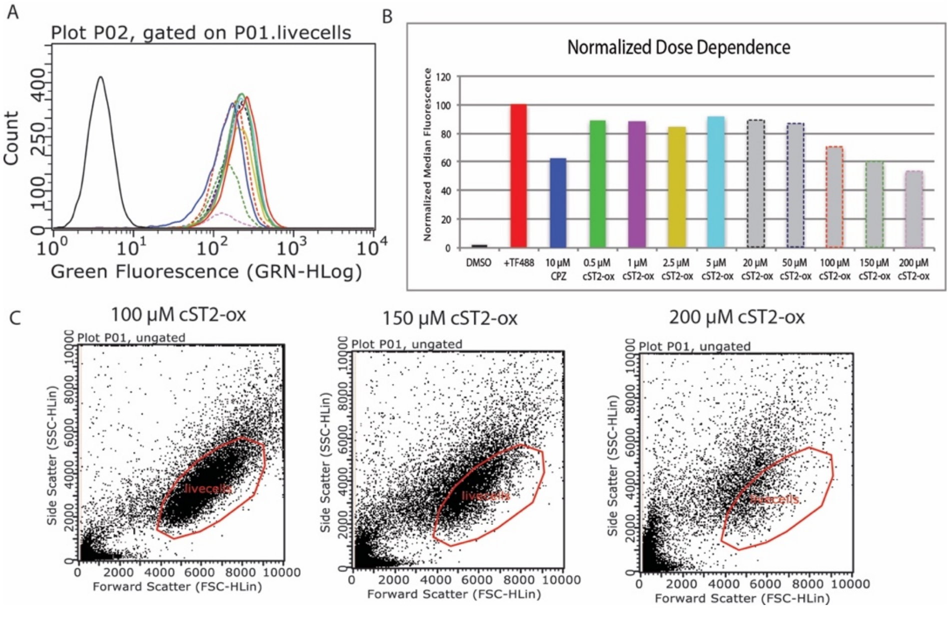


Figure 4.15 A.) Histogram of HeLa cells when treated with **cST2-*ox*** at various concentrations. B.) Corresponding normalized dose dependence showing decrease in transferrin uptake as a measure of fluorescence. C.) Forward and side scatter plots of HeLa cells at the highest three concentrations of **cST2-*ox***.

4.6 Structural analysis of the cST2-ox peptide

To understand the ability of**cST2-*ox*** to bind Eps15-EH2 and to aid in structure-based design, NMR studies were carried out to elucidate the structure of the peptide in solution. 1D-H^1, COSY, TOCSY, ROESY and HSQC experiments were carried out on **cST2-*ox*** in 9:1 H_2O:D_2O, at 1 mM peptide concentration at 298 K and pH 4. Using these spectra, the proton peaks were assigned (Table 4.3) and the carbon peaks were also assigned (Table 4.4). The degree of overall structure of the peptide was estimated from the amide proton and C_α proton shift change from the expected values for fully unfolded peptides (Figure 4.16). Many of the peaks are highly shifted from random coil values, allowing us to hypothesize that the peptide is well-folded. In addition, the ROESY spectra show cross-peaks between neighboring amino acid amide protons (Ala6 to Thr7, Thr7 to Asn8, and Phe10 to Cys11), indicative of a turn-like structure. Cross-peaks observed between the Asn8 β-protons and the Phe10 amide proton and β-protons also suggest the formation of a turn structure. In addition, cross-peaks were also observed between the Asn8 side-chain amide protons and the Phe10 β-protons. We surmised this turn would be capable of pre-orienting the NPF motif for binding to Eps15-EH2. We also observed a medium-range cross peak between the Arg4 side-chain Hε proton and the Thr6 β-proton indicative of a larger folded structure.

We used these ROESY cross-peaks to calculate a solution structure. We generated a list of distance constraints in CCPNMR by measuring ROESY peak intensities

(Table 4.5). Simulated annealing and refinement in the CNSsolve program using distance constraints resulted in an ensemble of conformations in aqueous solution using. In this ensemble, the NPF motif is exposed on one face of the peptide, and while not a canonical β-turn, a pronounced proline-induced turn is occurring over these three residues (Figure 4.17). Overlaying the backbone atoms of hot loop (from P2 to F9) results in a backbone RMSD of 0.562 Å (Figure 4.17A). Superimposing the NPF motif itself results in an all atom RMSD of 0.371 Å (Figure 4.17B). While the structure of the peptide does not appear to adopt one singular structure via NMR, it is apparent that the NPF motif attains a stable structure exposed on a single face of the peptide. This would seem to confirm the experimental binding results for which any disruption of the N7 or F9 residues within the peptides abolishes all binding affinity, but mutation of the W3 residue does not completely prevent binding to Eps15-EH2. Therefore, the purpose of the W3 residue may be more of a structural component. Another possibility is that this peptide undergoes an induced conformational stabilization upon binding of the NPF motif to Eps15-EH2. Overall, this structure can further be used to help inform future peptide design through N-methylation, bi-cyclization or other conformational restriction techniques.

Table 4.3 ^{1}H proton chemical shifts for **cST2-ox**.

cST2-ox	*NH*	*HA*	*HB*	*HG*	*HD*	*HE*	*HZ*	*HE3*	*HZ3*	*HH2*	*HZ2*	*NHE1*	*HD1*
CYS1	8.29	4.69	2.83/2.83										
PRO2		4.38	2.09	1.84/1.85	3.44/3.70								
TRP3	7.75	4.63	3.12/3.32					7.48	7.13	7.69	7.47	10.12	6.99
ARG4	7.67	4.13	1.66/1.75	1.51/1.51	3.10/3.13	7.13							
ALA5	8.29	4.00	1.41										
THR6	7.47	4.24	4.34	1.75									
ASN7	7.75	4.95	2.63/2.92		6.92/7.52								
PRO8		4.28	1.85/2.12	1.63	7.51								
PHE9	8.11	4.56	2.90/3.14		7.26	7.35	7.3						
CYS10	7.86	4.34	2.88/3.02										

Table 4.4 ^{13}C carbon chemical shifts for **cST2-ox** from HSQC experiment.

cST2-ox	CA	CB	CG	CD	CE	CZ	CE3	CZ3	CN2	CZ2	CD1
CYS1	66.53	32.31									
PRO2	60.87	24.16	24.24	48.10							
TRP3		26.61					118.04	119.37	121.97	111.92	124.30
ARG4	53.87	28.61	40.59								
ALA5	51.15										
THR6	58.84	52.87	19.02								
ASN7		36.63									
PRO8	61.35	28.96	29.18	47.94							
PHE9	55.29	35.87		128.96	128.72	127.22					
CYS10		32.86									

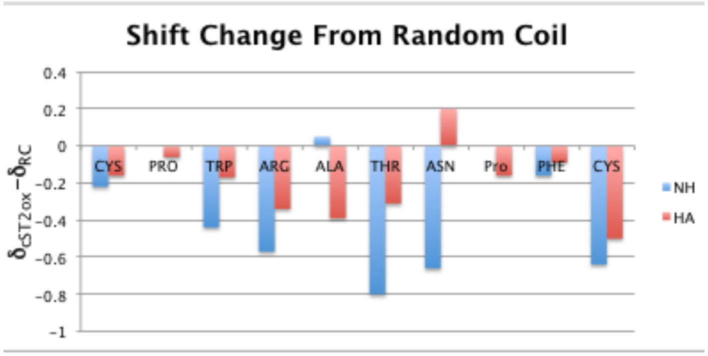


Figure 4.16 (*left*) ROESY cross peaks observed for **cST2-*ox***. (*right*) Shift changes from random coil can be calculated for **cST2-*ox*** with large shift changes observed that indicate a folded structure.

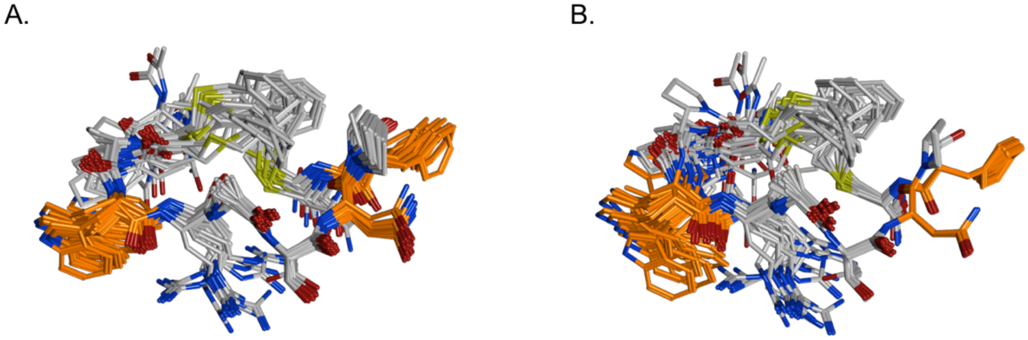


Figure 4.17 The top twenty structures from the refinement of **cST2-*ox*** are shown with the hot spot residues of W3, N7 and F9 highlighted in orange. A.) The backbone RMSD of the hot loop residues of P2 to F9 0.562 Å. B.) The total RMSD of the NPF motif is 0.371 Å.

Table 4.5 Distance Restraints for **cST2-ox**

Restraint	Proton 1			Proton 2			Value (Å)	Upper Limit	Lower Limit
1	2	CYS	HN	2	CYS	HA	3.4	4.1	2.7
2	2	CYS	HN	2	CYS	HB2	2.8	3.4	2.2
3	2	CYS	HN	2	CYS	HB1	2.8	3.4	2.2
4	3	PRO	HD2	2	CYS	HA	2.9	3.5	2.3
5	3	PRO	HD1	2	CYS	HA	3.5	4.2	2.8
6	3	PRO	HA	3	PRO	HB1	2.6	3.1	2.1
7	3	PRO	HA	3	PRO	HB2	3.6	4.3	2.9
8	3	PRO	HA	3	PRO	HD2	3.8	4.6	3
9	3	PRO	HA	3	PRO	HD1	4.4	5.3	3.5
10	3	PRO	HA	3	PRO	HG1	3.7	4.4	3
11	3	PRO	HA	3	PRO	HG2	3.7	4.4	3
12	4	TRP	HN	4	TRP	HB1	4.5	5.4	3.6
13	4	TRP	HE1	4	TRP	HD1	3.4	4.1	2.7
14	4	TRP	HE3	4	TRP	HB2	3.3	4	2.6
15	4	TRP	HE3	4	TRP	HB1	3.6	4.3	2.9
16	4	TRP	HD1	4	TRP	HB1	3.2	3.8	2.6
17	4	TRP	HD1	4	TRP	HB2	3.6	4.3	2.9
18	4	TRP	HE3	4	TRP	HA	3.3	4	2.6
19	4	TRP	HD1	4	TRP	HA	3.6	4.3	2.9
20	4	TRP	HE1	4	TRP	HZ2	3.7	4.4	3
21	4	TRP	HE3	5	ARG	HA	3.7	4.4	3
22	5	ARG	HN	4	TRP	HA	3.1	3.7	2.5
23	5	ARG	HN	4	TRP	HB2	4.2	5	3.4
24	5	ARG	HN	5	ARG	HA	3.2	3.8	2.6
25	5	ARG	HN	5	ARG	HG2	3.8	4.6	3
26	5	ARG	HN	5	ARG	HG1	3.8	4.6	3
27	5	ARG	HN	5	ARG	HB2	3.5	4.2	2.8
28	5	ARG	HN	5	ARG	HB1	3.9	4.7	3.1
29	5	ARG	HA	5	ARG	HD2	3.3	4	2.6
30	5	ARG	HA	5	ARG	HG1	3.3	4	2.6
31	5	ARG	HA	5	ARG	HG2	3.3	4	2.6
32	5	ARG	HA	5	ARG	HB2	3.2	3.8	2.6
33	5	ARG	HA	5	ARG	HB1	3.1	3.7	2.5
34	5	ARG	HE	5	ARG	HD2	3.2	3.8	2.6
35	5	ARG	HE	5	ARG	HD1	3.3	4	2.6
36	5	ARG	HE	5	ARG	HG2	3.5	4.2	2.8
37	5	ARG	HE	5	ARG	HG1	3.5	4.2	2.8
38	5	ARG	HE	5	ARG	HB2	4.4	5.3	3.5
39	5	ARG	HE	5	ARG	HB1	4.1	4.9	3.3

40	5	ARG	HE	7	THR	HB	3.2	3.8	2.6
41	6	ALA	HN	5	ARG	HA	2.7	3.2	2.2
42	6	ALA	HN	5	ARG	HB1	3.9	4.7	3.1
43	6	ALA	HN	5	ARG	HG2	4.3	5.2	3.4
44	6	ALA	HN	5	ARG	HG1	4.3	5.2	3.4
45	6	ALA	HN	6	ALA	HA	3.4	4.1	2.7
46	6	ALA	HN	6	ALA	HB#	2.9	3.5	2.3
47	6	ALA	HA	6	ALA	HB#	2.9	3.5	2.3
48	6	ALA	HN	7	THR	HN	3.5	4.2	2.8
49	7	THR	HN	6	ALA	HA	3.5	4.2	2.8
50	7	THR	HN	6	ALA	HB#	3.6	4.3	2.9
51	7	THR	HN	7	THR	HA	3.4	4.1	2.7
52	7	THR	HN	7	THR	HB	3.5	4.2	2.8
53	7	THR	HN	7	THR	HG1	3.6	4.3	2.9
54	7	THR	HA	7	THR	HG1	2.7	3.2	2.2
55	7	THR	HB	7	THR	HG1	2.7	3.2	2.2
56	8	ASN	HN	7	THR	HA	3.3	4	2.6
57	8	ASN	HN	7	THR	HN	3.1	3.7	2.5
58	8	ASN	HN	8	ASN	HB2	3.4	4.1	2.7
59	8	ASN	HN	8	ASN	HB1	3.5	4.2	2.8
60	8	ASN	HN	8	ASN	HA	3.3	4	2.6
61	8	ASN	HB2	8	ASN	HB1	2.4	2.9	1.9
62	8	ASN	HD21	8	ASN	HB1	3.2	3.8	2.6
63	8	ASN	HD21	8	ASN	HB2	3.7	4.4	3
64	8	ASN	HA	8	ASN	HB2	3.7	4.4	3
65	8	ASN	HN	9	PRO	HD1	4.3	5.2	3.4
66	8	ASN	HD22	10	PHE	HB2	3.9	4.7	3.1
67	8	ASN	HN	11	CYS	HA	3.7	4.4	3
68	9	PRO	HD2	8	ASN	HA	2.7	3.2	2.2
69	9	PRO	HA	9	PRO	HB1	2.6	3.1	2.1
70	9	PRO	HA	9	PRO	HB2	3.5	4.2	2.8
71	9	PRO	HA	9	PRO	HG2	3.1	3.7	2.5
72	9	PRO	HA	9	PRO	HG1	3.1	3.7	2.5
73	9	PRO	HA	9	PRO	HD1	3.8	4.6	3
74	9	PRO	HA	9	PRO	HD2	3.7	4.4	3
75	9	PRO	HD1	9	PRO	HG2	3.4	4.1	2.7
76	9	PRO	HD1	9	PRO	HG1	3.4	4.1	2.7
77	9	PRO	HD2	9	PRO	HG2	3.2	3.8	2.6
78	9	PRO	HD2	9	PRO	HG1	3.2	3.8	2.6
79	9	PRO	HD1	9	PRO	HB2	2.8	3.4	2.2
80	9	PRO	HD2	9	PRO	HB2	3.1	3.7	2.5

81	9	PRO	HD1	9	PRO	HB1	3.6	4.3	2.9
82	10	PHE	HN	9	PRO	HA	3.1	3.7	2.5
83	10	PHE	HN	10	PHE	HB2	3.1	3.7	2.5
84	10	PHE	HN	10	PHE	HB1	3.8	4.6	3
85	10	PHE	HN	8	ASN	HB1	4.5	5.4	3.6
86	10	PHE	HB1	8	ASN	HB1	2.9	3.5	2.3
87	10	PHE	HE#	10	PHE	HA	3.7	4.4	3
88	10	PHE	HD#	10	PHE	HA	2.8	3.4	2.2
89	10	PHE	HE#	10	PHE	HB2	3.5	4.2	2.8
90	10	PHE	HE#	10	PHE	HB1	3.8	4.6	3
91	10	PHE	HD#	10	PHE	HB2	2.9	3.5	2.3
92	10	PHE	HD#	10	PHE	HB1	3	3.6	2.4
93	10	PHE	HN	10	PHE	HA	3.2	3.8	2.6
94	10	PHE	HA	10	PHE	HB1	3.3	4	2.6
95	10	PHE	HN	11	CYS	HN	3.3	4	2.6
96	11	CYS	HN	10	PHE	HA	3.1	3.7	2.5
97	11	CYS	HN	10	PHE	HB1	4.4	5.3	3.5
98	11	CYS	HN	11	CYS	HA	3.2	3.8	2.6
99	11	CYS	HN	11	CYS	HB2	3.1	3.7	2.5
100	11	CYS	HN	11	CYS	HB1	3.8	4.6	3
101	11	CYS	HA	11	CYS	HB2	2.9	3.5	2.3
102	11	CYS	HA	11	CYS	HB1	3.1	3.7	2.5

4.7 Summary of Results and Future Directions

This thesis illustrated a novel approach for the identification of loop-mediated PPIs via LoopFinder. It identified a hot loop mediating the interaction between stonin2 and Eps15, and used it to design cyclic peptide inhibitors of the interaction. As expected, the linear hot loop epitope was unable to bind the target with high affinity. Our application of cysteine bis-alkylation for peptide cyclization allowed us to screen a panel of aryl linkers with varied substitution patterns, producing what we called a “locked loop library”. This small library of peptides maintained a constant hot loop epitope, but varied the linker identity in order to stabilize a range of different structural conformations. We found this to be a robust approach for screening panels of peptides and identified **cST2-*ox*** and **cST2-*mx*** as sub-micromolar binders. We hypothesize that these peptides attain the best conformation for binding. Straightforward amino acid substitution introduced unnatural amino acids in the place of tryptophan, phenylalanine or the cysteine residues. Here, added size of 1-naphthylalanine substituted for tryptophan resulted in a slight improvement (**cST4-*ox***). No improvement was observed for **cST2-*ox*** or **cST2-*mx*** when one or more penicillamines were substituted for cysteines. An inherent advantage of cyclic peptide design is the large diversity of amino acids that can be used for optimization. Further improvement of the peptide can be undertaken by full alanine scanning to identify whether any other amino acids besides the tryptophan, asparagine and phenylalanine, are responsible for the binding interaction. By further optimizing the amino acid sequence of the peptide, it may be possible to get closer to the naturally observed 0.15 μM affinity attained by the larger 39-mer peptide.[163]

NMR structural determination of the peptide can also help identify new avenues for peptide optimization. Analysis of the **cST2-*ox*** NMR structure revealed some interesting features of the peptide including the positioning of the alanine side chain pointing away from the binding face of the peptide. This may be an interesting position for substitution with D-homocysteine or D-cysteine to produce a bicyclic peptide via tris-alkylation with 1,3,5-tris(bromomethyl)-benzene. In addition, it could be seen from the ROESY data that the amide protons Ala6 toThr7, Thr7 to Asn8, and Phe10 to Cys11 are all in close proximity and have a strong interaction. It may be possible to introduce N-methylation at the other amide protons that may not be as responsible for attaining a proper conformation. Both bicyclization and N-methylation will further rigidify the peptide structure, an effect that may extend to an improvement in cell penetration.

A major limitation of using peptides as inhibitors of intracellular PPIs is the limited understanding of how to make peptides cell-penetrant. Even the pathways by which TAT and other well-established cell-penetrating peptides are able to enter the cells is poorly understood.[197] Methods for improving cell penetration of peptides are severely underdeveloped. As observed, poly-arginine TAT-peptide conjugation to **cST2-*ox*** resulted in widespread cell death at high concentrations, and at lower concentrations no inhibition of endocytosis was observed. Through further design strategies, peptide conformational space may be explored that allow for the penetration of the peptide allowing for biological inhibition of Eps15 activity.

While so far unsuccessful in identifying biologically active peptide inhibitors of Eps15, this method has allowed for some important steps in the targeting of endocytosis through Eps15 inhibition. We were able to use the LoopFinder method coupled with a small locked loop library to identify the cyclic peptide **cST2-*ox***, which was capable of binding Eps15-EH2 almost as tightly as the native interaction. **cST2-*ox*** represents the first sub-micromolar short peptide binder of Eps15-EH2. It also possesses inherent specificity over other EH domains. This peptide was successfully developed into a fluorescence polarization competition probe for further biochemical assays. This assay was used to screen peptide binders in a medium-throughput fashion, and it should only require minimal optimization to extend to a high-throughput assay. This type of assay can even be used to identify small molecules capable of displacing the FP probe. This work is allowing us to get closer to the discovery of new chemical tools that can be used to specifically halt clathrin-mediated endocytosis in a pathway specific manner and fill an important need in the field. Inhibitors such as these will increase the biological understanding of the balance among all types of endocytosis.

More generally, the development of a hot loop from LoopFinder into a cyclic peptide capable of sub-micromolar binding helps validate the approach the Kritzer lab is taking to discover new PPI inhibitors. As discussed in Chapter 3 there are a wide variety of hot loops identified by LoopFinder that are completely novel targets for inhibition. Cyclic peptides would make useful chemical probes to aid in the understanding of biological pathways. A list of all hot loops identified by LoopFinder that fulfill all three hot loop parameters are provided in the appendix. Other interesting and biomedically relevant targets are present in the dataset that can be starting points for cyclic peptide inhibitor design. These even include a wide range of other PPIs involved in vesicle trafficking. It was the goal of this work to develop a platform for hot loop discovery and cyclic peptide inhibitor identification that can be applied generally across a range of protein targets. We expect this strategy will develop many more functional tools for understanding biological pathways. If issues with cell penetration can be worked out, these molecules may even become lead molecules for drug development.

4.8 Materials and Methods

4.8.1 Cyclic Peptide Synthesis.

Reagents used were as follows. Fmoc-protected amino acids were purchased from either Anaspec, Chem Pep, or Novabiochem. The 1-[Bis(dimethylamino)methylene-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and 1-hydroxy-7-azabenzotriazole (HOAt) were purchased from Chem Pep. Rink amide linker was purchased from Novabiochem. The dimethylformamide (DMF), *N,N*-diisopropylethylamine (DIEA), acetic anhydride, trifluoroacetic acid (TFA), dichloromethane (DCM), triisopropylsilane (TIS), 2,6-Lutidine, ethanedithiol (EDT) and piperidine were purchased from Sigma Aldrich. The 5/6-carboxyfluorescein succinimidyl ester (NHS-Fluorescein) was purchased from Thermo Fisher Scientific.

The linear precursor peptides were synthesized using standard Fmoc peptide synthesis protocol on Rink amide resin (substitution = 0.44 mmol/g) either in batch format or on the Aapptec Apex 396 peptide synthesizer or by hand. Couplings were carried out using 5 equivalents (equiv.) of Fmoc protected amino acid, 4.8 equiv HATU, 4.8 equiv of HOAt, 13 equiv of DIEA. The coupling reactions were allowed to proceed for 45 minutes in DMF, after which Fmoc deprotection was carried out upon treatment of the peptide-bound resin with 20% (vol/vol) piperidine/DMF (for 5 minutes followed by another 5 minutes treatment). Upon completion of the peptide sequence assembly on resin and deprotection of the final Fmoc group, the N-terminal amine was acetylated by treatment with 5% (vol) acetic anhydride and 6% (vol) 2,6-lutidine in DMF. In the case of the fluoresceinated peptides, 2 equiv NHS-Fluorescein and 4 equiv of DIPEA were added to the resin in 3 mL DMF. The reaction was allowed to shake on an orbital shaker overnight, then rinsed thoroughly with DMF and DCM.

Cleavage and deprotection from the solid support was carried out using TFA/H_2O/TIS/EDT (94/2/2/2 vol/vol) for 3 hours at room temperature. The resin was then filtered and washed with excess TFA. The isolated TFA solution was then evaporated, and the peptide was precipitated upon addition of cold diethyl ether. The peptide suspension in ether was then centrifuged to pellet the peptide followed by decanting away the ether. The pellet was then washed two additional times with cold diethyl ether to remove as much TFA as possible. For the non-fluoresceinated peptides the isolated crude linear precursor was then directly dissolved in 50 mL 1:1 (vol/vol) acetonitrile/25 mM sodium bicarbonate at pH 8.0. The solution was then aliquoted into five new fractions (10 mL each) which were further diluted to 40 mL with 1:1 (vol/vol) acetonitrile/25 mM sodium bicarbonate at pH 8.0. Bis-alkylation of the cysteine residues was carried out upon treatment with 1.5 equiv of the dibromo-containing linkers (α, α'-dibromo-*o*-xylene, α, α'-dibromo-*m*-xylene, α, α'-dibromo-*p*-xylene, or 2,6-bis(bromomethyl)pyridine) or 3 equiv of α -bromotoluene and allowed to be treated for 2 hours. Reaction completion was monitored by matrix assisted laser desorption/ionization (MALDI, Bruker). The reactions were then frozen in liquid nitrogen and lyophilized to remove the acetonitrile/buffer solvent. Purification of the resulting cyclic peptides was achieved by high performance liquid chromatography (HPLC, Varian) on a reversed-phase C_8 column (Agilent; Column Type) to yield pure peptide. Purity was assessed using MALDI in conjunction with analytical scale HPLC on a C_{18} column (Vydac, Protein and Peptide C_{18}). In the case of the fluorescently labeled peptide, HPLC purification was carried out on the linear precursor to yield the purified linear precursor peptide which was then subjected to treatment for 2 hrs with 1.5 eqiv of α, α'-dibromo-*o*-xylene. Post cyclization reaction, the fluorescently labeled cyclic peptide was purified by HPLC under the same conditions as the previous peptides described above. The analytical HPLC as well as mass spectral data for the peptides tested in this study were all confirmed to be pure and properly characterized (Table 4.6).

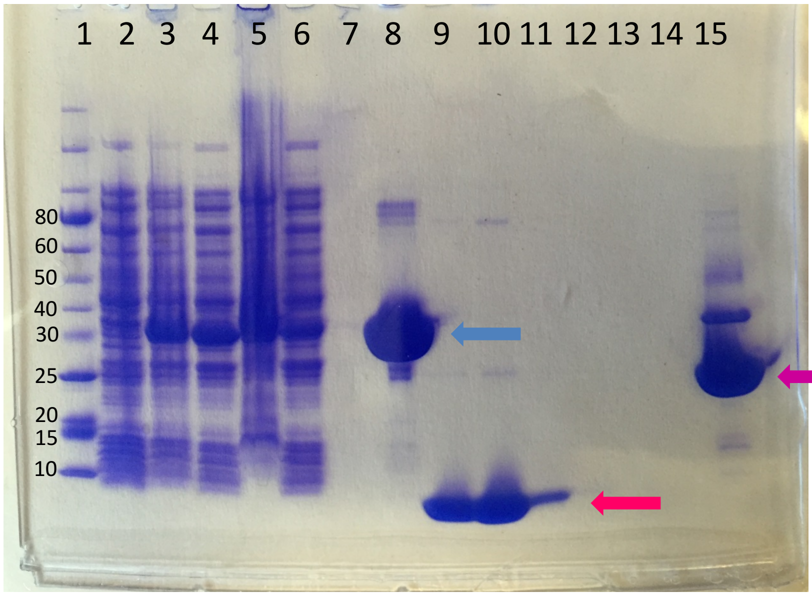
Concentrations were calculated by UV/Vis for all peptides containing a tryptophan residue and ***ox***, ***mx***, ***px***, and ***dbz*** linkers using an extinction coefficient of 5,500 M^{-1} cm^{-1}. For peptides containing a tryptophan and ***pyr*** linker an adjusted extinction coefficient of 7,440 $M^{-1}cm^{-1}$ was used. This extinction coefficient arises from the addition of the pyridine extinction coefficient estimated via a known concentration of 2,6-lutidine in MOPS buffer at pH 7.0. For peptides lacking a tryptophan residue, concentration was measured by mass. The concentration of FP probe was calculated by UV/Vis using an extinction coefficient of 70,000 $M^{-1}cm^{-1}$.

Table 4.6 Peptide characterization data.

Peptide	**Retention Time (min)**	**Expected Mass (m/z)**	**Observed Mass (m/z)**
lST2	14.5	1229.38	1229.67
cST2-*ox*	17.4	1336.58	1335.97
cST2-*mx*	17.4	1336.58	1336.11
cST2-*px*	17.4	1336.58	1336.25
cST2-*pyr*	19.1	1337.58	1337.26
cST2-*dbz*	15.0	1414.63	1414.54
cST2-W3A-*ox*	14.5	1221.54	1221.38
cST2-N7A-*ox*	16.8	1293.58	1293.29
cST2-F9A-*ox*	13.8	1260.55	1260.5
pent-cST2-*ox*	18.4	1374.60	1374.85
pent-cST2b-*ox*	18.5	1331.59	1331.75
cST4-*ox*	16.7	1347.59	1347.88
cST4-*mx*	16.7	1347.59	1347.98
cST5-*ox*	17.0	1386.60	1387.05
cST5-*mx*	17.0	1386.60	1386.93
cST6-*ox*	17.9	1397.60	1398.35
cST6-*mx*	17.9	1397.60	1398.07
cST7-*ox*	13.6	1392.65	1393.28
cST7-*mx*	13.6	1392.65	1393.11
cST8-*ox*	13.5	1364.62	1365.09
cST8-*mx*	13.5	1364.62	1365.4
cST9-*ox*	13.4	1364.62	1364.66
cST9-*mx*	13.2	1364.62	1364.71
cST1-*ox*	16.2	1239.53	1240.01
cST1-*mx*	16.2	1239.53	1239.85
cST1-*pyr*	14.0	1240.53	1240.79
cST1-*px*	16.1	1239.53	1240.05
Flu-cST2-*ox*	32.8	1794.70	1795.57

4.8.2 Cloning, Expression, and Purification of Eps15-EH2.

The Eps15-EH2 domain was expressed was a GST-fusion protein from pGEX-GST-Eps15-EH2 plasmid (Figure 4.18). The plasmid was transformed into BL-21 (DE3) *E. coli*. 1L of LB broth media was inoculated with a single colony of BL-21 (DE3) containing the above plasmid. The cells were grown at 37 ˚C and induced with 1 mM isopropyl β -D-1-thiogalactopyranoside at an OD_{600} of 0.8. Cells were then allowed to grow for 20 hours at room temperature. Protein was purified by treatment of the lysate with glutathione agarose (Gold Biotechnology) to bind the fusion protein. The resin was washed with thrombin digest buffer (50 mM Tris, 150 mM NaCl, 2.5 mM $CaCl_2$ at pH 8.0) and 100 units of Thrombin (Sigma Aldrich) was added and allowed to react for 2 hours to cleave the Esp15-EH2 purified protein from the resin bound GST. The eluted protein fractions were run through a benzamidine agarose column (Sigma Aldrich) to remove contaminant thrombin from the elution buffer. Purity was analyzed to be >95% by SDS-Page electrophoresis. Concentrations were measured by absorbance at 280 nm, using an extinction coefficient of 13980 M^{-1} cm^{-1}. An identical method was used to purify EHD1 via expression from a pGEX-GST-EHD1 plasmid and Reps1-EH via expression from a pGEX-GST-Resp1-EH plasmid in bacteria (Figure 4.19).

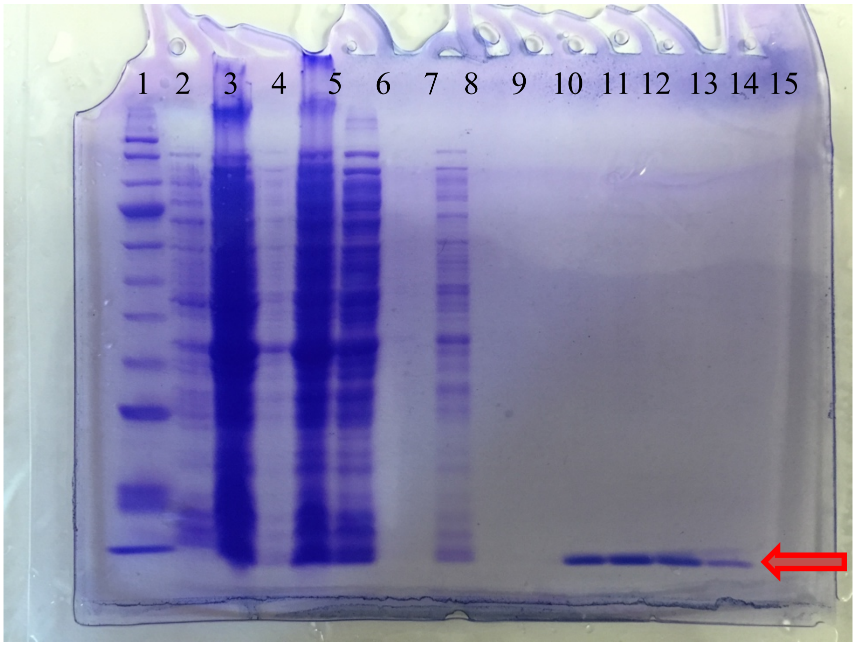


GST-EPS15-EH$_2$ Fusion : 37 kDa

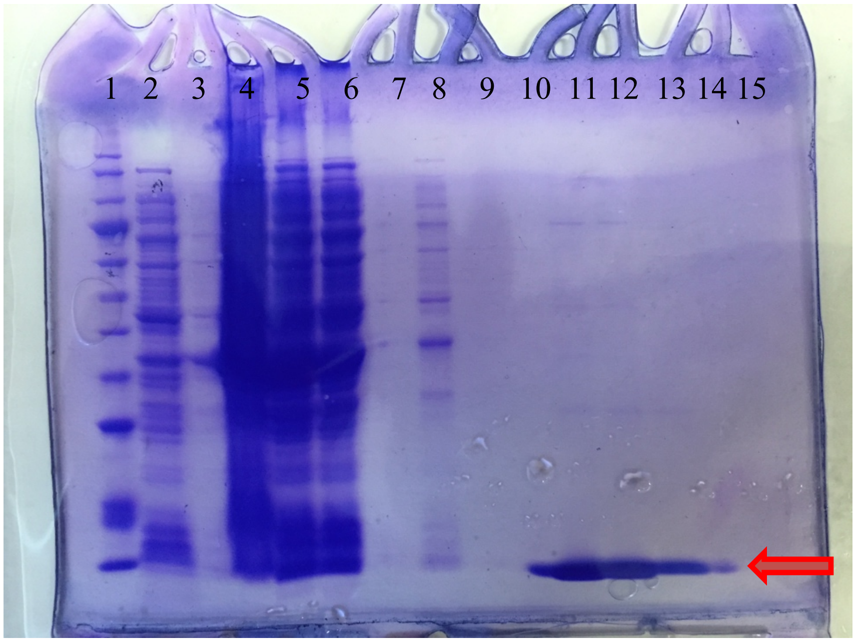
GST : 27 kDA

Eps15-EH$_2$: 10 kDA

Figure 4.18 Purification of Eps15-EH2 via glutathione agarose affinity chromatography. Full length GST-Eps15-EH2 is shown (blue arrow) before treatment with thrombin. Elution after thrombin cleavage results in pure Eps15-EH2 (red arrow). The beads after elution are analyzed and shown to retain the GST (purple arrow).



EHD1 EH Domain : 10 kDA



REPS1 EH Domain : 10 kDA

Figure 4.19 EHD1 (top) and Reps1 (bottom) EH domains were purified by glutathione agarose affinity chromatography.

4.8.3 Isothermal Titration Calorimetry Experiments.

To assess binding of the peptides to the Eps15-EH2 protein binding partner isothermal titration calorimetry (ITC) was carried out using the MicroITC$_{200}$ (GE). The peptide and protein samples were dialyzed into the same buffer (25 mM MOPS, 150 mM NaCl, 1 mM $CaCl_2$ at pH 7.0) for four hours, followed up by replacement of fresh buffer for an overnight dialysis, all at 4˚C. The cell temperature was set to 25 ˚C. The sample cell was loaded with 0.04 mM Eps15-EH2 and the syringe was loaded with 0.4 mM peptide stocks. The stirring speed of the injector was set to 750 rpms with a reference power of 10 mcal/mol and 20 injections. The first injection was 0.2 μL with duration of 0.4 sec, spacing of 180 sec and filter period of 5 sec. The subsequent 19 injections were all 2 μL injections for duration of 4 sec, with 180 sec spacing and filter period of 5 sec. The experiments were then analyzed using the MicroCal Analysis tool in order to calculate the association constant and dissociation constant for each peptide binding to Eps15-EH$_2$. ITC data for each peptide in the **cST2** is provided in the Appendix.

4.8.4 Fluorescence Polarization Binding and Competition Assays

To more quickly test the binding of peptides and additional molecules to Eps15-EH2, a fluorescence polarization assay was developed. The probe designed was a fluorescent variant of **cST2-*o*x** in which the N-terminus of the peptide had a linker followed by the fluorescein fluorophore (Flu- β Ala- β Ala-Cys-Pro-Trp-

Arg-Ala-Thr-Asn-Pro-Phe-Cys-NH_2 with the *o*-xylene linker between the two cysteine residues). Concentration of fluorescently labeled peptide was determined by UV/Vis with an extinction coefficient of 77,700 M^{-1} cm^{-1}. To measure direct binding of the probe to Eps15-EH2, a serial dilution of protein was added to 40 nM probe. Fluorescence polarization was calculated after half-hour incubation period at room temperature using an Infinite F200 microplate reader (Tecan) with an excitation wavelength of 485 nm and an emission wavelength of 535 nm. The anisotropy data was plotted using KaleidaGraph and the curve was fit to the following equation:

Equation 4.1 Equation to calculate the K_d of probe by FP.

$$r = [\left(\frac{p + [c] + K_d \pm \sqrt{(-p - [c] - K_d)^2 - 4(p \times [c])}}{2p}\right) \times (r_\infty - r_0))] + r_0$$

here, the p = probe concentration, $[c]$ = protein concentration, r_∞= anisotropy of bound state, r_0 = anisotropy of free state and r = anisotropy. The K_d value for the probe was measured using the above curve fit and calculated to be 0.50 ± 0.4 μM. This probe was used to develop a competition assay used to screen for Eps15-EH2 binding ligands. The competition assay was run using 40 nM of probe and 1.56 μM Eps15-EH2. Serial dilutions of peptides to be tested were tested and the fluorescence anisotropy was measured using the same methods as previously discussed. In the case of the competition assay, the fraction bound at each concentration of peptide was calculated using the following equation:

Equation 4.2 Equation to calculate the fraction of probe bound to the target protein by FP.

$$FB = \frac{(r - r_0)}{(r_\infty - r_0)}$$

where , r_∞= anisotropy of bound state, r_0 = anisotropy of free state and r = anisotropy. The resulting fraction bound measurement was then plotted using KaleidaGraph and fit to the following equation:

Equation 4.3 Equation to calculate the IC_{50} of peptides displacing probe by FP.

$$FB = 0 + \frac{(1-0)}{(1 + (\frac{[p]}{IC_{50}})^y}$$

from which the IC_{50} for each peptide can be calculated. All fluorescence polarization assays were carried out in 25 mM MOPS, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM DTT at pH 7.0). FP competition assay data for three independent trials of each peptide are provided in the Appendix.

4.8.5 Transferrin Uptake Assays

HeLa cells were grown at 37ºC in 10% CO_2 and seeded at two densities in DMEM (Gibco) with 10% FBS: 7.5 x 10^4 cells per well to be ready the next day, and 3.75 x 10^4 cells per well to be ready in two days. HeLa cells were grown to ~80% confluence in a six well plate (Cell Treated Thermo Scientific). The media was removed via aspiration and 2 mL of the appropriate serum free DMEM was added to each well. For the background HeLa cell fluorescence control and negative control, only DMEM and DMSO was added. For the positive control, 2 uL of a 10 µM stock of chlorpromazine (Sigma Aldrich) was added to 2 mL of serum free DMEM. Various concentrations of peptide stocks of DMSO can be added to 2 mL of DMEM. Importantly, all DMEM media was standardized to contain the same amount of DMSO. The cells were then incubated for 1 hour, after which they were removed from the incubator and the media was aspirated away. The same DMEM containing DMSO, Control or peptide was prepared, and 1 µL of a 5 mg/mL stock of Alexa-488 labelled Transferrin (Life Technologies) was added to each sample of media. Subsequently the media was added to each corresponding well. The cells were then incubated for 30-minutes. After 30-minutes, the media was removed via aspiration and 2 mL of ice cold DMEM to stop endocytosis. The cold DMEM was removed immediately, and 1 mL of DMEM containing 2 mg/mL unlabeled human transferrin (Sigma Aldrich) was added to each well and allowed to incubate at room temperature for 2 minutes. The media was then removed and 1 mL of DMEM with Trypsin was added to each well to lift the cells from the plate during a 5-minute incubation time at 37ºC. The cells can be seen lifted from the plate under the microscope, and 0.5 mL of serum free DMEM was added to each well. The media was then transferred to a 1.5 mL Eppendorf tube and spun down for 3 minutes at 3,000 rpm to pellet the cells. The media was removed and the cells were washed with ice cold PBS, then pelleted again. After the PDB was removed 300 µL of 1:10 para-formaldehyde (Sigma Aldrich) in PBS was used to resuspended the cells. The cells were then allowed to incubate on ice for 10 minutes during this fixation step. After 10 minutes, the cells were spun down for 3 minutes at 3,000 rpm to pellet the cells. The cells were then resuspended in 175 µL of PBS and added to a round bottom 96-well plate (Falcon). The fluorescence of the cells were then measured by flow cytometry using the EMD Millipore Guava easyCyte.

4.8.6 Collection of NMR Spectra

All experiments were conducted at 25 ºC. 1D-^{1}H , 2D-COSY, TOCSY, NOESY, ROESY and ^{1}H-^{13}C HSQC experiments were conducted on either Bruker Avance 500 MHz spectrometer (Tufts University, Department of Chemistry, Medford, MA), and/or subsequently conducted on Bruker Avance 800 MHz spectrometer with a 5 mm TCI cryoprobe with z-axis pulsed magnetic field gradients (Brandeis University, NMR Facility, Waltham, MA) all of which are located in the Appendix. Water suppression was achieved using gradient suppression with selective irradiation at the o1 frequency of 2349 Hz (600 MHz spectrometer) or 3763 Hz (800 MHz spectrometer).

All spectra were processed using the Topsin software package (Bruker). The spectra were then imported into CCPNMR program in order to assign peaks within the spectra and generate ROE-derived distance constraints as previously discussed.[200] From the Brandeis ROESY spectra, 80 distance constraints were generated and exported into a CNSSolve compatible format.[201]

These 80 distance constraints were used in the simulated annealing step with CNSSolve. Simulated annealing was run using a high temperature (5000) annealing stage of 1,000 steps, with two slow cool annealing stages the first of which was 1,000 steps and second stage of 3,000 steps. A final minimization stage of 10 cycles with 200 steps was used. Trial and accepted structures were printed. This method resulted in 100 lowest-energy structures that could be visualized in PyMol. Violations to the distance constraints were calculated using a python script developed by Hongtao Yu (Lin Lab). Adjustments to the upper and lower bounds of all violations > 0.3 Å were made, and the annealing round was run again to produce the 100 lowest-energy structures. No gross violations were observed, and the 100 lowest-energy annealed and accepted structures were used as input for the refinement stage. High temperature annealing stage was done at 1,000 with 1,000 steps followed by two slow-cool annealing stages the first with 2,000 steps the second with 6,000 steps. For final minimization 10 cycles with 200 steps was used. 50 accepted structures were printed and analyzed via VMD to result in a RMSD of 0.408 Å.

5 Appendix

Table 5.1 Hot Loops that satisfy all three parameters.

PDBID	Structure Title	Chain	Partner	Residues	First	Last	Termini Distance	Hot Loop Sequence	ΔΔG1	ΔΔG2	ΔΔG3	ΔΔG4	ΔΔG5	ΔΔG6	ΔΔG7	ΔΔG8
1axi	STRUCTURAL PLASTICITY AT THE HGH:HGHBP INTERFACE	B	A	7	D164	M170	5.53	DIQKGWM	1.00	4.50	0.08	0.29	0.00	2.41	0.00	
1b9l	7,8-DIHYDRONEOPTERIN TRIPHOSPHATE EPIMERASE	B	A	8	H104	S111	4.79	HALRYADS	0.00	0.00	0.67	0.00	4.50	0.00	1.59	1.00
1bii	THE CRYSTAL STRUCTURE OF H-2DD MHC CLASS I IN COMPLEX WITH2 THE HIV-1 DERIVED PEPTIDE P18-110	B	A	8	S355	F362	5.83	SFSKDWSF	0.62	1.38	0.00	0.00	0.00	2.91	0.00	0.00
1bsu	STRUCTURAL AND ENERGETIC ORIGINS OF INDIRECT READOUT IN2 SITE-SPECIFIC DNA CLEAVAGE BY A RESTRICTION ENDONUCLEASE	B	A	8	I23	I30	5.55	IISAEGKI	1.46	1.29	0.00	0.00	0.00	0.00	0.17	4.07
1bxk	DTDP-GLUCOSE 4,6-DEHYDRATASE FROM E. COLI	B	A	7	P556	A562	5.03	PSSPYSA	2.40	1.00	1.00	0.74	0.00	0.00	1.00	
1cs1	CYSTATHIONINE GAMMA-SYNTHASE (CGS) FROM ESCHERICHIA COLI	B	A	7	N15	G21	6.15	NDDEQYG	0.82	0.00	0.00	0.45	1.06	2.50	0.00	
1dlp	STRUCTURAL CHARACTERIZATION OF THE NATIVE FETUIN-BINDING2 PROTEIN SCILLA CAMPANULATA AGGLUTININ (SCAFET): A NOVEL3 TWO-DOMAIN LECTIN	F	C	5	Q188	R192	4.7	QPNGR	1.49	1.00	1.00	0.00	0.00			
1dum	NMR STRUCTURE OF [F5Y, F16W] MAGAININ 2 BOUND TO2 PHOSPHOLIPID VESICLES	A	B	6	K4	A9	5.71	KYLHSA	2.05	1.00	1.00	0.00	0.00	0.56		
1egp	PROTEINASE INHIBITOR EGLIN C WITH HYDROLYSED REACTIVE CENTER	B	A	8	Y56	V63	4.1	YNPGTNVV	2.25	0.00	0.57	0.00	0.00	1.00	0.56	0.73
1etr	REFINED 2.3 ANGSTROMS X-RAY CRYSTAL STRUCTURE OF BOVINE THROMBIN2 COMPLEXES FORMED WITH THE BENZAMIDINE AND ARGININE-BASED THROMBIN3 INHIBITORS NAPAP, 4-	L	H	6	L3	E8	5.58	LRPLFE	0.88	1.00	4.50	0.00	0.00	2.04		
1fth	CRYSTAL STRUCTURE OF STREPTOCOCCUS PNEUMONIAE ACYL CARRIER2 PROTEIN SYNTHASE (3'5'-ADP COMPLEX)	A	C	8	S1104	T1111	5.49	SHTDQFVT	0.29	0.00	0.70	0.00	0.00	4.50	0.00	4.50
1fxo	THE STRUCTURAL BASIS OF THE CATALYTIC MECHANISM AND2 REGULATION OF GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE3 (RMLA). TMP COMPLEX.	C	D	8	L27	P34	6.03	LLPVYDKP	4.50	0.00	0.60	0.00	1.49	0.66	0.32	4.50
1hao	COMPLEX OF HUMAN ALPHA-THROMBIN WITH A 15MER OLIGONUCLEOTIDE2 GGTTGGTGTGGTTGG (BASED ON NMR MODEL OF DNA)	L	H	6	L3	E8	5.62	LRPLFE	0.33	1.00	4.50	0.45	0.00	1.69		
1hap	COMPLEX OF HUMAN ALPHA-THROMBIN WITH A 15MER OLIGONUCLEOTIDE2 GGTTGGTGTGGTTGG (BASED ON X-RAY MODEL OF DNA)	L	H	6	L3	E8	5.51	LRPLFE	0.12	1.00	0.66	0.64	0.00	2.13		
1hwg	1:2 COMPLEX OF HUMAN GROWTH HORMONE WITH ITS SOLUBLE2 BINDING PROTEIN	B	C	7	V144	H150	6	VSLTGIH	0.00	1.21	0.10	1.31	0.00	0.31	1.57	
1hx6	P3, THE MAJOR COAT PROTEIN OF THE LIPID-CONTAINING2 BACTERIOPHAGE PRD1.	A	B	7	R329	D335	5.86	RRRIATD	1.00	0.00	0.00	0.00	0.00	4.50	4.50	
1im9	CRYSTAL STRUCTURE OF THE HUMAN NATURAL KILLER CELL2 INHIBITORY RECEPTOR KIR2DL1 BOUND TO ITS MHC LIGAND HLA-CW4	A	B	8	R234	F241	5.62	RPAGDGTF	1.00	4.50	0.00	0.00	0.77	0.00	0.00	0.00
1im9	CRYSTAL STRUCTURE OF THE HUMAN NATURAL KILLER CELL2 INHIBITORY RECEPTOR KIR2DL1 BOUND TO ITS MHC LIGAND HLA-CW4	F	E	8	S55	F62	5.75	SFSKDWSF	1.06	1.20	0.00	0.00	0.00	3.10	0.00	0.00
1iq6	(R)-HYDRATASE FROM A. CAVIAE INVOLVED IN PHA BIOSYNTHESIS	A	B	6	A26	D31	4.82	AALSED	0.00	0.00	1.27	0.26	1.29	0.89		
1jmt	X-RAY STRUCTURE OF A CORE U2AF65/U2AF35 HETERODIMER	A	B	8	R133	I140	6	RWFNGQPI	2.76	2.85	2.16	0.18	0.00	0.00	0.00	0.00
1jnp	CRYSTAL STRUCTURE OF MURINE TCL1 AT 2.5 RESOLUTION	B	A	6	L276	R281	5.53	LYPKDR	0.02	1.00	1.00	2.78	0.00	0.00		
1jr0	CHOLERA TOXIN B-PENTAMER WITH LIGAND BMSC-0011	E	D	6	L31	E36	4.83	LAGKRE	1.00	0.00	0.00	0.00	1.40	1.83		
1ju5	TERNARY COMPLEX OF AN CRK SH2 DOMAIN, CRK-DERIVED2 PHOPHOPEPTIDE, AND ABL SH3 DOMAIN BY NMR SPECTROSCOPY	C	A	8	Y93	C100	4.73	YNHNGEWC	0.00	1.33	1.15	0.62	0.00	1.00	1.00	0.00
1k1d	CRYSTAL STRUCTURE OF D-HYDANTOINASE	A	D	7	E276	G282	5.47	EKPNFEG	0.00	1.00	0.83	4.36	1.00	0.16	0.00	
1kgy	CRYSTAL STRUCTURE OF THE EPHB2-EPHRINB2 COMPLEX	F	B	8	S1321	E1328	6.01	SPNLWGLE	0.00	0.00	0.73	1.38	3.10	0.00	0.00	1.04
1kke	CRYSTAL STRUCTURE OF REOVIRUS ATTACHMENT PROTEIN SIGMA12 TRIMER	A	B	8	I281	T288	5.95	INSSGQLT	0.58	0.89	0.00	0.00	0.00	1.16	1.40	0.83
1krr	GALACTOSIDE ACETYLTRANSFERASE IN COMPLEX WITH ACETYL-2 COENZYME A	C	B	7	P116	R122	6.07	PVHHELR	3.02	0.65	1.39	0.43	0.41	0.70	0.00	
1ku6	FASCICULIN 2-MOUSE ACETYLCHOLINESTERASE COMPLEX	B	A	8	S26	M33	5.37	SRRHPPKM	0.00	0.86	0.00	0.65	0.51	2.61	0.15	0.91
1le8	CRYSTAL STRUCTURE OF THE MATA1/MATALPHA2-3A HETERODIMER2 BOUND TO DNA COMPLEX	A	B	6	L95	K100	5.69	LNSKEK	0.00	0.64	0.10	1.45	1.68	0.00		
1mas	PURINE NUCLEOSIDE HYDROLASE	B	A	8	E265	L272	5.27	ELTGKLTL	2.77	0.00	0.00	0.00	0.00	0.79	2.44	0.00
1ncp	STRUCTURAL CHARACTERIZATION OF A 39 RESIDUE SYNTHETIC2 PEPTIDE CONTAINING THE TWO ZINC BINDING DOMAINS FROM THE3 HIV-1 P7 NUCLEOCAPSID PROTEIN BY CD AND NMR	N	C	8	K2	E9	6.1	KCFNCGKE	4.50	0.10	1.00	1.00	0.00	0.00	1.00	1.00
1nfk	STRUCTURE OF THE NUCLEAR FACTOR KAPPA-B (NF-KB) P502 HOMODIMER	B	A	6	V303	A308	5.75	VHRQFA	0.00	1.20	0.59	0.00	1.00	1.15		
1nvm	CRYSTAL STRUCTURE OF A BIFUNCTIONAL ALDOLASE-DEHYDROGENASE :2 SEQUESTERING A REACTIVE AND VOLATILE INTERMEDIATE	A	C	6	V278	L283	5.75	VDRETL	0.84	0.40	2.53	0.00	0.00	4.50		
1o7d	THE STRUCTURE OF THE BOVINE LYSOSOMAL A-MANNOSIDASE2 SUGGESTS A NOVEL MECHANISM FOR LOW PH ACTIVATION	C	E	7	V570	S576	5.61	VPAVGFS	0.00	4.50	0.01	0.77	0.00	1.00	0.62	
1pzg	T.GONDII LDH1 COMPLEXED WITH APAD AND SULFATE AT 1.6 ANGSTROMS	B	A	6	S260	E265	4.66	SFLNDE	0.00	0.73	4.01	0.02	0.34	3.17		
1q4u	CRYSTAL STRUCTURE OF 4-HYDROXYBENZOYL COA THIOESTERASE FROM2 ARTHROBACTER SP. STRAIN SU COMPLEXED WITH 4-HYDROXYBENZYL COA	A	B	8	R57	H64	4.79	RQRWGLVH	3.36	0.00	4.16	1.84	0.00	0.00	0.00	1.57
1q5a	S-SHAPED TRANS INTERACTIONS OF CADHERINS MODEL BASED ON FITTING C-2 CADHERIN (1L3W) TO 3D MAP OF DESMOSOMES OBTAINED BY ELECTRON3 TOMOGRAPHY	A	B	8	V81	V88	5.64	VSENGSPV	1.00	4.50	0.00	0.06	0.00	0.00	1.00	1.00
1q5b	LAMBDA-SHAPED TRANS AND CIS INTERACTIONS OF CADHERINS MODEL BASED ON2 FITTING C-CADHERIN (1L3W) TO 3D MAP OF DESMOSOMES OBTAINED BY3 ELECTRON TOMOGRAPHY	B	C	8	A80	P87	5.9	AVSENGSP	0.00	0.05	4.50	0.00	0.05	0.00	4.50	1.00
1q5b	LAMBDA-SHAPED TRANS AND CIS INTERACTIONS OF CADHERINS MODEL BASED ON2 FITTING C-CADHERIN (1L3W) TO 3D MAP OF DESMOSOMES OBTAINED BY3 ELECTRON TOMOGRAPHY	B	C	8	V81	V88	5.64	VSENGSPV	0.05	4.50	0.00	0.05	0.00	4.50	1.00	1.00
1q5b	LAMBDA-SHAPED TRANS AND CIS INTERACTIONS OF CADHERINS MODEL BASED ON2 FITTING C-CADHERIN (1L3W) TO 3D MAP OF DESMOSOMES OBTAINED BY3 ELECTRON TOMOGRAPHY	A	C	7	Q41	P47	6	QGADNPP	1.00	0.00	0.00	1.00	1.00	1.00	1.00	
1q5b	LAMBDA-SHAPED TRANS AND CIS INTERACTIONS OF CADHERINS MODEL BASED ON2 FITTING C-CADHERIN (1L3W) TO 3D MAP OF DESMOSOMES OBTAINED BY3 ELECTRON TOMOGRAPHY	A	C	8	G42	G49	5.92	GADNPPQG	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.00
1qks	CYTOCHROME CD1 NITRITE REDUCTASE, OXIDISED FORM	B	A	8	M287	Y294	4.34	MTYDEQEY	0.76	0.00	0.00	0.00	0.17	0.25	4.50	0.87
1qle	CRYO-STRUCTURE OF THE PARACOCCUS DENITRIFICANS FOUR-SUBUNIT2 CYTOCHROME C OXIDASE IN THE COMPLETELY OXIDIZED STATE3 COMPLEXED WITH AN ANTIBODY FV	L	H	8	H90	L97	6.01	HHYGTPPL	0.00	4.50	1.00	0.00	0.00	0.53	4.50	0.22
1rcu	X-RAY STRUCTURE OF TM1055 NORTHEAST STRUCTURAL GENOMICS2 CONSORTIUM TARGET VT76	D	B	6	Y145	R150	5.51	YLDNRR	0.00	0.00	1.00	4.50	0.66	0.00		
1rve	THE CRYSTAL STRUCTURE OF ECORV ENDONUCLEASE AND OF ITS2 COMPLEXES WITH COGNATE AND NON-COGNATE DNA FRAGMENTS	B	A	8	I23	I30	5.21	IISAEGKI	0.89	1.65	0.00	0.00	0.00	0.00	0.00	3.07
1s5d	CHOLERA HOLOTOXIN WITH AN A-SUBUNIT Y30S MUTATION, CRYSTAL FORM 2	E	D	6	L31	E36	4.72	LAGKRE	1.00	0.00	0.00	0.00	1.98	1.67		
1sr4	CRYSTAL STRUCTURE OF THE HAEMOPHILUS DUCREYI CYTOLETHAL2 DISTENDING TOXIN	A	B	8	P120	F127	4.35	PGKHREYF	1.00	0.00	0.00	0.00	2.75	1.05	0.00	0.76
1t6u	NICKEL SUPEROXIDE DISMUTASE (NISOD) NATIVE 1.30 A STRUCTURE	C	E	8	H1	V8	5.8	HCDLPCGV	0.81	1.00	0.00	1.93	0.00	0.00	0.00	4.50
1ul3	CRYSTAL STRUCTURE OF PII FROM SYNECHOCYSTIS SP. PCC 6803	B	C	8	V100	K107	4.26	VRIRTGEK	0.62	0.98	2.15	1.08	0.00	0.00	0.00	0.00

PDBID	Structure Title	Chain	Partner	Residues	First	Last	Termini Distance	Hot Loop Sequence	ΔΔG1	ΔΔG2	ΔΔG3	ΔΔG4	ΔΔG5	ΔΔG6	ΔΔG7	ΔΔG8
1vgw	CRYSTAL STRUCTURE OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL2 SYNTHASE	B	A	8	R146	S153	5.6	RAESGQIS	0.29	4.34	0.00	0.00	0.00	0.00	2.72	3.48
1wmh	CRYSTAL STRUCTURE OF A PB1 DOMAIN COMPLEX OF PROTEIN KINASE2 C IOTA AND PAR6 ALPHA	A	B	8	W61	P68	5.37	WIDEEGDP	0.00	0.00	2.06	0.00	0.00	0.00	1.90	0.96
1wt5	THE CRYSTAL STRUCTURE OF A HUMANIZED ANTIBODY FV 528	B	D	7	G100	D106	4.7	GGPYFFD	0.00	0.00	0.00	1.00	3.95	1.00	0.00	
1xef	CRYSTAL STRUCTURE OF THE ATP/MG2+ BOUND COMPOSITE DIMER OF HLYB-NBD	B	A	6	L636	E641	5.93	LDYESE	0.00	1.46	1.56	0.93	1.58	0.00		
1xfs	X-RAY CRYSTAL STRUCTURE OF PROTEIN NE0264 FROM NITROSOMONAS EUROPAEA.2 NORTHEAST STRUCTURAL GENOMICS CONSORTIUM TARGET NER5.	A	B	8	L92	A99	4.71	LVKNYRPA	0.70	0.00	0.00	0.00	0.56	1.72	4.50	0.00
1xkp	CRYSTAL STRUCTURE OF THE VIRULENCE FACTOR YOPN IN COMPLEX WITH ITS2 HETERODIMERIC CHAPERONE SYCN-YSCB	A	B	7	Q33	V39	5	QFRGESV	0.00	1.48	0.00	0.00	1.98	0.00	0.87	
1yfn	VERSATILE MODES OF PEPTIDE RECOGNITION BY THE AAA+ ADAPTOR2 PROTEIN SSPB- THE CRYSTAL STRUCTURE OF A SSPB-RSEA COMPLEX	F	B	6	P4	W9	5.74	PAPHQW	4.50	0.00	0.74	0.00	0.00	1.00		
2a19	PKR KINASE DOMAIN- EIF2ALPHA- AMP-PNP COMPLEX.	B	A	7	C485	T491	5.65	CDTAFET	0.00	2.11	0.66	0.00	1.04	1.64	0.00	
2a26	CRYSTAL STRUCTURE OF THE N-TERMINAL, DIMERIZATION DOMAIN OF2 SIAH INTERACTING PROTEIN	B	A	7	A20	R26	6.03	ATRKRVR	0.00	0.00	4.50	0.00	0.89	0.61	0.00	
2a2m	CRYSTAL STRUCTURE OF A PUTATIVETENA FAMILY TRANSCRIPTIONAL REGULATOR2 (BT_3146) FROM BACTEROIDES THETAIOTAOMICRON VPI-5482 AT 1.88 A3 RESOLUTION	B	A	7	Y120	L126	5.62	YHQTWHL	0.00	4.50	0.00	0.44	0.98	0.69	0.00	
2a2m	CRYSTAL STRUCTURE OF A PUTATIVETENA FAMILY TRANSCRIPTIONAL REGULATOR2 (BT_3146) FROM BACTEROIDES THETAIOTAOMICRON VPI-5482 AT 1.88 A3 RESOLUTION	B	A	7	H121	R127	5.73	HQTWHLR	4.50	0.00	0.44	0.98	0.69	0.00	0.65	
2ae8	CRYSTAL STRUCTURE OF IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE FROM2 STAPHYLOCOCCUS AUREUS SUBSP. AUREUS N315	B	C	8	I94	A101	5.33	IPXDETLA	0.00	1.01	N/A	4.50	0.00	0.00	0.85	0.00
2afh	CRYSTAL STRUCTURE OF NUCLEOTIDE-FREE AV2-AV1 COMPLEX	A	F	6	D122	G127	5.32	DIVFGG	0.00	3.99	0.72	1.38	0.00	0.00		
2aj7	CRYSTAL STRUCTURE OF A PUTATIVE CONTRACTILE PROTEIN (BH3618) FROM2 BACILLUS HALODURANS AT 1.67 A RESOLUTION	A	B	7	F40	Y46	5.53	FAAGTPY	2.04	0.31	0.00	0.00	0.40	0.64	2.26	
2ass	CRYSTAL STRUCTURE OF THE SKP1-SKP2-CKS1 COMPLEX	C	B	6	M3038	W3043	5.71	MSESEW	0.00	1.11	1.50	1.77	3.02	0.00		
2ast	CRYSTAL STRUCTURE OF SKP1-SKP2-CKS1 IN COMPLEX WITH A P27 PEPTIDE	C	B	6	M3038	W3043	5.75	MSESEW	0.00	1.39	1.86	2.10	2.92	0.00		
2axi	HDM2 IN COMPLEX WITH A BETA-HAIRPIN	B	A	7	F21	E27	5.95	FEXLDWE	2.23	1.00	N/A	1.59	0.00	0.00	0.00	
2bfd	REACTIVITY MODULATION OF HUMAN BRANCHED-CHAIN ALPHA-2 KETOACID DEHYDROGENASE BY AN INTERNAL MOLECULAR SWITCH	B	A	8	R118	F125	5.72	RYRSGDLF	0.50	0.00	0.00	1.00	0.00	0.03	1.04	2.98
2cvo	CRYSTAL STRUCTURE OF PUTATIVE N-ACETYL-GAMMA-GLUTAMYL-2 PHOSPHATE REDUCTASE (AK071544) FROM RICE (ORYZA SATIVA)	D	B	6	A260	E265	5.87	ANLYTE	0.00	0.00	2.00	2.99	0.74	0.00		
2czc	CRYSTAL STRUCTURE OF GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE FROM2 PYROCOCCUS HORIKOSHII OT3	D	C	8	R264	N271	4.95	RDLHREWN	1.13	3.50	0.00	1.57	0.00	1.92	1.00	1.00
2dcn	CRYSTAL STRUCTURE OF 2-KETO-3-DEOXYGLUCONATE KINASE FROM SULFOLOBUS2 TOKODAII COMPLEXED WITH 2-KETO-6-PHOSPHOGLUCONATE (ALPHA-FURANOSE3 FORM)	F	B	5	H95	P99	4.3	HYPVP	2.30	1.09	0.20	0.86	0.92			
2e4m	CRYSTAL STRUCTURE OF HEMAGGLUTININ SUBCOMPONENT COMPLEX (HA-2 33/HA-17) FROM CLOSTRIDIUM BOTULINUM SEROTYPE D STRAIN 4947	B	A	8	Q97	V104	4.3	QDYISRNV	1.09	1.25	1.46	0.93	0.47	0.00	0.67	0.00
2fbn	PLASMODIUM FALCIPARUM PUTATIVE FK506-BINDING PROTEIN2 PFL2275C, C-TERMINAL TPR-CONTAINING DOMAIN	B	A	6	M133	L138	5.11	MYFGFL	0.00	0.00	1.07	0.00	2.14	4.28		
2fd6	STRUCTURE OF HUMAN UROKINASE PLASMINOGEN ACTIVATOR IN COMPLEX WITH2 UROKINASE RECEPTOR AND AN ANTI-UPAR ANTIBODY AT 1.9 A	A	U	8	S21	I28	5.7	SNKYFSNI	0.00	0.31	0.00	2.11	3.83	0.92	0.00	1.06
2fe8	SARS CORONAVIRUS PAPAIN-LIKE PROTEASE: STRUCTURE OF A VIRAL2 DEUBIQUITINATING ENZYME	A	B	8	T266	H273	5.31	TGNYQCGH	0.00	0.00	0.00	0.81	0.99	0.05	0.00	4.50
2hlc	HL COLLAGENASE STRUCTURE AT 1.7A RESOLUTION	A	B	6	I35	V40	5	ITQRRV	0.00	0.63	0.37	3.81	0.65	1.96		
2hlc	HL COLLAGENASE STRUCTURE AT 1.7A RESOLUTION	B	A	7	I35	W41	5.94	ITQRRVW	0.00	1.00	0.64	3.93	1.17	0.55	0.00	
2hq7	CRYSTAL STRUCTURE OF PROTEIN RELATED TO GENERAL STRESS PROTEIN2 26(GS26) OF B.SUBTILIS (PYRIDOXINEPHOSPHATE OXIDASE FAMILY)3 (NP_350077.1) FROM CLOSTRIDIUM	B	A	8	N25	N32	6.16	NGENGYPN	0.33	0.00	0.00	0.00	0.00	2.07	1.00	4.35
2hzs	STRUCTURE OF THE MEDIATOR HEAD SUBMODULE MED8C/18/20	A	B	8	V92	G99	5.87	VFNGSSTG	0.00	1.19	0.74	0.00	2.70	0.77	1.33	0.00
2i9b	CRYSTAL STRUCTURE OF ATF-UROKINASE RECEPTOR COMPLEX	A	E	8	S21	I28	5.83	SNKYFSNI	1.13	0.54	0.00	1.78	3.25	0.44	0.53	1.00
2iij	STRUCTURE OF HUMAN ASF1A IN COMPLEX WITH HISTONE H3	B	A	6	R129	R134	5.12	RIRGER	1.00	1.00	0.19	0.00	1.00	1.00		
2j9c	STRUCTURE OF GLNK1 WITH BOUND EFFECTORS INDICATES2 REGULATORY MECHANISM FOR AMMONIA UPTAKE	B	C	8	V100	E107	4.49	VRVRTKEE	0.78	1.05	1.22	1.33	0.00	0.50	0.00	0.00
2jb0	CRYSTAL STRUCTURE OF THE MUTANT H573A OF THE NUCLEASE2 DOMAIN OF COLE7 IN COMPLEX WITH IM7	A	B	5	D52	Y56	4.56	DLIYY	1.07	0.00	0.00	1.21	1.14			
2jdi	GROUND STATE STRUCTURE OF F1-ATPASE FROM BOVINE HEART2 MITOCHONDRIA (BOVINE F1-ATPASE CRYSTALLISED IN THE ABSENCE3 OF AZIDE)	H	G	7	F20	V26	6.16	FASPTQV	0.00	0.00	0.00	0.77	0.34	1.35	2.37	
2jra	A NOVEL DOMAIN-SWAPPED SOLUTION NMR STRUCTURE OF PROTEIN RPA2121 FROM2 RHODOPSEUDOMONAS PALUSTRIS. NORTHEAST STRUCTURAL GENOMICS TARGET RPT6	A	B	8	I46	Y53	4.91	IAHGDDRY	1.38	0.00	2.17	0.00	0.00	0.00	0.00	4.50
2jxc	STRUCTURE OF THE EPS15-EH2 STONIN2 COMPLEX	B	A	8	W309	L316	5.74	WRATNPFL	1.00	0.56	0.00	0.14	4.50	0.00	1.00	0.00
2kpz	HUMAN NEDD4 3RD WW DOMAIN COMPLEX WITH THE HUMAN T-CELL LEUKEMIA VIRUS2 1 GAG-PRO POLIPROTEIN DERIVED PEPTIDE SDPQIPPPYVEP	A	B	8	I30	T37	4.55	IDHNTKTT	4.50	0.00	1.00	0.00	0.00	0.09	0.00	4.50
2ksp	MECHANISM FOR THE SELECTIVE INTERACTION OF C-TERMINAL EH-DOMAIN2 PROTEINS WITH SPECIFIC NPF-CONTAINING PARTNERS	B	A	7	P147	E153	6.1	PYNPFEE	0.00	0.70	0.00	1.13	1.88	1.16	0.00	
2ldx	CHARACTERIZATION OF THE ANTIGENIC SITES ON THE REFINED 3-2 ANGSTROMS RESOLUTION STRUCTURE OF MOUSE TESTICULAR LACTATE3 DEHYDROGENASE C4	C	B	8	V293	T300	5.83	VLGESGIT	0.00	0.00	0.00	4.48	4.50	0.00	0.40	1.00
2ltz	SMURF2 WW3 DOMAIN IN COMPLEX WITH A SMAD7 DERIVED PEPTIDE	A	B	8	N307	Y314	5.99	NTATGRVY	0.00	4.50	0.00	0.00	0.00	1.00	0.00	1.00
2ns1	CRYSTAL STRUCTURE OF THE E. COLI AMMONIA CHANNEL AMTB COMPLEXED WITH2 THE SIGNAL TRANSDUCTION PROTEIN GLNK	B	A	8	E44	F51	5.12	ELYRGAEF	0.00	0.42	1.55	3.04	0.00	0.00	0.00	1.22
2nwt	NMR STRUCTURE OF PROTEIN UPF0165 PROTEIN AF_2212 FROM2 ARCHAEOGLOBUS FULGIDUS; NORTHEAST STRUCTURAL GENOMICS3 CONSORTIUM TARGET GR83	A	B	8	A7	F14	5.67	AVYENGVF	0.00	1.00	4.50	0.45	0.00	0.00	0.00	4.14
2o56	CRYSTAL STRUCTURE OF A MEMBER OF THE ENOLASE SUPERFAMILY FROM2 SALMONELLA TYPHIMURIUM	H	F	8	L80	Q87	5.85	LKKTFWGQ	0.00	0.00	0.00	0.14	0.75	1.67	0.00	4.50
2ogk	CRYSTAL STRUCTURE OF PROTEIN AF2318 FROM ARCHAEGLOBUS2 FULGIDUS, PFAM DUF54	C	D	7	V15	R21	6.01	VHSTEDR	0.00	0.61	0.85	2.95	0.06	0.30	0.00	
2oiz	CRYSTAL STRUCTURE OF THE TRYPTAMINE-DERIVED (INDOL-3-ACETAMIDE)-TTQ2 ADDUCT OF AROMATIC AMINE DEHYDROGENASE	D	B	8	R147	L154	5.47	RPGYEFFL	3.11	0.00	0.00	1.61	0.00	4.50	0.00	0.00
2oqb	CRYSTAL STRUCTURE OF THE N-TERMINAL DOMAIN OF COACTIVATOR-ASSOCIATED2 METHYLTRANSFERASE 1 (CARM1)	B	A	8	S92	F99	5.7	SRVGRQSF	0.00	0.70	4.50	0.00	0.16	1.00	1.32	0.00
2ovp	STRUCTURE OF THE SKP1-FBW7 COMPLEX	B	A	8	A2305	R2312	3.67	AQTCRYWR	0.00	0.00	4.50	1.00	1.41	0.00	0.00	0.00
2peh	CRYSTAL STRUCTURE OF THE UHM DOMAIN OF HUMAN SPF45 IN COMPLEX WITH2 SF3B155-ULM5	B	D	8	R375	V382	5.65	RYFGGRVV	4.50	0.82	2.28	0.00	0.00	0.00	0.00	0.00
2ppb	CRYSTAL STRUCTURE OF THE T. THERMOPHILUS RNAP POLYMERASE ELONGATION2 COMPLEX WITH THE NTP SUBSTRATE ANALOG AND ANTIBIOTIC STREPTOLYDIGIN	E	D	7	V13	L19	5.93	VDSKYRL	0.00	0.00	1.29	1.00	1.79	0.22	4.50	

PDBID	Structure Title	Chain	Partner	Residues	First	Last	Termini Distance	Hot Loop Sequence	ΔΔG1	ΔΔG2	ΔΔG3	ΔΔG4	ΔΔG5	ΔΔG6	ΔΔG7	ΔΔG8
2q4o	ENSEMBLE REFINEMENT OF THE CRYSTAL STRUCTURE OF A LYSINE2 DECARBOXYLASE-LIKE PROTEIN FROM ARABIDOPSIS THALIANA GENE AT2G37210	B	A	7	A161	I167	5.6	AVEEGFI	4.50	0.00	0.00	1.93	0.00	4.50	4.50	
2qeu	CRYSTAL STRUCTURE OF PUTATIVE CARBOXYMUCONOLACTONE DECARBOXYLASE2 (YP_555818.1) FROM BURKHOLDERIA XENOVORANS LB400 AT 1.65 A RESOLUTION	C	A	6	W117	R122	5.18	WGKTGR	2.97	0.00	1.23	0.00	0.00	4.50		
2qmh	STRUCTURE OF V267F MUTANT HPRK/P	F	D	7	E204	I210	4.55	EIRGLGI	0.77	1.09	0.66	0.00	1.64	0.00	0.90	
2qnr	HUMAN SEPTIN 2 IN COMPLEX WITH GDP	B	A	8	R256	V263	5.05	RLYPWGVV	0.14	0.00	1.30	0.54	2.19	0.00	4.50	0.00
2qya	CRYSTAL STRUCTURE OF AN UNCHARACTERIZED CONSERVED PROTEIN2 FROM METHANOPYRUS KANDLERI	B	A	8	H9	V16	5.47	HGTGDTVV	1.59	0.00	1.11	0.00	0.82	1.29	0.00	0.00
2rsk	RNA APTAMER AGAINST PRION PROTEIN IN COMPLEX WITH THE PARTIAL BINDING2 PEPTIDE	D	B	5	G37	K41	4.92	GQWNK	0.00	1.00	1.00	4.50	0.00			
2vq5	X-RAY STRUCTURE OF NORCOCLAURINE SYNTHASE FROM THALICTRUM2 FLAVUM IN COMPLEX WITH DOPAMINE AND HYDROXYBENZALDEHYDE	B	A	5	L132	V136	4.57	LDLGV	4.50	0.66	0.76	0.00	0.00			
2vvm	THE STRUCTURE OF MAO-N-D5, A VARIANT OF MONOAMINE OXIDASE2 FROM ASPERGILLUS NIGER.	B	A	7	W97	W103	5.19	WHQSHVW	1.24	2.04	0.00	0.15	4.50	0.00	0.00	
2wam	CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS UNKNOWN2 FUNCTION PROTEIN RV2714	B	A	6	F90	F95	4.32	FKTDHF	0.96	0.60	0.12	1.62	1.07	2.06		
2x7n	MECHANISM OF EIF6S ANTI-ASSOCIATION ACTIVITY	D	B	8	F21	I28	5.74	FVRGDSKI	0.00	0.00	0.00	0.00	3.81	0.00	1.00	1.00
2xsx	CRYSTAL STRUCTURE OF HUMAN BETA ENOLASE ENOB	A	B	7	P398	L404	5.99	PCRSERL	4.50	1.00	1.00	0.00	1.00	0.00	1.00	
2y0n	CRYSTAL STRUCTURE OF THE COMPLEX BETWEEN DOSAGE2 COMPENSATION FACTORS MSL1 AND MSL3	H	D	8	P573	P580	5.78	PVVAFGRP	0.00	0.00	0.00	0.00	2.75	0.00	4.50	1.00
2y0n	CRYSTAL STRUCTURE OF THE COMPLEX BETWEEN DOSAGE2 COMPENSATION FACTORS MSL1 AND MSL3	H	D	8	V574	L581	5.66	VVAFGRPL	0.00	0.00	0.00	2.75	0.00	4.50	1.00	0.73
2ylb	STRUCTURE OF SALMONELLA TYPHIMURIUM HFQ AT 1.15 A	A	F	7	L45	S51	6.1	LLKNTVS	1.05	0.00	0.00	0.00	0.00	1.34	3.49	
2z3h	CRYSTAL STRUCTURE OF BLASTICIDIN S DEAMINASE (BSD)2 COMPLEXED WITH DEAMINOHYDROXY BLASTICIDIN S	C	B	8	V46	P53	5.41	VYHFTGGP	0.00	0.00	3.20	0.00	1.56	0.00	0.00	4.10
2zih	CRYSTAL STRUCTURE OF YEAST VPS74	D	C	8	K199	M206	5.81	KNFFLFDM	0.00	0.10	0.00	2.52	0.66	0.62	0.00	1.00
2zih	CRYSTAL STRUCTURE OF YEAST VPS74	C	D	7	N200	M206	6.17	NFFLFDM	0.00	1.00	1.83	1.41	1.00	0.00	0.00	
2zii	CRYSTAL STRUCTURE OF YEAST VPS74-N-TERM TRUNCATION VARIANT	A	B	8	K199	M206	5.83	KNFFLFDM	0.00	0.24	0.00	2.48	1.09	0.00	0.00	1.00
2zki	CRYSTAL STRUCTURE OF HYPOTHETICAL TRP REPRESSOR BINDING2 PROTEIN FROM SUL FOLOBUS TOKODAII (ST0872)	C	B	8	A116	H123	6.15	ASTVHGGH	0.00	0.00	0.61	4.50	1.08	0.00	0.00	0.61
3ajv	SPLICING ENDONUCLEASE FROM AEROPYRUM PERNIX	A	D	6	V152	L157	5.51	VDRTGL	0.00	0.74	2.32	0.00	0.00	1.06		
3ba3	CRYSTAL STRUCTURE OF PYRIDOXAMINE 5'-PHOSPHATE OXIDASE-LIKE PROTEIN2 (NP_783940.1) FROM LACTOBACILLUS PLANTARUM AT 1.55 A RESOLUTION	A	B	8	T20	D27	5.37	TAVNNEAD	0.08	0.00	0.00	0.00	0.25	0.64	4.50	0.95
3bdx	CRYSTAL STRUCTURE OF THE UNSTABLE AND HIGHLY FIBRILLOGENIC PRO7SER2 MUTANT OF THE RECOMBINANT VARIABLE DOMAIN 6AJL2	C	A	8	S93	V100	5.91	SYDSSNHV	0.00	1.03	0.00	0.00	0.00	0.62	2.31	1.38
3bt1	STRUCTURE OF UROKINASE RECEPTOR, UROKINASE AND VITRONECTIN COMPLEX	A	U	8	S21	I28	5.75	SNKYFSNI	0.24	0.34	1.00	1.67	3.30	0.40	0.00	0.84
3bxw	CRYSTAL STRUCTURE OF STABILIN-1 INTERACTING CHITINASE-LIKE PROTEIN,2 SI-CLP	A	B	8	I209	L216	4.55	ITPGTDQL	4.50	0.00	0.00	0.00	0.00	2.61	0.00	0.65
3c5x	CRYSTAL STRUCTURE OF THE PRECURSOR MEMBRANE PROTEIN- ENVELOPE PROTEIN2 HETERODIMER FROM THE DENGUE 2 VIRUS AT LOW PH	A	C	6	P243	Q248	4.53	PHAKKQ	0.00	0.70	0.01	1.00	1.97	0.00		
3cdx	CRYSTAL STRUCTURE OF2 SUCCINYLGLUTAMATEDESUCCINYLASE/ASPARTOACYLASE FROM3 RHODOBACTER SPHAEROIDES	F	E	8	S31	G38	5.68	SRNNSGWG	1.39	0.00	0.15	0.24	3.25	0.00	1.16	0.00
3cea	CRYSTAL STRUCTURE OF MYO-INOSITOL 2-DEHYDROGENASE (NP_786804.1) FROM2 LACTOBACILLUS PLANTARUM AT 2.40 A RESOLUTION	D	C	8	V275	V282	4.87	VFNDQGVV	0.00	0.72	0.94	0.05	0.49	0.00	4.19	0.29
3cfi	NANOBODY-AIDED STRUCTURE DETERMINATION OF THE EPSI:EPSJ PSEUDOPILIN2 HETERDIMER FROM VIBRIO VULNIFICUS	C	B	8	K97	W104	5.63	KWLGGRDW	0.00	0.00	1.14	0.00	0.00	0.73	1.58	2.10
3chb	CHOLERA TOXIN B-PENTAMER COMPLEXED WITH GM1 PENTASACCHARIDE	H	G	8	S30	M37	5.54	SLAGKREM	0.00	1.00	0.00	0.00	0.00	4.50	1.67	0.00
3co2	MLOTIK1 ION CHANNEL CYCLIC-NUCLEOTIDE BINDING DOMAIN MUTANT	A	D	7	V282	V288	6.04	VATPNPV	0.48	0.00	0.24	1.00	0.00	4.50	0.83	
3dd2	CRYSTAL STRUCTURE OF AN RNA APTAMER BOUND TO HUMAN THROMBIN	L	H	6	L3	E8	5.36	LRPLFE	0.33	0.00	4.50	0.75	0.00	2.77		
3dem	CUB1-EGF-CUB2 DOMAIN OF HUMAN MASP-1/3	A	B	8	H139	Y146	5.37	HNYIGGYY	2.04	0.00	1.47	4.50	0.00	0.00	0.00	0.00
3e8m	STRUCTURE-FUNCTION ANALYSIS OF 2-KETO-3-DEOXY-D-GLYCERO-D-GALACTO-2 NONONATE-9-PHOSPHATE (KDN) PHOSPHATASE DEFINES A NEW CLAD WITHIN THE3 TYPE C0 HAD SUBFAMILY	D	C	8	F21	E28	5.5	FYDQTGNE	1.00	1.40	2.79	0.18	0.00	0.00	0.00	0.00
3eev	CRYSTAL STRUCTURE OF CHLORAMPHENICOL ACETYLTRANSFERASE VCA0300 FROM2 VIBRIO CHOLERAE O1 BIOVAR ELTOR	C	B	8	H78	T85	4.91	HRSDWIST	2.22	0.00	0.06	0.29	2.17	1.55	1.61	0.27
3euw	CRYSTAL STRUCTURE OF A MYO-INOSITOL DEHYDROGENASE FROM CORYNEBACTERIUM2 GLUTAMICUM ATCC 13032	A	C	8	K260	E267	5.79	KHNAESTE	0.00	1.00	0.22	0.00	0.00	0.00	2.66	1.14
3f3f	CRYSTAL STRUCTURE OF THE NUCLEOPORIN PAIR NUP85-SEH1, SPACE GROUP P21	A	D	6	A333	E338	5.64	ATYSNE	0.00	0.69	0.13	4.50	0.00	1.00		
3f9k	TWO DOMAIN FRAGMENT OF HIV-2 INTEGRASE IN COMPLEX WITH LEDGF IBD	A	I	6	Y143	Q148	5.29	YNPQSQ	0.00	1.21	4.13	1.00	0.00	0.00		
3gke	CRYSTAL STRUCTURE OF DICAMBA MONOOXYGENASE	A	B	6	I48	F53	4.51	ICPHRF	1.42	0.00	0.00	1.59	2.02	1.71		
3gke	CRYSTAL STRUCTURE OF DICAMBA MONOOXYGENASE	B	C	8	I48	P55	5.35	ICPHRFAP	0.82	0.00	0.00	1.98	1.73	1.64	0.00	0.73
3h8v	HUMAN UBIQUITIN-ACTIVATING ENZYME 5 IN COMPLEX WITH ATP	A	B	8	Y285	F292	4.66	YNAMQDFF	0.00	0.00	0.00	0.05	0.70	2.30	0.00	2.47
3hdg	CRYSTAL STRUCTURE OF THE N-TERMINAL DOMAIN OF AN2 UNCHARACTERIZED PROTEIN (WS1339) FROM WOLINELLA3 SUCCINOGENES	B	A	7	E15	R21	5.25	EDDTDAR	0.00	1.95	1.06	0.50	1.30	0.00	0.00	
3hi6	CRYSTAL STRUCTURE OF INTERMEDIATE AFFINITY I DOMAIN OF2 INTEGRIN LFA-1 WITH THE FAB FRAGMENT OF ITS ANTIBODY AL-57	H	A	8	S99	A106	5.93	SYDFWSNA	0.00	2.00	1.79	1.27	2.94	0.12	0.00	0.00
3hsh	CRYSTAL STRUCTURE OF HUMAN COLLAGEN XVIII TRIMERIZATION DOMAIN2 (TETRAGONAL CRYSTAL FORM)	B	A	8	F26	L33	4.52	FVAEQEEL	2.55	0.00	0.00	0.00	0.00	2.34	0.00	1.00
3hvd	THE PROTECTIVE ANTIGEN COMPONENT OF ANTHRAX TOXIN FORMS2 FUNCTIONAL OCTAMERIC COMPLEXES	H	G	8	S221	T228	5.95	SSPEKWST	0.00	0.00	4.50	2.25	0.00	2.22	0.00	0.00
3iaa	CRYSTAL STRUCTURE OF CALG2, CALICHEAMICIN GLYCOSYLTRANSFERASE, TDP2 BOUND FORM	B	A	7	S133	F139	5.61	SNEHYSF	0.00	0.00	0.67	2.69	0.00	2.60	0.00	
3ijj	TERNARY COMPLEX OF MACROPHAGE MIGRATION INHIBITORY FACTOR (MIF) BOUND2 BOTH TO 4-HYDROXYPHENYLPYRUVATE AND TO THE ALLOSTERIC INHIBITOR3 AV1013 (R-	B	A	7	V106	T112	5.4	VGWNNST	4.50	0.00	0.93	0.22	0.00	0.33	0.71	
3ik5	SIVMAC239 NEF IN COMPLEX WITH TCR ZETA ITAM 1 POLYPEPTIDE (A63-R80)	B	A	6	Y72	L77	5.03	YNELNL	2.37	0.00	1.38	1.88	0.00	1.27		
3j09	HIGH RESOLUTION HELICAL RECONSTRUCTION OF THE BACTERIAL P-TYPE ATPASE2 COPPER TRANSPORTER COPA	B	A	7	V49	T55	5.94	VNLATET	0.09	4.50	1.00	0.56	1.00	0.00	0.00	

PDBID	Structure Title	Chain	Partner	Residues	First	Last	Termini Distance	Hot Loop Sequence	ΔΔG1	ΔΔG2	ΔΔG3	ΔΔG4	ΔΔG5	ΔΔG6	ΔΔG7	ΔΔG8
3jqy	CRYSTAL STRUCUTRE OF THE POLYSIA SPECIFIC ACETYLTRANSFERASE NEUO	B	A	6	F150	K155	5.8	FDIHSK	4.50	0.85	1.75	0.56	0.00	0.96		
3jrv	STRUCTURE OF POXVIRUS K7 PROTEIN IN COMPLEX WITH RNA HELICASE DDX3	C	A	6	S83	R88	5.76	SFFSDR	0.98	2.92	2.89	0.04	1.33	0.00		
3l2h	CRYSTAL STRUCTURE OF PUTATIVE SUGAR PHOSPHATE ISOMERASE (AFE_0303)2 FROM ACIDITHIOBACILLUS FERROOXIDANS ATCC 23270 AT 1.85 A RESOLUTION	C	A	8	Y138	N145	5.04	YRNNGEWN	1.18	0.00	0.00	0.00	0.00	2.34	1.02	0.81
3lm2	CRYSTAL STRUCTURE OF PUTATIVE KINASE. (17743352) FROM AGROBACTERIUM2 TUMEFACIENS STR. C58 (DUPONT) AT 1.70 A RESOLUTION	A	B	7	V65	L71	5.36	VVHNKPL	0.00	2.63	1.51	0.00	1.55	0.00	0.93	
3mdi	CRYSTAL STRUCTURE OF THE 25KDA SUBUNIT OF HUMAN CLEAVAGE FACTOR IM IN2 COMPLEX WITH RNA UGUAAA	B	A	6	K29	E34	5.98	KPLTLE	0.00	3.03	1.69	0.30	0.71	0.00		
3mmz	CRYSTAL STRUCTURE OF PUTATIVE HAD FAMILY HYDROLASE FROM STREPTOMYCES2 AVERMITILIS MA-4680	D	B	8	L29	E36	5.6	LIDSDGRE	0.83	0.66	4.50	0.00	0.00	0.00	0.00	0.00
3n6q	CRYSTAL STRUCTURE OF YGHZ FROM E. COLI	G	H	8	M102	S109	6.03	MWPGPYGS	0.00	0.00	4.50	0.00	2.11	1.00	0.00	0.00
3nvn	MOLECULAR MECHANISM OF GUIDANCE CUE RECOGNITION	A	B	8	G199	H206	5.84	GPSSLSSH	0.00	0.00	1.48	0.86	0.83	1.15	0.24	0.50
3nvq	MOLECULAR MECHANISM OF GUIDANCE CUE RECOGNITION	A	B	8	G271	V278	4.58	GGESSLSV	0.00	0.00	4.50	1.25	0.89	0.71	0.71	0.66
3ovg	THE CRYSTAL STRUCTURE OF AN AMIDOHYDROLASE FROM MYCOPLASMA SYNOVIAE2 WITH ZN ION BOUND	C	B	8	Y135	V142	6.15	YNYNGPVV	1.00	0.06	0.65	0.72	0.00	1.47	1.00	0.00
3p0g	STRUCTURE OF A NANOBODY-STABILIZED ACTIVE STATE OF THE BETA22 ADRENOCEPTOR	A	B	8	A134	Y141	5.65	AITSPFKY	0.00	0.87	0.00	4.50	0.10	0.69	0.00	0.00
3pqr	CRYSTAL STRUCTURE OF METARHODOPSIN II IN COMPLEX WITH A C-TERMINAL2 PEPTIDE DERIVED FROM THE GALPHA SUBUNIT OF TRANSDUCIN	B	A	6	L344	L349	5.17	LKDVGL	2.47	0.00	0.00	0.88	0.00	0.67		
3psm	.98A CRYSTAL STRUCTURE OF A DIMERIC PLANT DEFENSIN SPE10	B	A	6	C35	R40	5.53	CRDDFR	0.00	2.13	0.89	2.43	0.00	0.00		
3pyi	STRUCTURE OF THE N-TERMINAL DOMAIN OF C. ELEGANS SAS-6	A	B	8	S150	G157	4.86	SKTPISKG	0.08	0.00	0.81	4.50	0.00	1.09	0.93	0.00
3q62	CRYSTAL STRUCTURE OF 3-HYDROXYDECANOYL-(ACYL CARRIER PROTEIN)2 DEHYDRATASE FROM YERSINIA PESTIS	A	B	7	H71	V77	6.17	HFIGDPV	1.65	1.14	0.47	0.00	2.26	0.00	0.00	
3qht	CRYSTAL STRUCTURE OF THE MONOBODY YSMB-1 BOUND TO YEAST SUMO	C	A	6	S77	Y82	4.43	SYYDLY	0.00	2.31	0.62	0.86	0.00	0.46		
3qmn	CRYSTAL STRUCTURE OF 4'-PHOSPHOPANTETHEINYL TRANSFERASE ACPS FROM2 VIBRIO CHOLERAE O1 BIOVAR ELTOR	C	B	7	S111	A117	5.79	SDERHYA	0.53	0.00	3.15	1.29	0.00	1.00	0.00	
3qs1	CRYSTAL STRUCTURE OF KNI-10006 COMPLEX OF PLASMEPSIN I (PMI) FROM2 PLASMODIUM FALCIPARUM	C	B	7	Q277	M283	6.19	QFISLCM	0.00	0.96	4.50	0.00	3.24	0.00	0.00	
3rwr	CRYSTAL STRUCTURE OF THE HUMAN XRCC4-XLF COMPLEX	A	E	8	K99	F106	4.88	KNLKDVSF	1.00	0.00	0.30	0.00	0.00	0.30	3.03	1.64
3u28	CRYSTAL STRUCTURE OF A CBF5-NOP10-GAR1 COMPLEX FROM SACCHAROMYCES2 CEREVISIAE	C	A	8	L82	F89	5.22	LGPLNEVF	1.07	0.00	0.45	0.92	0.00	0.00	4.50	0.00
3uly	CRYSTAL STRUCTURE OF BROX BRO1 DOMAIN IN COMPLEX WITH THE C-TERMINAL2 TAILS OF CHMP5	B	A	8	L207	P214	5.84	LVDEFGLP	0.00	0.00	0.00	0.63	1.97	0.00	1.81	0.87
3va1	CRYSTAL STRUCTURE OF THE MAMMALIAN MDC1 FHA DOMAIN	B	A	8	Q99	V106	4.3	QIVKPPRV	1.00	0.00	4.50	4.50	0.00	0.11	0.00	0.00
3vr0	CRYSTAL STRUCTURE OF PYROCOCCUS FURIOSUS PBAB, AN ARCHAEAL PROTEASOME2 ACTIVATOR	A	B	8	L72	V79	5.57	LIRKNSVV	1.00	0.54	0.00	0.44	0.00	4.50	0.45	0.90
3vyr	CRYSTAL STRUCTURE OF THE HYPC-HYPD COMPLEX	A	B	8	V42	A49	6.09	VIVHTGFA	0.00	0.89	0.00	2.25	0.58	0.00	1.85	0.00
3vyr	CRYSTAL STRUCTURE OF THE HYPC-HYPD COMPLEX	A	B	8	I43	I50	5.84	IVHTGFAI	0.89	0.00	2.25	0.58	0.00	1.85	0.00	1.01
3vyt	CRYSTAL STRUCTURE OF THE HYPC-HYPD-HYPE COMPLEX (FORM I INWARD)	A	B	8	I43	I50	5.69	IVHTGFAI	0.00	0.00	2.04	0.54	0.00	1.27	0.00	4.50
3zrl	IDENTIFICATION OF 2-(4-PYRIDYL)THIENOPYRIDINONES AS GSK-3BETA2 INHIBITORS	X	A	5	P199	L203	4.75	PHRLL	1.00	0.00	0.00	3.76	0.91			
3zyi	NETRING2 IN COMPLEX WITH NGL2	B	A	8	C69	L76	5.63	CSHENPYL	0.00	0.00	0.00	0.00	2.16	0.00	1.03	4.50
4a0l	STRUCTURE OF DDB1-DDB2-CUL4B-RBX1 BOUND TO A 12 BP ABASIC2 SITE CONTAINING DNA-DUPLEX	B	A	6	Q102	H107	5.72	QTSILH	0.00	1.00	4.50	1.00	1.00	0.00		
4akt	PATG MACROCYCLASE IN COMPLEX WITH PEPTIDE	C	A	5	P6	Y10	4.69	PFPAY	0.05	1.21	0.67	0.00	1.33			
4ani	STRUCTURAL BASIS FOR THE INTERMOLECULAR COMMUNICATION BETWEEN2 DNAK AND GRPE IN THE DNAK CHAPERONE SYSTEM FROM3 GEOBACILLUS KAUSTOPHILUS HTA426	G	D	8	I370	F377	5.52	IETMGGVF	0.00	1.28	1.00	4.50	0.00	0.00	0.00	0.00
4bwy	P4 PROTEIN FROM BACTERIOPHAGE PHI8 (R32)	E	F	5	R77	G81	4.24	RDLIG	0.81	1.09	0.53	4.50	0.00			
4c92	CRYSTAL STRUCTURE OF THE YEAST LSM1-7 COMPLEX	C	B	8	A32	I39	5.51	AFDSHCNI	0.00	2.50	1.24	0.58	0.00	0.00	4.50	0.00
4c9r	XENOPUS ZNRF3 ECTODOMAIN IN COMPLEX WITH XENOPUS RSPO2 FU1-2 FU2 CRYSTAL FORM I	B	A	8	C46	L53	5.42	CSKDNGCL	0.00	0.84	0.00	1.56	1.01	0.00	0.00	1.59
4cdg	CRYSTAL STRUCTURE OF THE BLOOM'S SYNDROME HELICASE BLM IN2 COMPLEX WITH NANOBODY	C	A	8	V120	F127	5.96	VADYDMGF	0.00	0.00	2.34	1.00	0.36	0.00	0.00	3.24
4d9i	CRYSTAL STRUCTURE OF HOLO DIAMINOPROPIONATE AMMONIA LYASE FROM2 ESCHERICHIA COLI	B	A	7	E387	K393	5.5	EVVWEGK	1.12	0.00	1.41	3.42	0.73	0.00	0.00	
4dih	X-RAY STRUCTURE OF THE COMPLEX BETWEEN HUMAN ALPHA THROMBIN AND2 THROMBIN BINDING APTAMER IN THE PRESENCE OF SODIUM IONS	L	H	6	L3	E8	5.74	LRPLFE	0.26	1.00	3.46	0.87	0.00	2.06		
4dii	X-RAY STRUCTURE OF THE COMPLEX BETWEEN HUMAN ALPHA THROMBIN AND2 THROMBIN BINDING APTAMER IN THE PRESENCE OF POTASSIUM IONS	L	H	6	L3	E8	5.71	LRPLFE	0.55	1.00	3.83	0.57	0.00	1.88		
4ecd	2.5 ANGSTROM RESOLUTION CRYSTAL STRUCTURE OF BIFIDOBACTERIUM LONGUM2 CHORISMATE SYNTHASE	A	B	8	T236	R243	5.18	TYVESDRR	0.00	0.00	0.00	2.47	1.00	0.97	0.00	1.00
4f88	X-RAY CRYSTAL STRUCTURE OF PLYC	H	A	8	A30	I37	5.32	AFINGVEI	0.00	0.00	2.51	1.00	0.00	0.68	0.00	0.85
4g6u	CDIA-CT/CDII TOXIN AND IMMUNITY COMPLEX FROM ESCHERICHIA COLI	A	B	8	Y244	L251	5.59	YALSGREL	1.05	0.00	0.30	0.35	0.00	1.50	4.50	0.00
4glt	CRYSTAL STRUCTURE OF GLUTATHIONE S-TRANSFERASE MFLA_2116 (TARGET EFI-2 507160) FROM METHYLOBACILLUS FLAGELLATUS KT WITH GSH BOUND	B	A	7	L60	I66	6.01	LYDSRVI	0.00	3.95	1.18	0.00	0.00	0.69	0.00	
4grw	STRUCTURE OF A COMPLEX OF HUMAN IL-23 WITH 3 NANOBODIES (LLAMA VHHS)	B	A	7	S241	F247	5.96	STPHSYF	0.00	0.00	0.35	0.00	1.07	2.48	4.50	
4hpo	CRYSTAL STRUCTURE OF RV144-ELICITED ANTIBODY CH58 IN COMPLEX WITH V22 PEPTIDE	P	H	5	H173	Y177	4.7	HALFY	2.12	0.00	0.00	1.29	1.00			
4huj	CRYSTAL STRUCTURE OF A HYPOTHETICAL PROTEIN SMA0349 FROM SINORHIZOBIUM2 MELILOTI	B	A	7	I94	K100	5.85	IDFPAFK	0.65	0.00	2.10	0.88	0.00	0.88	0.00	
4ief	COMPLEX OF PORPHYROMONAS GINGIVALIS RGPB PRO- AND MATURE DOMAINS	B	A	8	V551	G558	4.51	VEKYKKDG	0.00	1.05	0.00	0.56	1.20	0.00	4.50	0.00
4j07	CRYSTAL STRUCTURE OF A PROBABLE RIBOFLAVIN SYNTHASE, BETA CHAIN RIBH2 (6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE, DMRL SYNTHASE, LUMAZINE3 SYNTHASE) FROM	E	C	7	R84	F90	5.25	RGDTPHF	0.00	0.00	0.91	0.91	4.50	4.50	0.00	
4j07	CRYSTAL STRUCTURE OF A PROBABLE RIBOFLAVIN SYNTHASE, BETA CHAIN RIBH2 (6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE, DMRL SYNTHASE, LUMAZINE3 SYNTHASE) FROM	A	D	7	A105	P111	5.23	ALDTSTP	0.00	4.50	0.00	0.00	0.80	0.00	1.00	
4j2c	GARP-SNARE INTERACTION	B	A	6	K38	L43	4.68	KLYYGL	0.00	0.00	2.12	2.80	0.00	0.67		

PDBID	Structure Title	Chain	Partner	Residues	First	Last	Termini Distance	Hot Loop Sequence	ΔΔG1	ΔΔG2	ΔΔG3	ΔΔG4	ΔΔG5	ΔΔG6	ΔΔG7	ΔΔG8
4jqs	CRYSTAL STRUCTURE OF A HYPOTHETICAL PROTEIN (BACUNI_01602) FROM2 BACTEROIDES UNIFORMIS ATCC 8492 AT 2.30 A RESOLUTION	C	A	7	L205	V211	5.14	LGIKDVV	0.00	0.00	1.41	1.33	1.00	0.00	4.50	
4jqs	CRYSTAL STRUCTURE OF A HYPOTHETICAL PROTEIN (BACUNI_01602) FROM2 BACTEROIDES UNIFORMIS ATCC 8492 AT 2.30 A RESOLUTION	A	B	5	G206	V210	4.14	GIKDV	0.00	1.39	3.18	0.22	4.11			
4k6l	STRUCTURE OF TYPHOID TOXIN	E	D	7	Q46	Y52	5.59	QIDTSPY	1.77	0.70	0.00	0.10	0.00	3.73	0.00	
4k8y	ATOMIC RESOLUTION CRYSTAL STRUCTURES OF KALLIKREIN-RELATED PEPTIDASE 42 COMPLEXED WITH SUNFLOWER TRYPSIN INHIBITOR (SFTI-1)	B	A	8	C3	I10	5.39	CTKSIPPI	0.00	0.79	0.00	0.09	0.83	0.00	2.60	1.75
4kng	CRYSTAL STRUCTURE OF HUMAN LGR5-RSPO1-RNF43	P	F	8	L64	Q71	6.13	LERNDIRQ	1.17	0.00	1.66	0.00	2.04	0.00	0.14	1.55
4kng	CRYSTAL STRUCTURE OF HUMAN LGR5-RSPO1-RNF43	P	F	8	E65	V72	5.23	ERNDIRQV	0.00	1.66	0.00	2.04	0.00	0.14	1.55	0.00
4kt1	COMPLEX OF R-SPONDIN 1 WITH LGR4 EXTRACELLULAR DOMAIN	E	A	8	A104	C111	6.09	ACFSHNFC	0.00	0.00	1.70	0.11	0.00	0.69	3.82	0.00
4m1u	THE CRYSTAL STRUCTURE OF STX2 AND A DISACCHARIDE LIGAND	F	E	8	K12	F19	5.67	KYNEDDTF	0.74	2.20	0.74	0.27	0.00	1.23	0.00	0.00
4mur	CRYSTAL STRUCTURE OF VANCOMYCIN RESISTANCE D,D-DIPEPTIDASE/D,D-2 PENTAPEPTIDASE VANXYC D59S MUTANT	B	A	8	A25	V32	5.68	APFSDHDV	0.00	0.00	0.00	1.00	1.37	2.28	1.00	0.00
4mx3	CRYSTAL STRUCTURE OF PKA RIALPHA HOMODIMER	B	A	6	K118	M123	5.95	KDYKTM	0.68	4.50	1.88	0.00	0.00	0.71		
4nek	PUTATIVE ENOYL-COA HYDRATASE/CARNITHINE RACEMASE FROM MAGNETOSPIRILLUM2 MAGNETICUM AMB-1	A	F	8	F128	P135	5.76	FINLGIVP	0.00	1.36	0.81	0.10	0.00	0.21	1.29	4.50
4ofi	CRYSTAL STRUCTURE OF DUF (KIRRE) D1	A	B	5	T115	F119	4.92	TKDDF	4.29	0.00	0.00	2.73	0.72			
4oig	DENGUE VIRUS NON-STRUCTURAL PROTEIN NS1	B	A	7	K227	S233	6.06	KSHTLWS	4.50	0.11	0.78	0.00	1.00	1.00	0.44	
4olp	LIGAND-FREE STRUCTURE OF THE GRPU MICROCOMPARTMENT SHELL PROTEIN FROM2 PECTOBACTERIUM WASABIAE	C	D	6	L42	L47	5.58	LPVPDL	0.00	1.00	0.11	4.50	1.14	0.00		
4oww	STRUCTURAL BASIS OF SOSS1 IN COMPLEX WITH A 35NT SSDNA	C	A	8	Q86	L93	4.93	QDSAFGNL	0.86	0.00	0.76	0.00	1.78	0.00	2.28	0.00

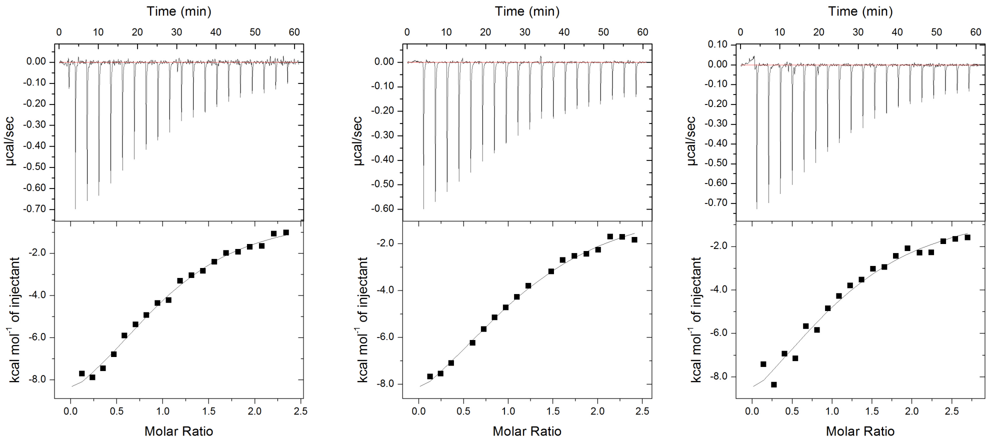


Figure 5.1 Binding affinities for **lST2** measured by ITC. **lST2** has an average K_d of 18.20 ± 3.38 μM for three independent trials.

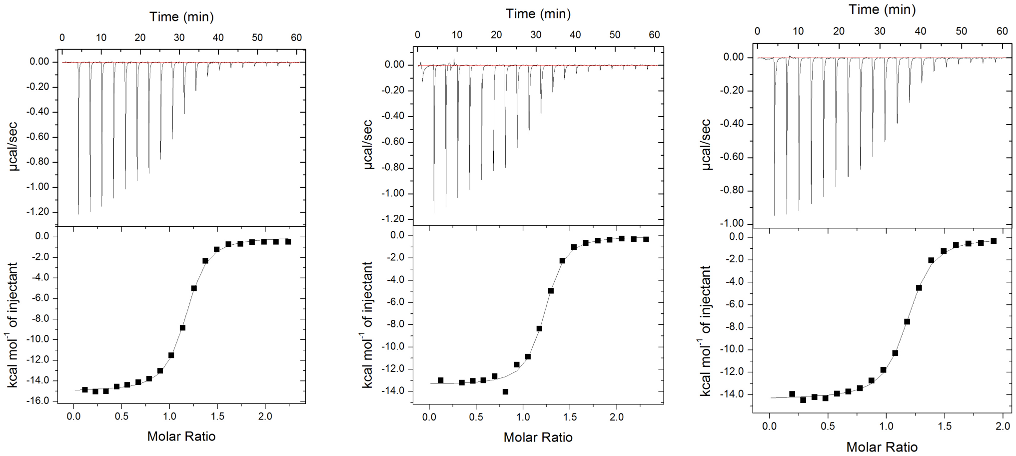


Figure 5.2 Binding affinities for **cST2-*ox*** measured by ITC. **cST2-*ox*** has an average K_d of 0.33 ± 0.01 μM for three independent trials.

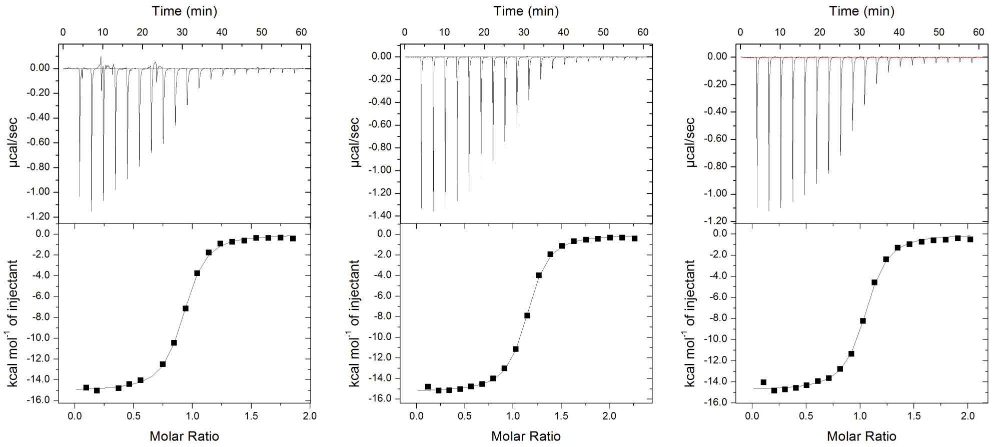


Figure 5.3 Binding affinities for **cST2-*mx*** measured by ITC. **cST2-*mx*** has an average K_d of 0.37 ± 0.02 µM for three independent trials.

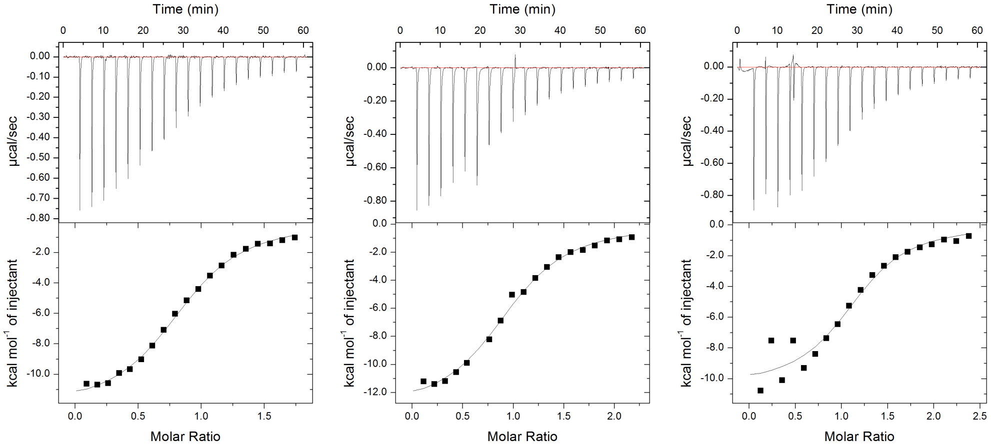


Figure 5.4 Binding affinities for **cST2-*px*** measured by ITC. **cST2-*px*** has an average K_d of 3.67 ± 0.17 µM for three independent trials.

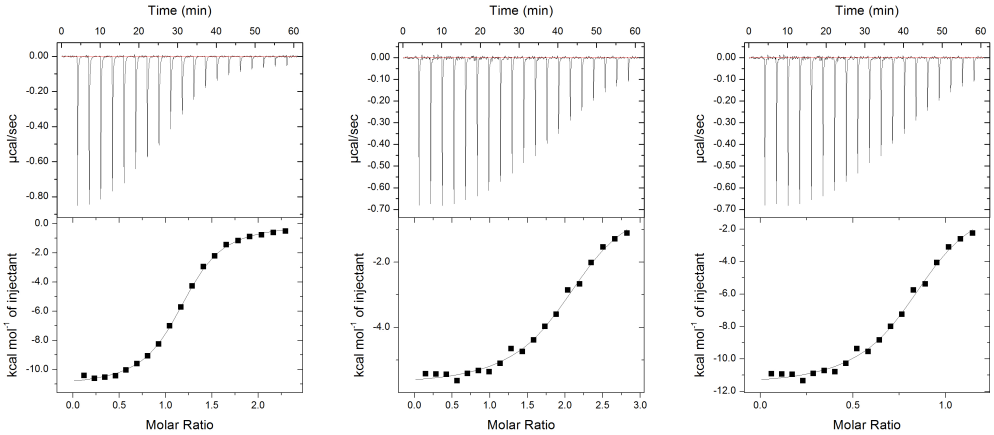


Figure 5.5 Binding affinities for **cST2-*pyr*** were measured by ITC. **cST2-*pyr*** has an average K_d of 1.53 ± 0.05 μM for three independent trials.

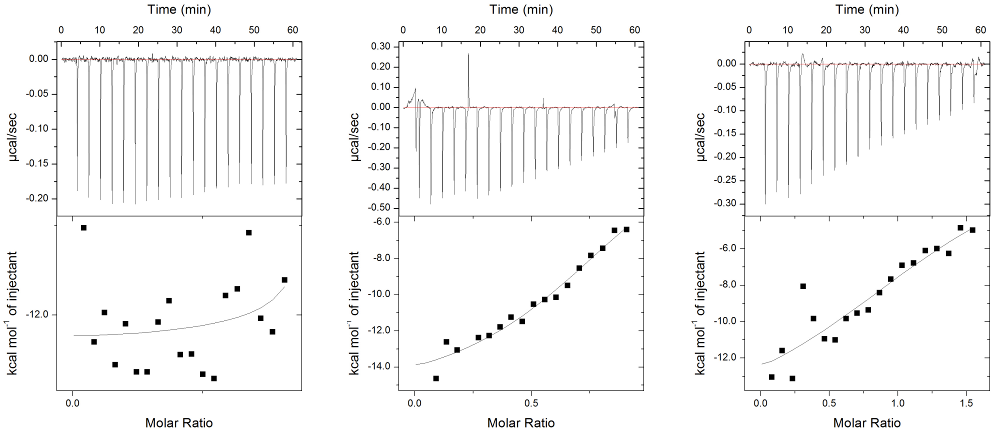


Figure 5.6 Binding affinities for **cST2-*dbz*** measured by ITC. Due to the weak binding of the peptide, no K_d was calculated from the poor curve fit.

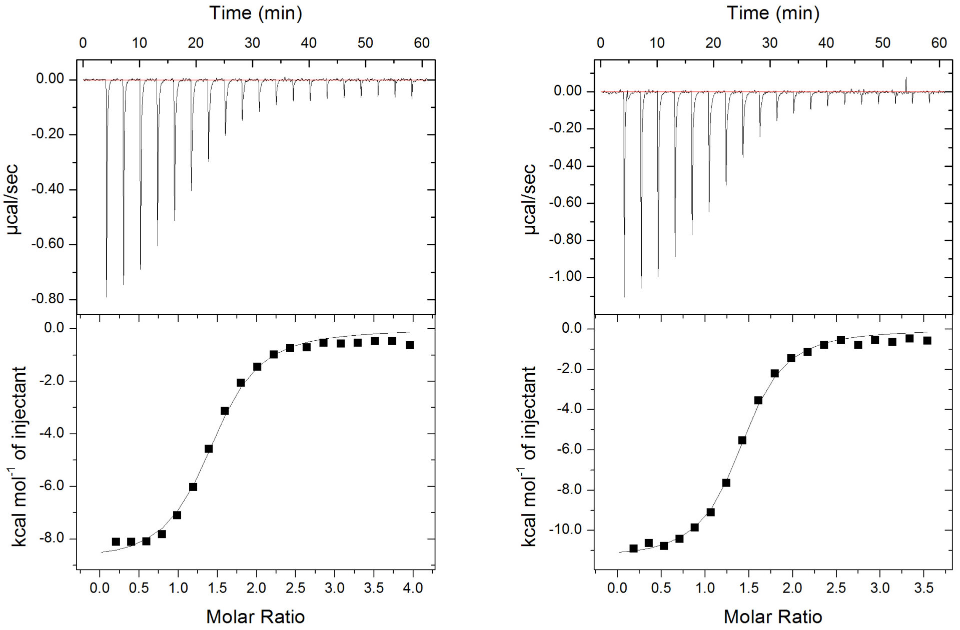


Figure 5.7 Binding affinities for **cST2-W3A-*ox*** measured by ITC. For this peptide, measured in duplicate due to inconsistencies with concentration measurement, the K_d was found to be 1.70 ± 0.28 μM.

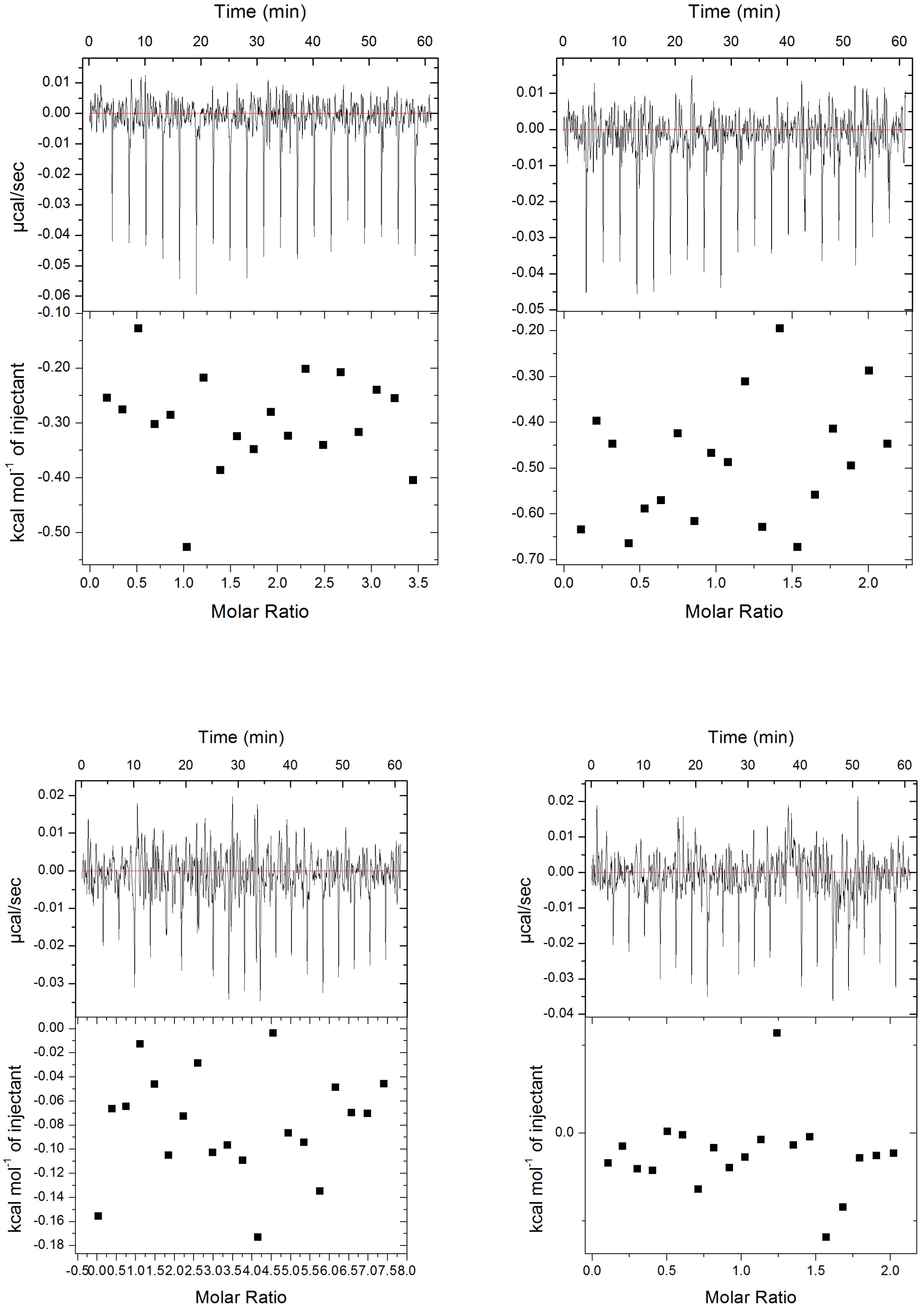


Figure 5.8 **cST2-N7A-*ox*** (top) or **cST2-F9A-*ox*** (bottom) had no observable binding to Eps15-EH2 by ITC.

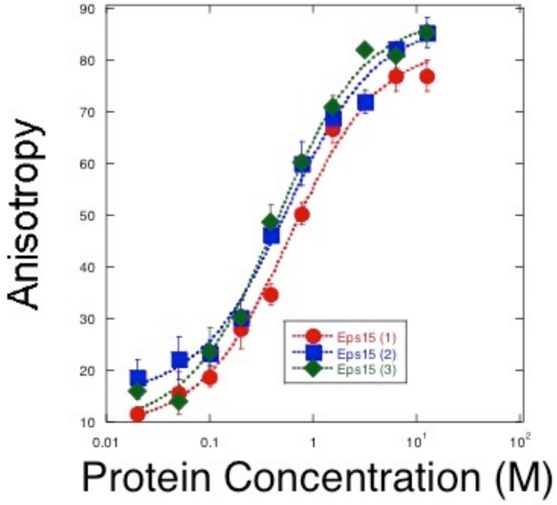


Figure 5.9 Direct Binding of **Flu-cST2-*ox*** to Eps15-EH2 measured by fluorescence polarization yields a calculated K_d of 0.51 ± 0.06 μM.

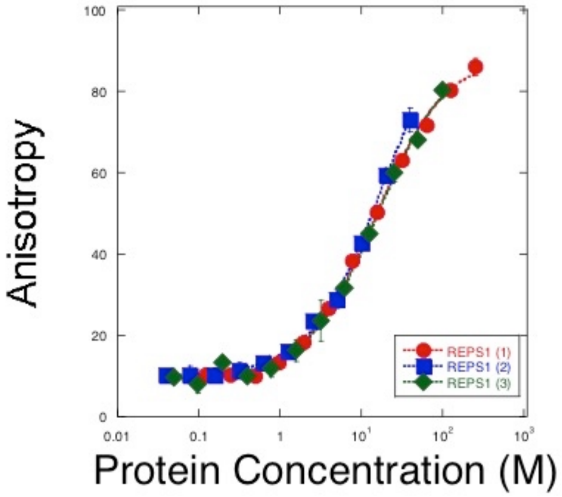


Figure 5.10 Direct Binding of **Flu-cST2-*ox*** to REPS1-EH measured by fluorescence polarization yields a calculated K_d of 15.8 ± 1.1 μM.

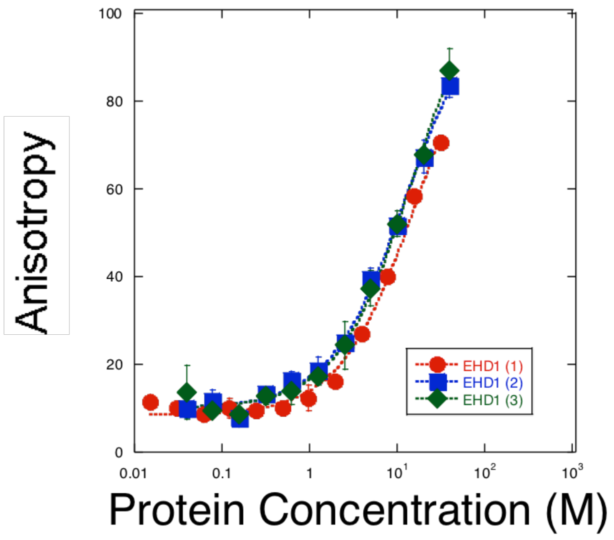


Figure 5.11 Direct binding of **Flu-cST2-*ox*** to EHD1-EH was measured by fluorescence anisotropy yielding a K_d of 14.8 ± 1.7 μM.

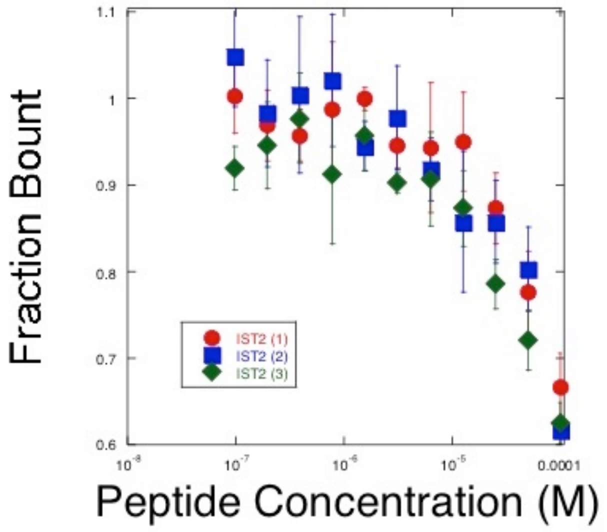


Figure 5.12 A FP competition assay was performed in triplicate to calculate an IC_{50} for **IST2**, but inhibition was not measurable at concentrations within the range tested.

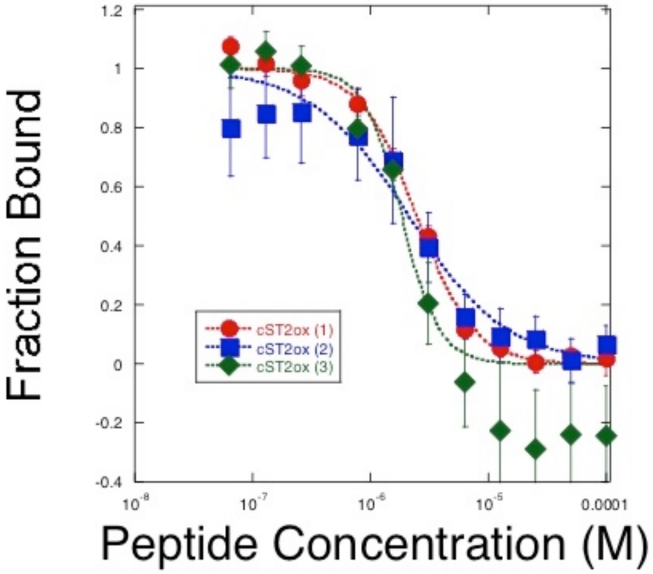


Figure 5.13 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-*ox*** that was calculated to be 2.15 ± 0.26 μM.

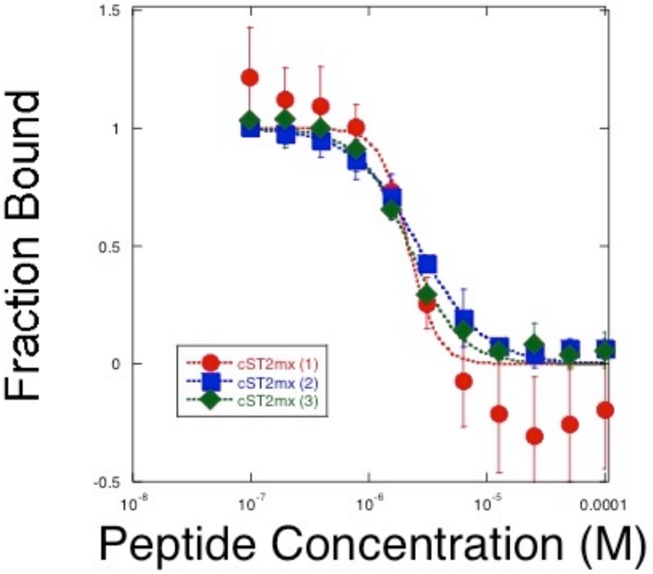


Figure 5.14 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-*mx*** that was calculated to be 2.33 ± 0.22 μM.

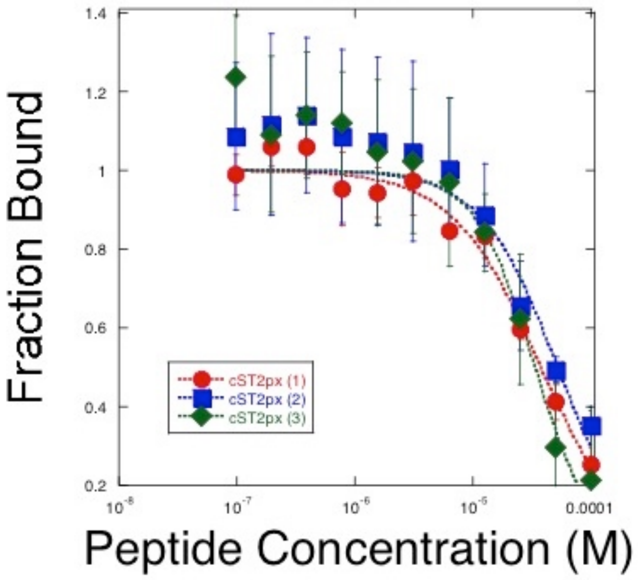


Figure 5.15 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-*px*** that was calculated to be 41.3 ± 7.9 μM.

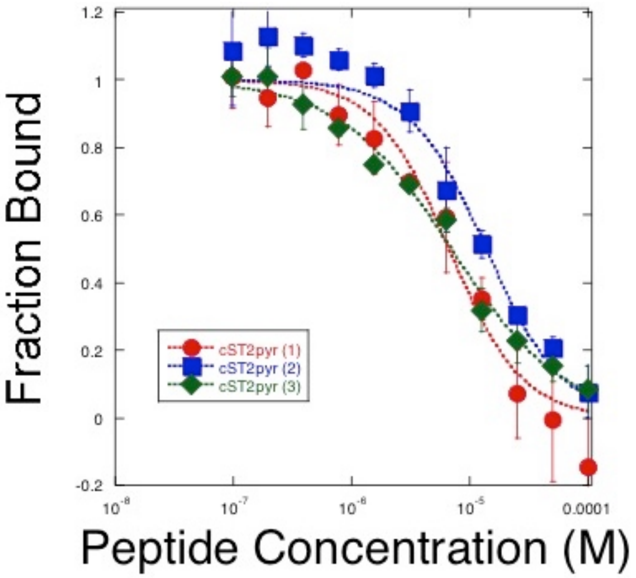


Figure 5.16 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-*pyr*** that was calculated to be 9.09 ± 3.26 μM.

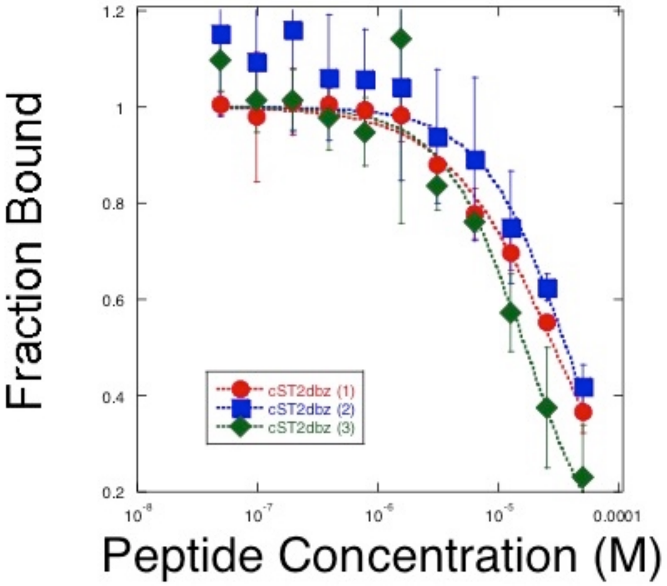


Figure 5.17 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-*dbz*** that was calculated to be 27.5 ± 8.3 μM.

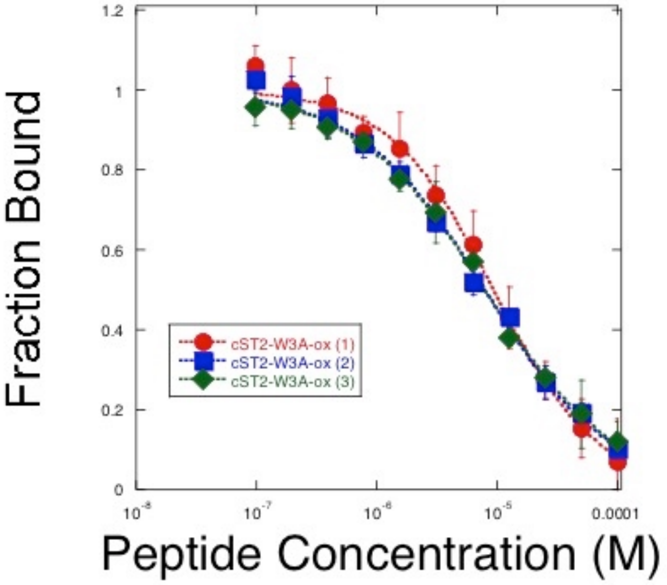


Figure 5.18 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-W3A-*ox*** that was calculated to be 8.34 ± 0.67 μM indicating the significance of the tryptophan residue for peptide binding.

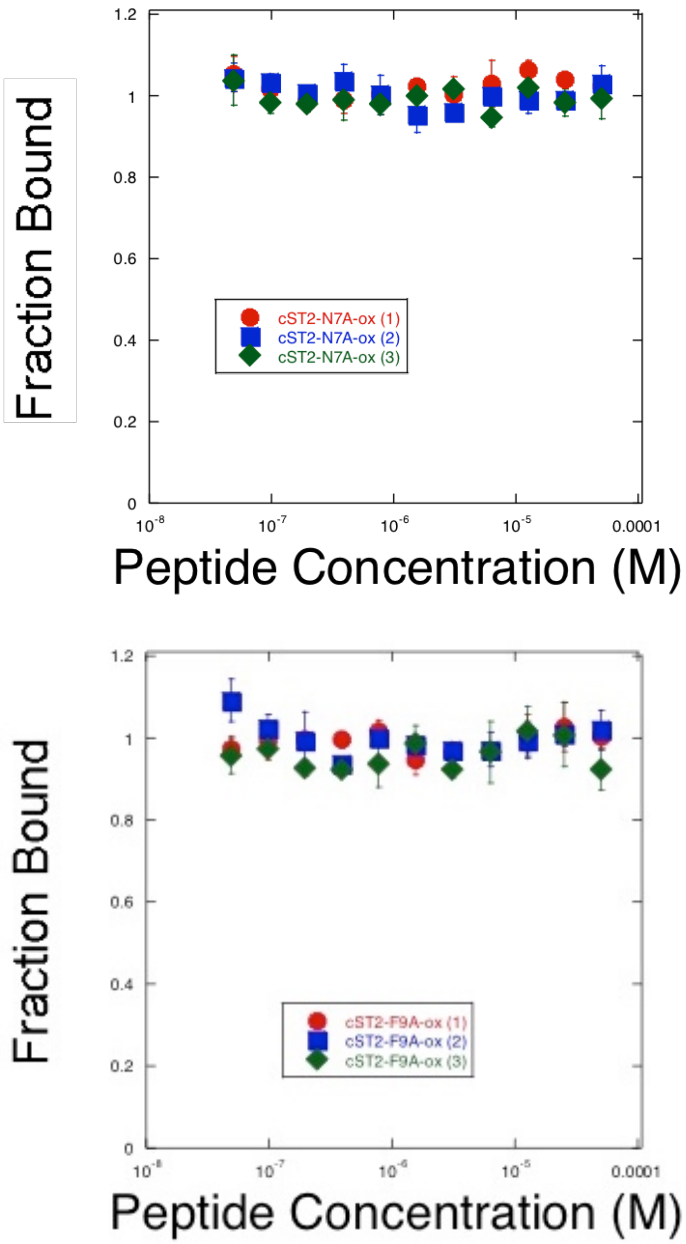


Figure 5.19 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST2-N7A-*ox*** and **cST2-F9A-*ox***. The lack of binding indicates that the proper epitope for Eps14-EH2 recognition resides in the canonical NPF motif.

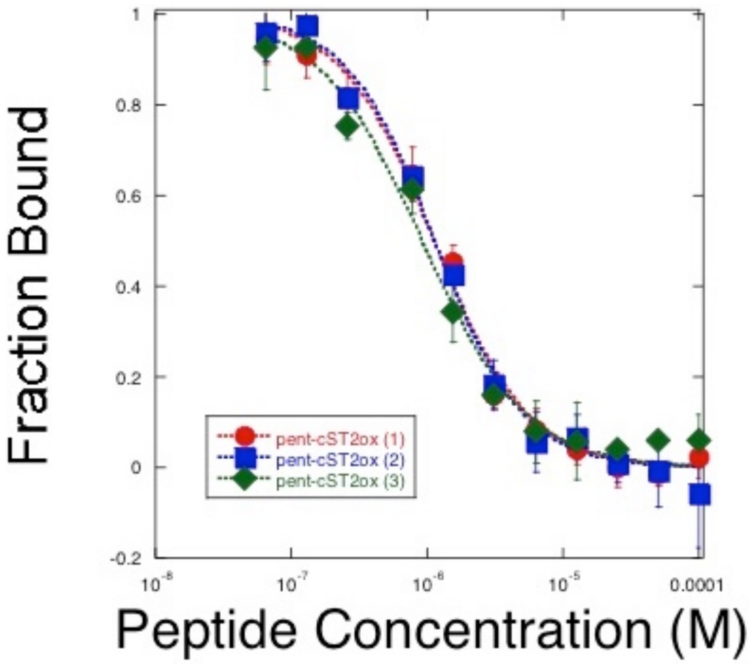


Figure 5.20 A FP competition assay was performed in triplicate to calculate an IC_{50} for **pent-cST2-*ox*** that was calculated to be 1.06 ± 0.10 μM.

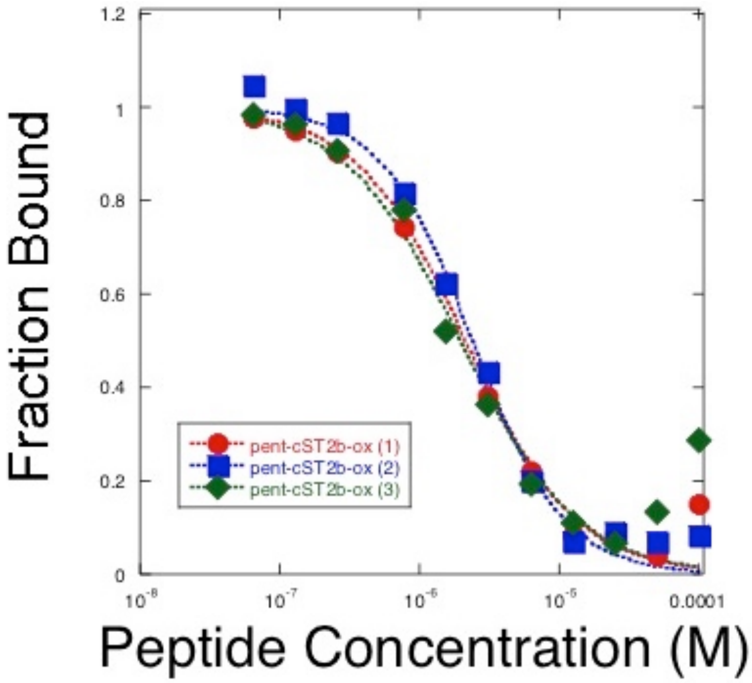


Figure 5.21 A FP competition assay was performed in triplicate to calculate an IC_{50} for **pent-cST2b-*ox*** that was calculated to be 2.16 ± 0.18 μM

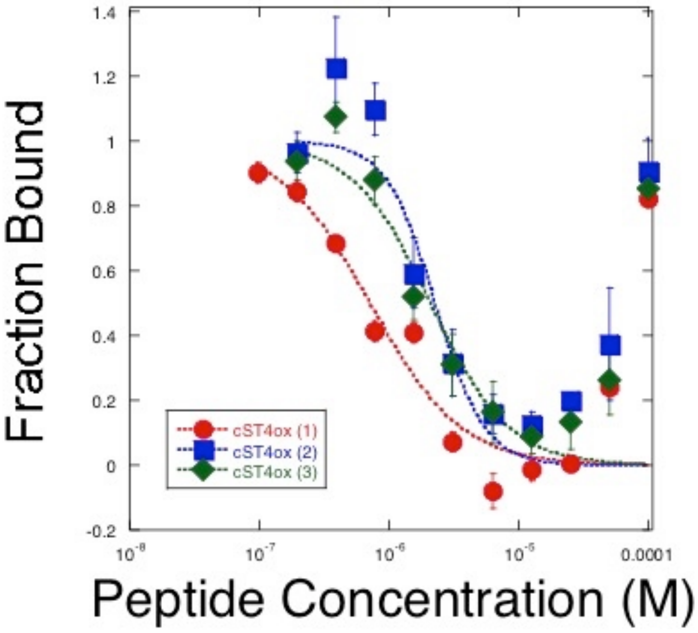


Figure 5.22 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST4-*ox*** that was calculated to be 1.69 ± 0.70 µM.

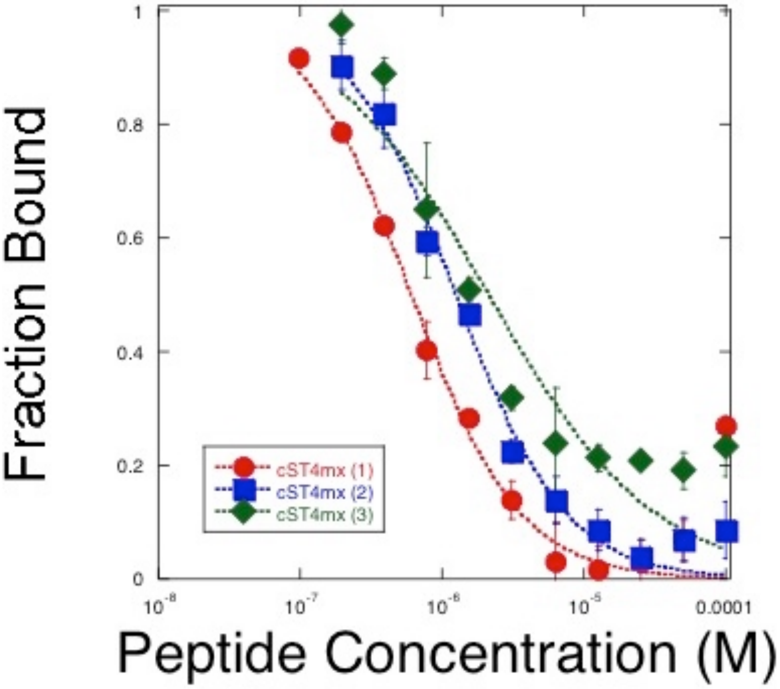


Figure 5.23 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST4-*mx*** that was calculated to be 1.33 ± 0.62 µM.

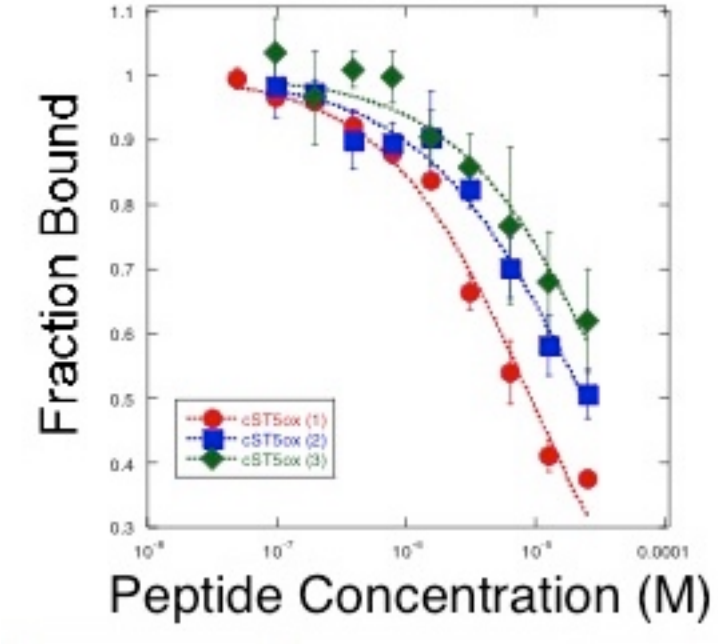


Figure 5.24 A FP competition assay was performed in triplicate to calculate an IC$_{50}$ for **cST5-*ox*** that was calculated to be 24.4 ± 12.7 μM.

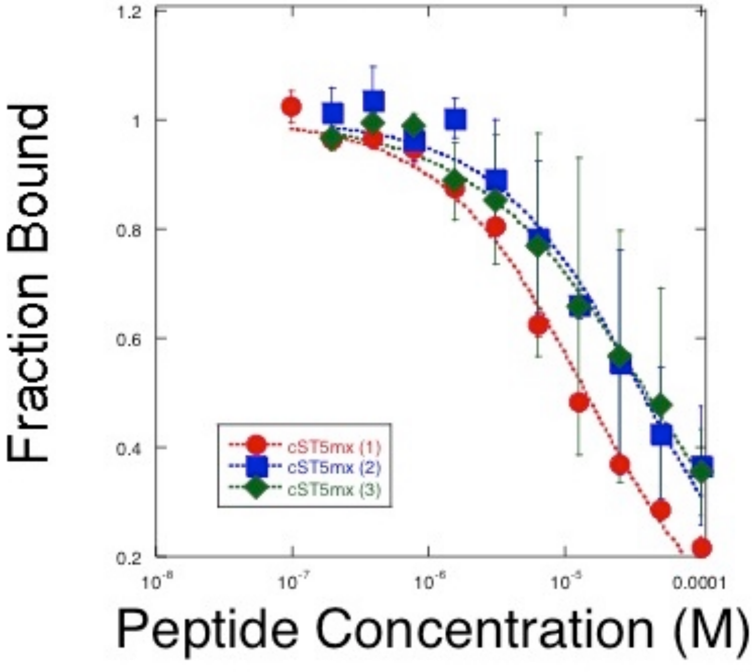


Figure 5.25 A FP competition assay was performed in triplicate to calculate an IC$_{50}$ for **cST5-*mx*** that was calculated to be 30.1 ± 11.3 μM.

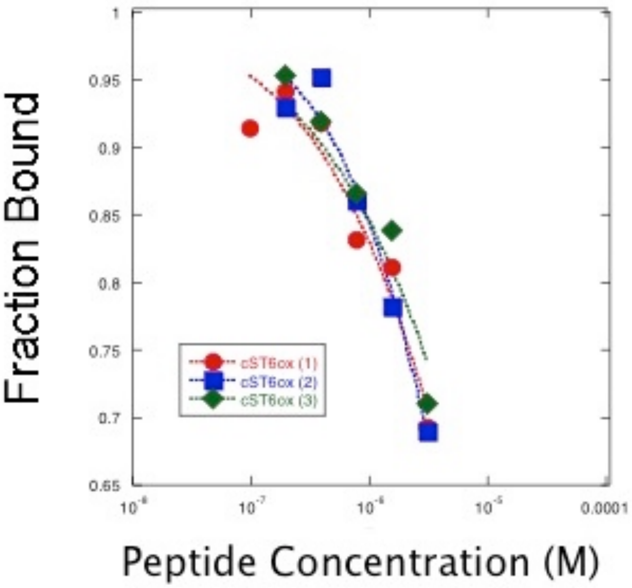


Figure 5.26 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST6-*ox*** that was calculated to be > 25 μM.

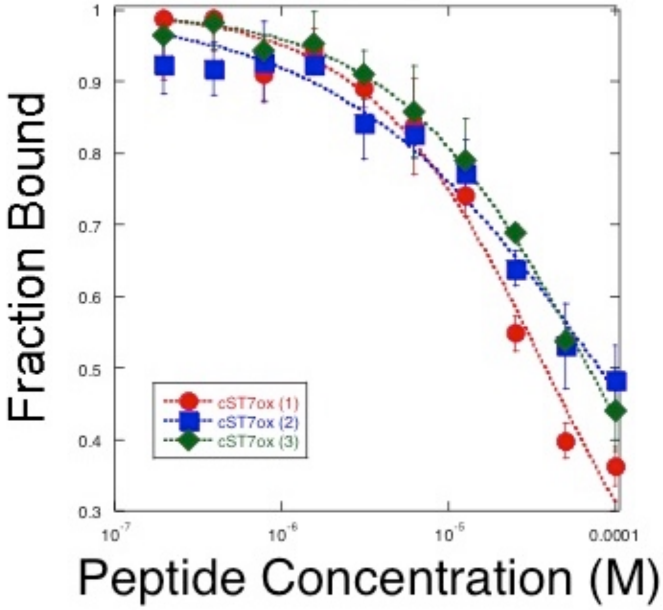


Figure 5.27 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST7-*ox*** that was calculated to be 68.1 ± 24.3 μM.

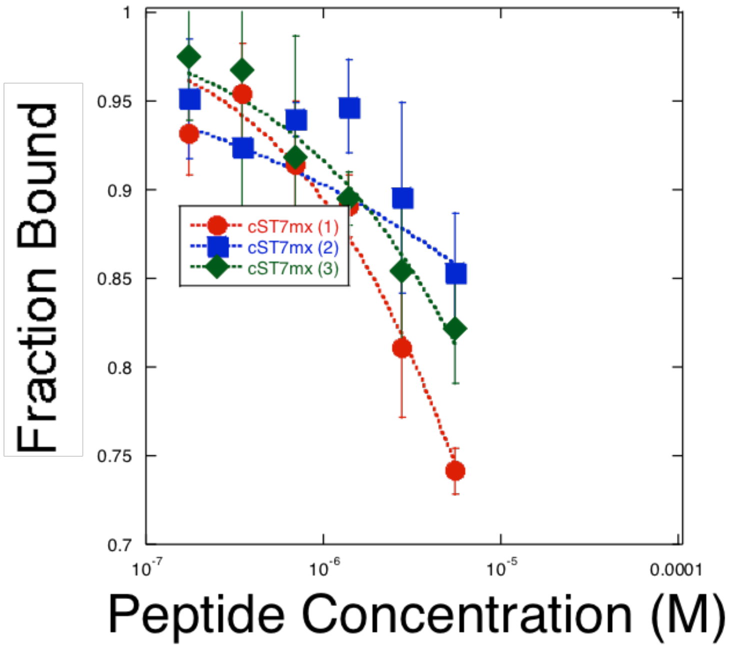


Figure 5.28 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST7-*mx*** that was calculated to be > 30 µM.

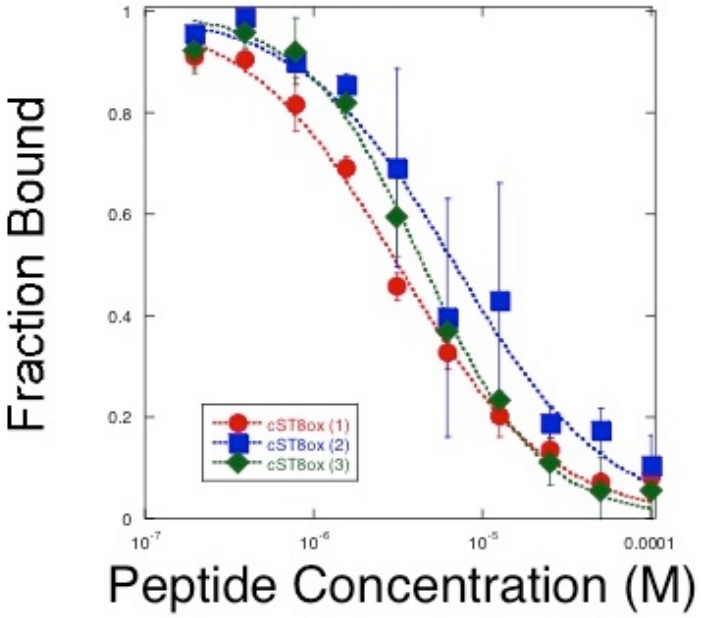


Figure 5.29 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST8-*ox*** that was calculated to be 4.83 ± 1.53 µM.

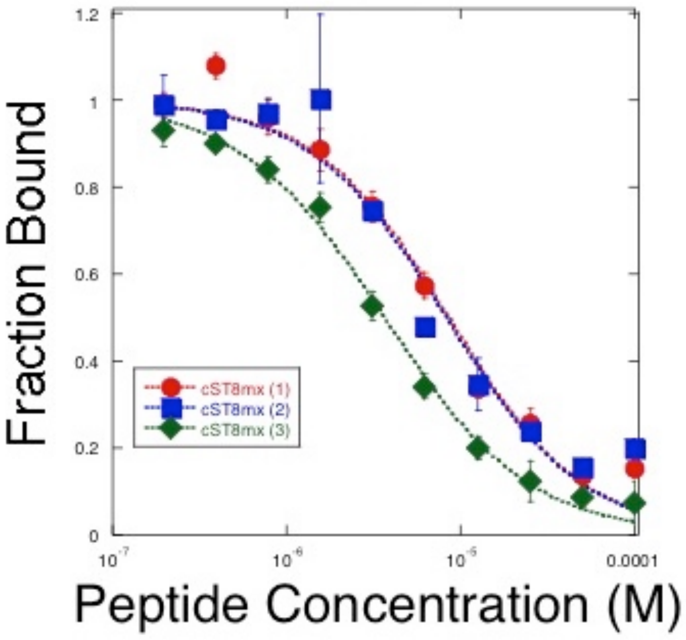


Figure 5.30 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST8-*mx*** that was calculated to be 6.82 ± 2.24 μM.

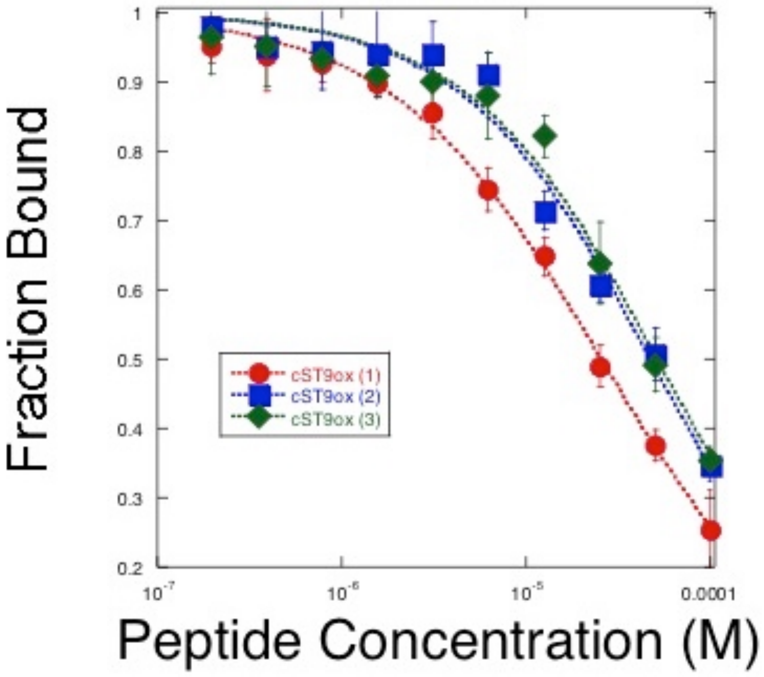


Figure 5.31 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST9-*ox*** that was calculated to be 41.3 ± 11.3 μM.

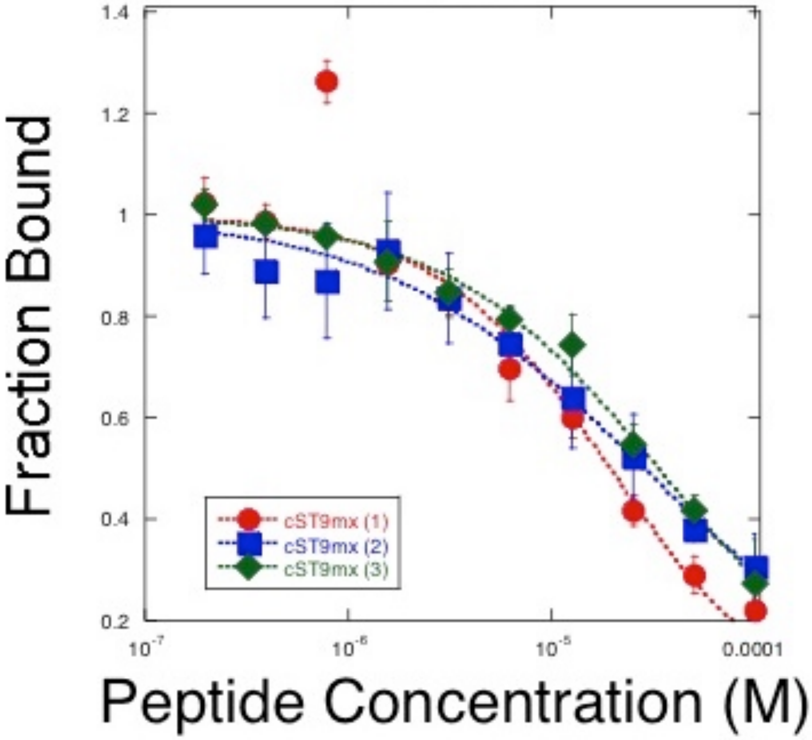
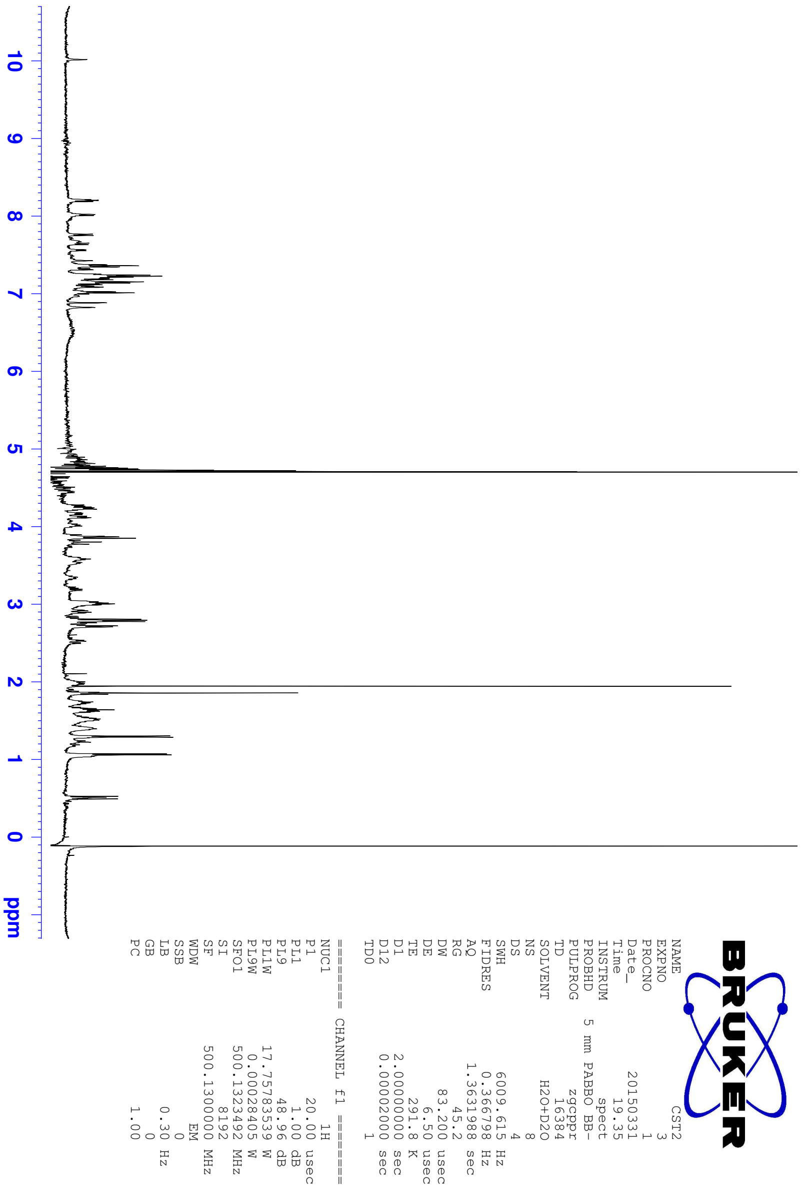


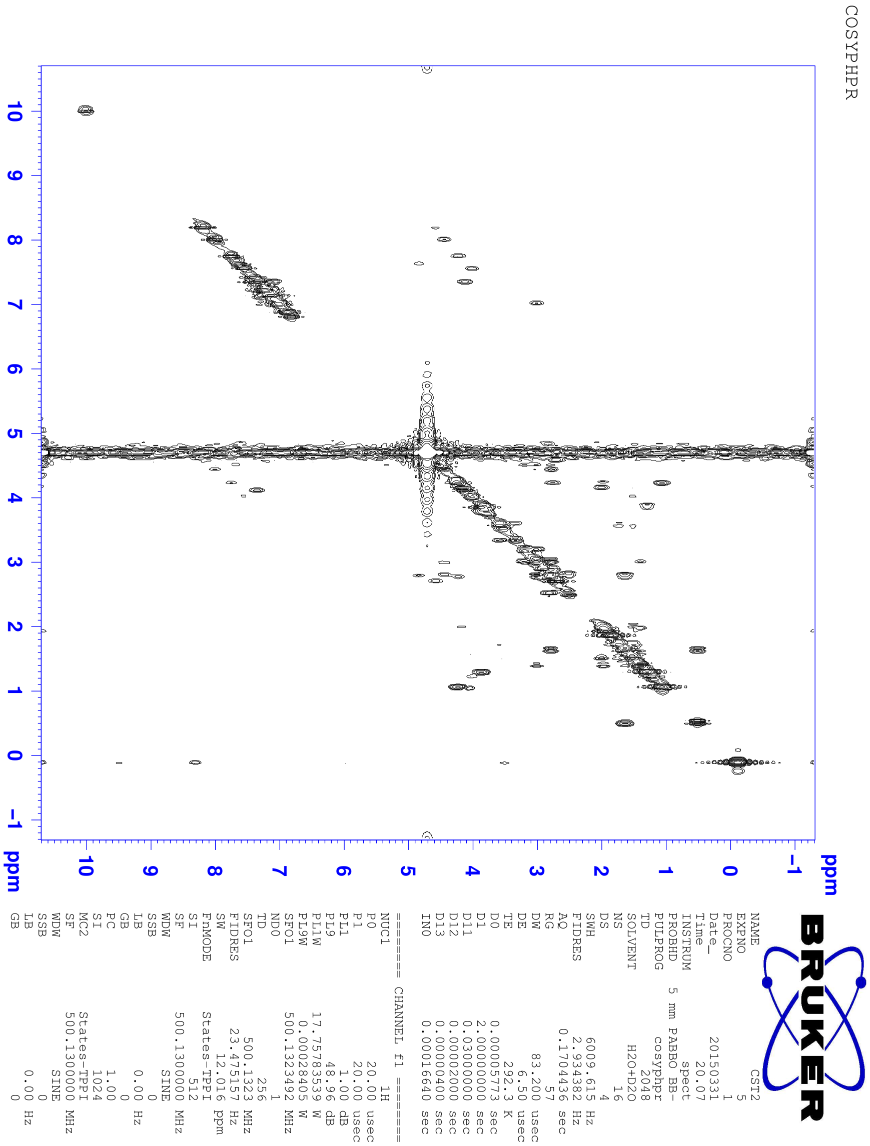
Figure 5.32 A FP competition assay was performed in triplicate to calculate an IC_{50} for **cST9-*mx*** that was calculated to be 27.0 ± 8.3 μM.



```
NAME                CST2
EXPNO                  3
PROCNO                 1
Date_           20150331
Time               19.35
INSTRUM            spect
PROBHD   5 mm PABBO BB-
PULPROG           zgcppr
TD                 16384
SOLVENT          H2O+D2O
NS                     8
DS                     4
SWH             6009.615 Hz
FIDRES          0.366798 Hz
AQ             1.3631988 sec
RG                  45.2
DW                83.200 usec
DE                  6.50 usec
TE                 291.8 K
D1            2.00000000 sec
D12           0.00002000 sec
TD0                    1

======== CHANNEL f1 ========
NUC1                  1H
P1                 20.00 usec
PL1                 1.00 dB
PL9                48.96 dB
PL1W         17.75783539 W
PL9W          0.00028405 W
SFO1         500.1323492 MHz
SI                  8192
SF           500.1300000 MHz
WDW                   EM
SSB                    0
LB                  0.30 Hz
GB                     0
PC                  1.00
```

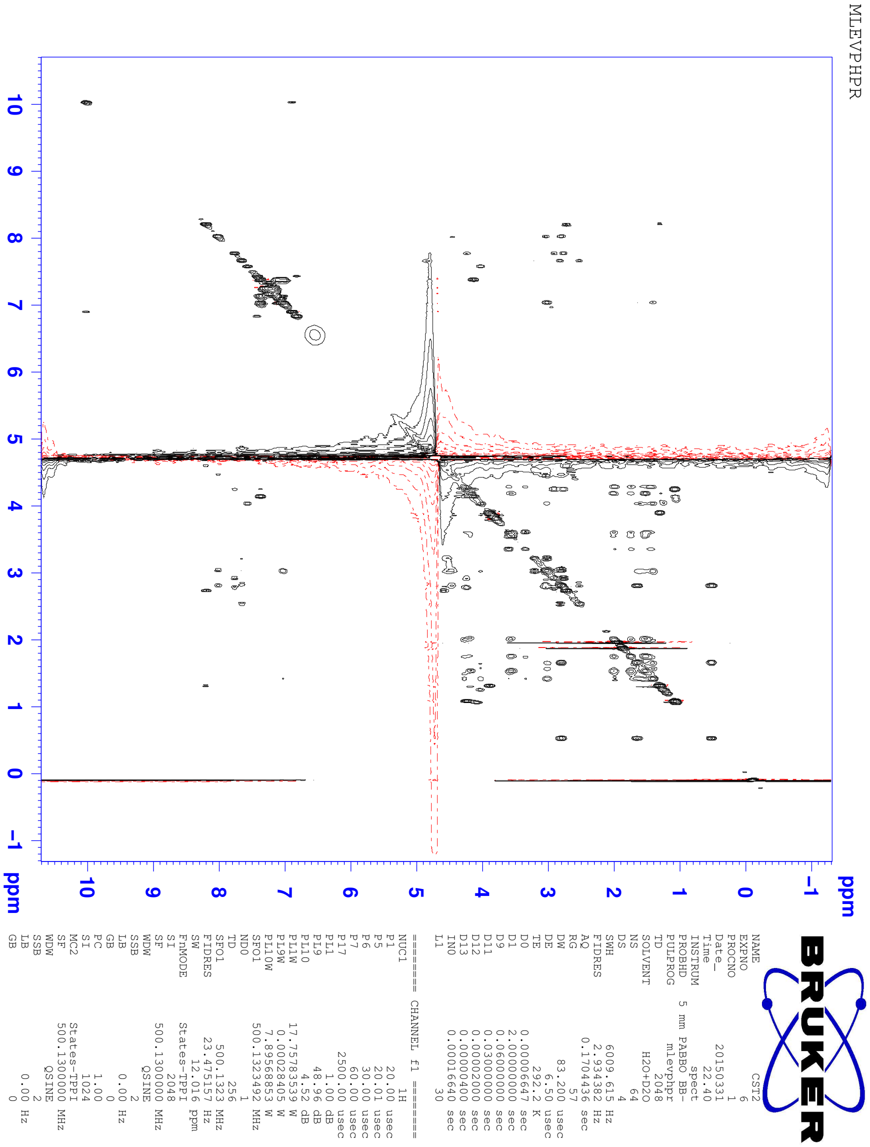
Figure 5.33 1D-^{1}H spectra of **cST2-*ox*** in 9:1 H_2O:D_2O with the zgcppr pulse program on the Tufts Chemistry 500 MHz NMR.



```
NAME                 CST2
EXPNO                   5
PROCNO                  1
Date_            20150331
Time                20.07
INSTRUM             spect
PROBHD   5 mm PABBO BB-
PULPROG          cosyphpr
TD                   2048
SOLVENT           H2O+D2O
NS                     16
DS                      4
SWH              6009.615 Hz
FIDRES           2.934382 Hz
AQ              0.1704436 sec
RG                     57
DW                 83.200 usec
DE                   6.50 usec
TE                  292.3 K
D0             0.00005773 sec
D1             2.00000000 sec
D11            0.03000000 sec
D12            0.00002000 sec
D13            0.00000400 sec
IN0            0.00016640 sec

======== CHANNEL f1 ========
NUC1                   1H
P0                  20.00 usec
P1                  20.00 usec
PL1                  1.00 dB
PL9                 48.96 dB
PL1W          17.75783539 W
PL9W           0.00028405 W
SFO1          500.1323492 MHz
ND0                     1
TD                    256
SFO1             500.1323 MHz
FIDRES          23.475157 Hz
SW                 12.016 ppm
FnMODE        States-TPPI
SI                    512
SF            500.1300000 MHz
WDW                  SINE
SSB                     0
LB                   0.00 Hz
GB                      0
PC                   1.00
SI                   1024
MC2           States-TPPI
SF            500.1300000 MHz
WDW                  SINE
SSB                     0
LB                   0.00 Hz
GB                      0
```

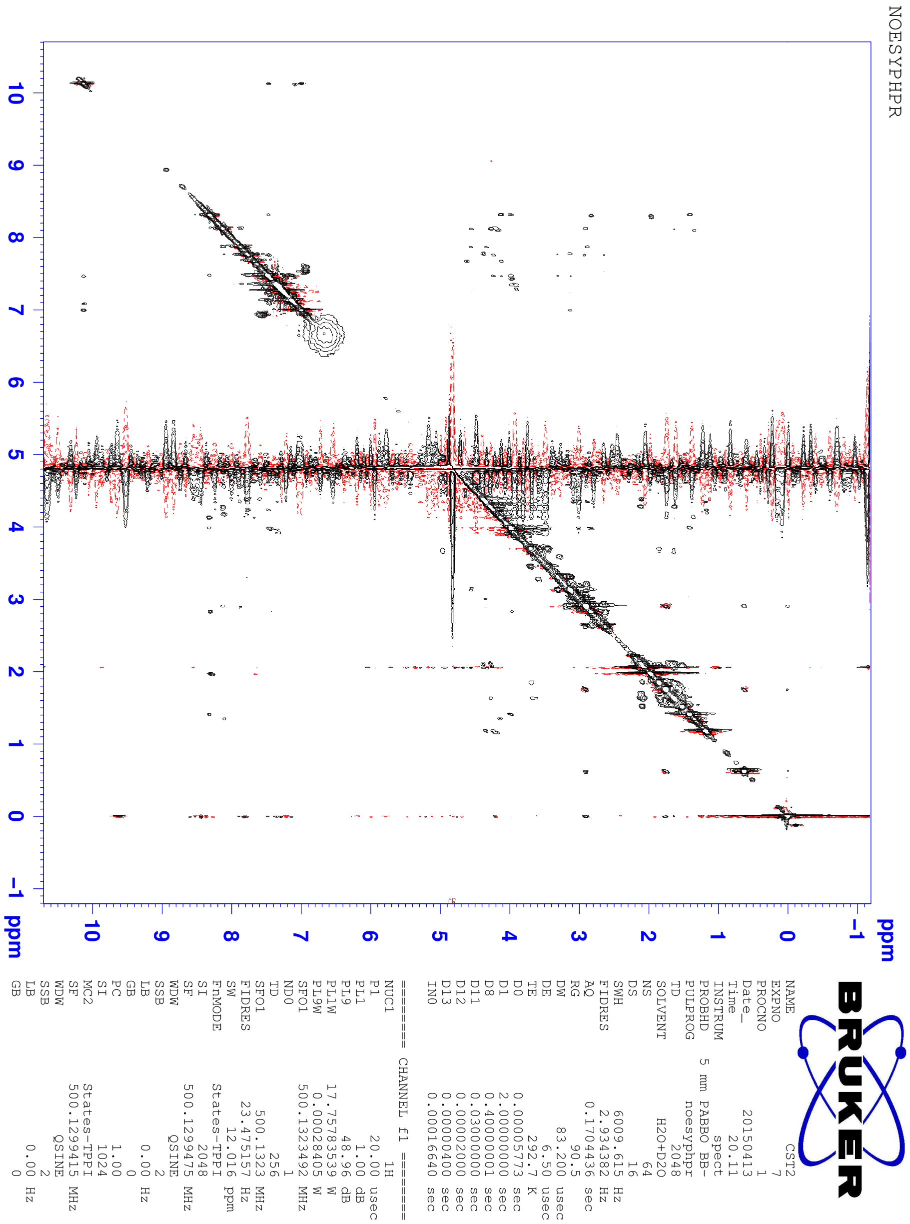
Figure 5.34 COSY spectra of **cST2-*ox*** in 9:1 H_2O:D_2O with the cosyphpr pulse program on the Tufts Chemistry 500 MHz NMR.



```
NAME                CST2
EXPNO                  6
PROCNO                 1
Date_           20150331
Time               22.40
INSTRUM            spect
PROBHD   5 mm PABBO BB-
PULPROG         mlevphpr
TD                  2048
SOLVENT          H2O+D2O
NS                    64
DS                     4
SWH             6009.615 Hz
FIDRES          2.934382 Hz
AQ             0.1704436 sec
RG                    57
DW                83.200 usec
DE                  6.50 usec
TE                 292.2 K
D0            0.00006647 sec
D1            2.00000000 sec
D9            0.06000000 sec
D11           0.03000000 sec
D12           0.00002000 sec
D13           0.00000400 sec
IN0           0.00016640 sec
L1                    30

======== CHANNEL f1 ========
NUC1                  1H
P1                 20.00 usec
P5                 20.01 usec
P6                 30.00 usec
P7                 60.00 usec
P17              2500.00 usec
PL1                 1.00 dB
PL9                48.96 dB
PL10                4.52 dB
PL1W         17.75783539 W
PL9W          0.00028405 W
PL10W         7.89568853 W
SFO1         500.1323492 MHz
ND0                    1
TD                   256
SFO1            500.1323 MHz
FIDRES         23.475157 Hz
SW                12.016 ppm
FnMODE       States-TPPI
SI                  2048
SF           500.1300000 MHz
WDW                QSINE
SSB                    2
LB                  0.00 Hz
GB                     0
PC                  1.00
SI                  1024
MC2          States-TPPI
SF           500.1300000 MHz
WDW                QSINE
SSB                    2
LB                  0.00 Hz
GB                     0
```

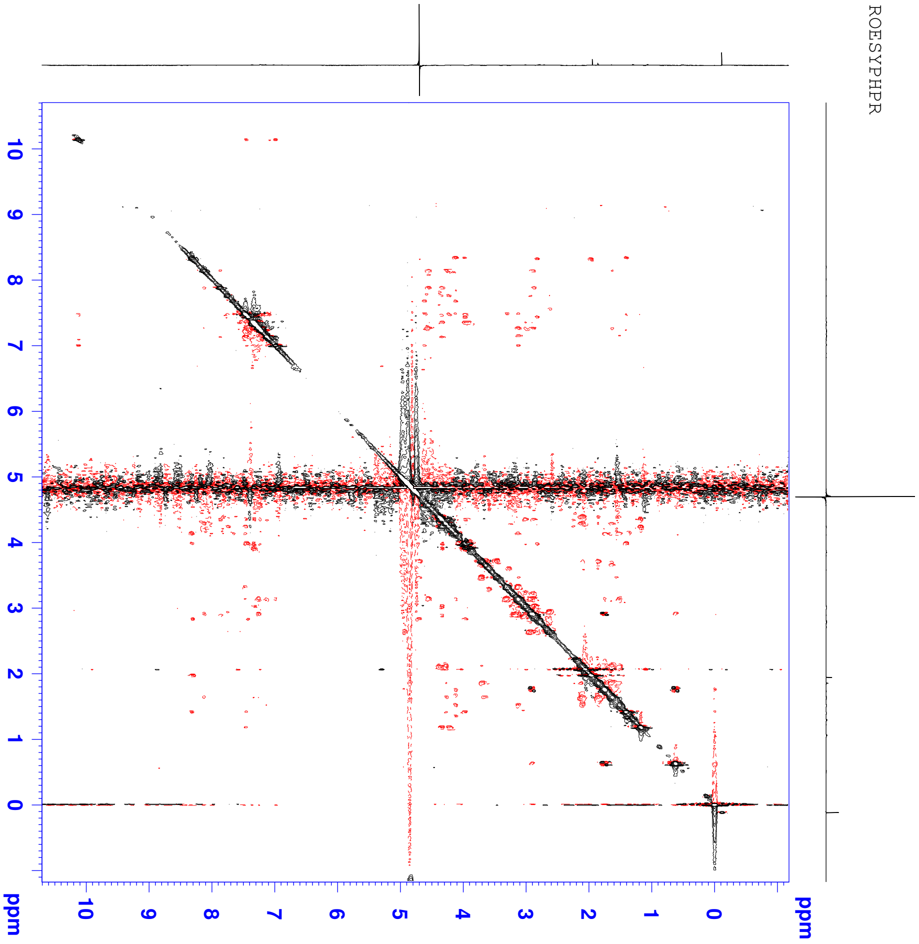
Figure 5.35 TOCSY spectra of **cST2-*ox*** in 9:1 H_2O:D_2O with the mlevphpr pulse program on the Tufts Chemistry 500 MHz NMR.



```
NAME                CST2
EXPNO                  7
PROCNO                 1
Date_           20150413
Time               20.11
INSTRUM            spect
PROBHD   5 mm PABBO BB-
PULPROG        noesyphpr
TD                  2048
SOLVENT          H2O+D2O
NS                    64
DS                    16
SWH             6009.615 Hz
FIDRES          2.934382 Hz
AQ             0.1704436 sec
RG                  90.5
DW                83.200 usec
DE                  6.50 usec
TE                 292.7 K
D0            0.00005773 sec
D1            2.00000000 sec
D8            0.40000001 sec
D11           0.03000000 sec
D12           0.00002000 sec
D13           0.00000400 sec
IN0           0.00016640 sec

======== CHANNEL f1 ========
NUC1                  1H
P1                 20.00 usec
PL1                 1.00 dB
PL9                48.96 dB
PL1W         17.75783539 W
PL9W          0.00028405 W
SFO1         500.1323492 MHz
ND0                    1
TD                   256
SFO1            500.1323 MHz
FIDRES         23.475157 Hz
SW                12.016 ppm
FnMODE       States-TPPI
SI                  2048
SF           500.1299475 MHz
WDW                QSINE
SSB                    2
LB                  0.00 Hz
GB                     0
PC                  1.00
SI                  1024
MC2          States-TPPI
SF           500.1299415 MHz
WDW                QSINE
SSB                    2
LB                  0.00 Hz
GB                     0
```

Figure 5.36 NOESY spectra of **cST2-*ox*** in 9:1 H_2O:D_2O with the noesyphpr pulse program on the Tufts Chemistry 500 MHz NMR.



```
NAME                CST2
EXPNO                  8
PROCNO                 1
Date_           20150514
Time               18.48
INSTRUM            spect
PROBHD   5 mm PABBO BB-
PULPROG        roesyphpr
TD                  2048
SOLVENT          H2O+D2O
NS                    64
DS                     4
SWH             6009.615 Hz
FIDRES          2.934382 Hz
AQ             0.1704436 sec
RG                   203
DW                83.200 usec
DE                  6.50 usec
TE                 290.5 K
D0            0.00006647 sec
D1            2.00000000 sec
D11           0.03000000 sec
D12           0.00002000 sec
D13           0.00000400 sec
IN0           0.00016640 sec

======== CHANNEL f1 ========
NUC1                  1H
P1                 20.00 usec
P15            400000.00 usec
PL1                 1.00 dB
PL9                48.96 dB
PL11               16.92 dB
PL1W         17.75783539 W
PL9W          0.00028405 W
PL11W         0.45434943 W
SFO1         500.1323492 MHz
ND0                    1
TD                   256
SFO1            500.1323 MHz
FIDRES         23.475157 Hz
SW                12.016 ppm
FnMODE       States-TPPI
SI                  1024
SF           500.1299327 MHz
WDW                QSINE
SSB                    2
LB                  0.00 Hz
GB                     0
PC                  1.00
SI                  1024
MC2          States-TPPI
SF           500.1299365 MHz
WDW                QSINE
SSB                    2
LB                  0.00 Hz
GB                     0
```

Figure 5.37 ROESY spectra of **cST2-*ox*** in 9:1 H_2O:D_2O with the roesyphpr pulse program on the Tufts Chemistry 500 MHz NMR.

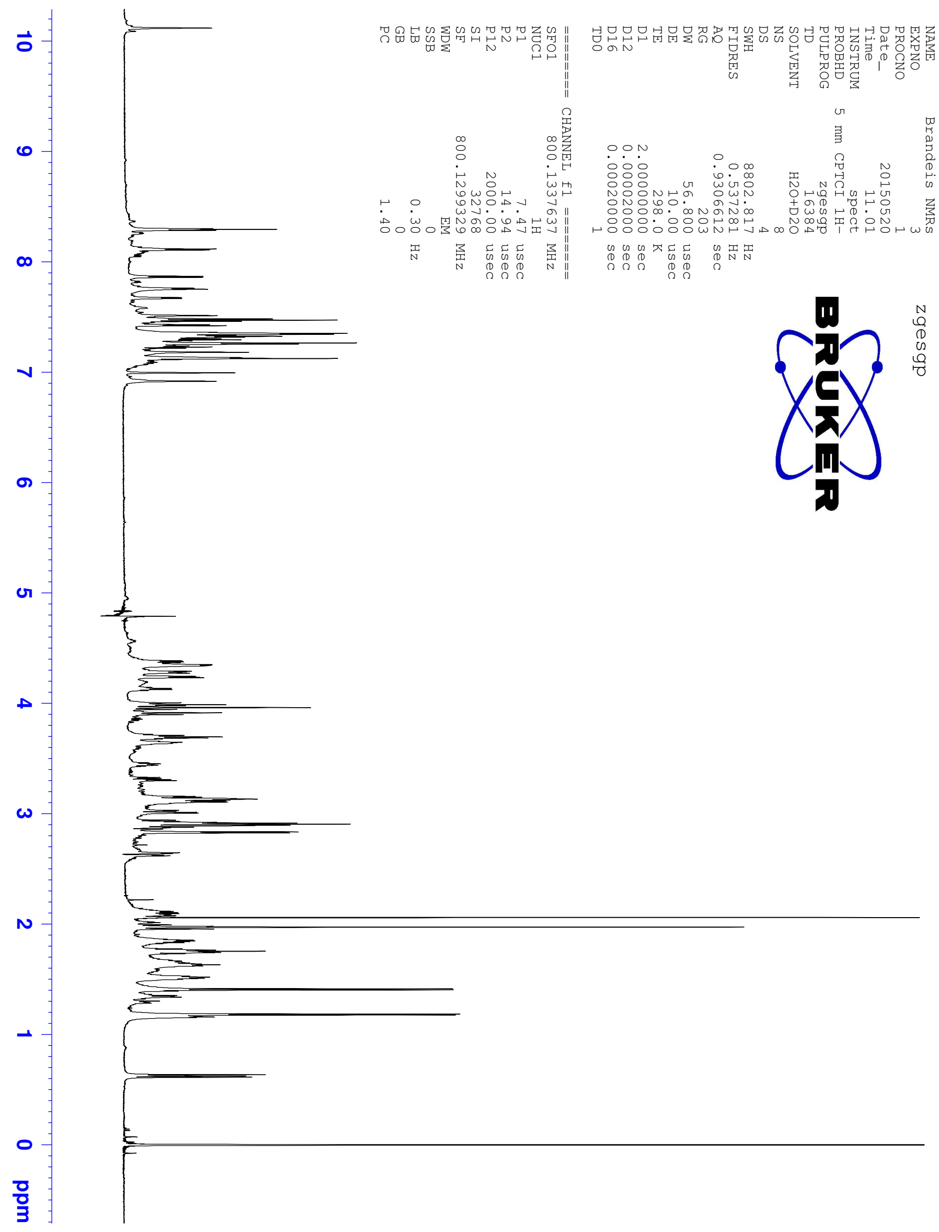


Figure 5.38 1D-^{1}H spectra of **cST2-*ox*** in 9:1 H_2O:D_2O at 25 ºC with the zgesgp pulse program on the Brandeis 800 MHz NMR.

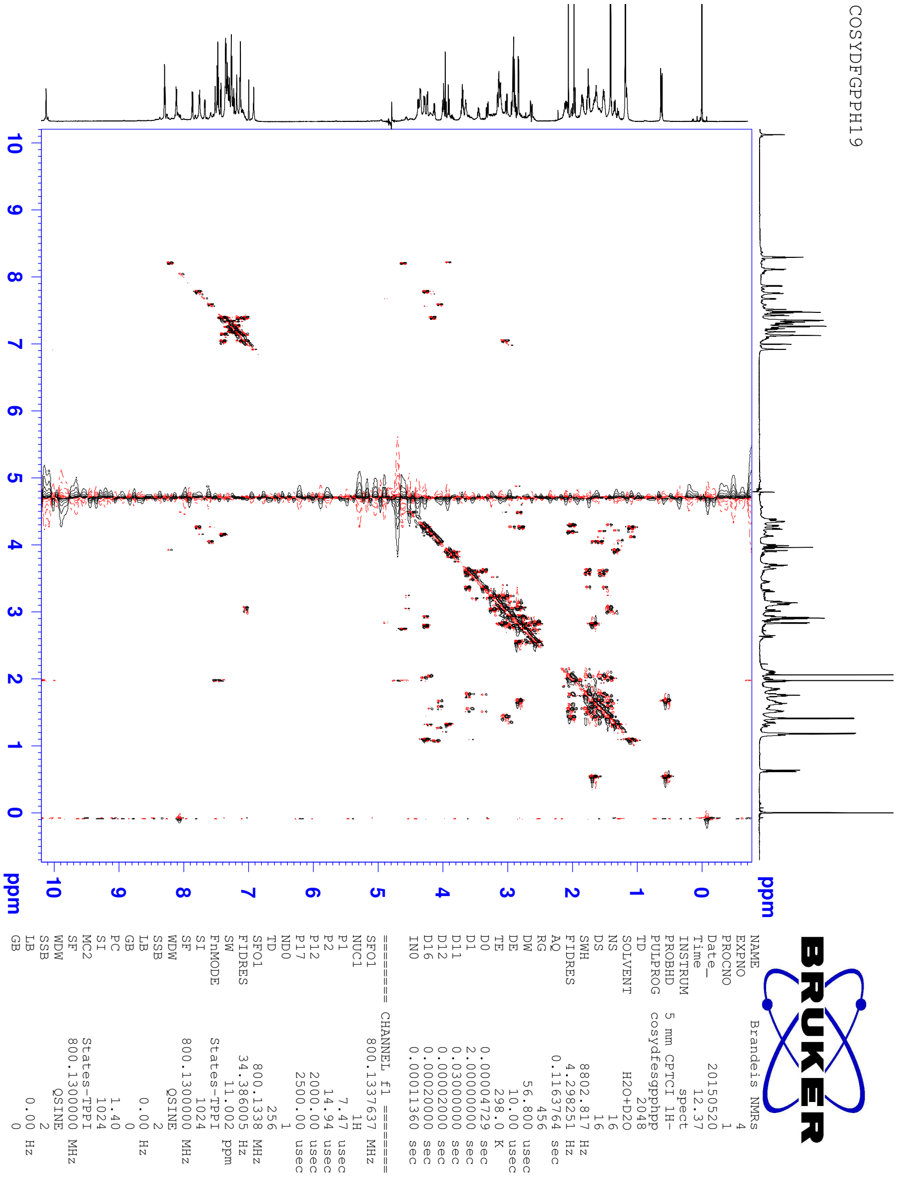
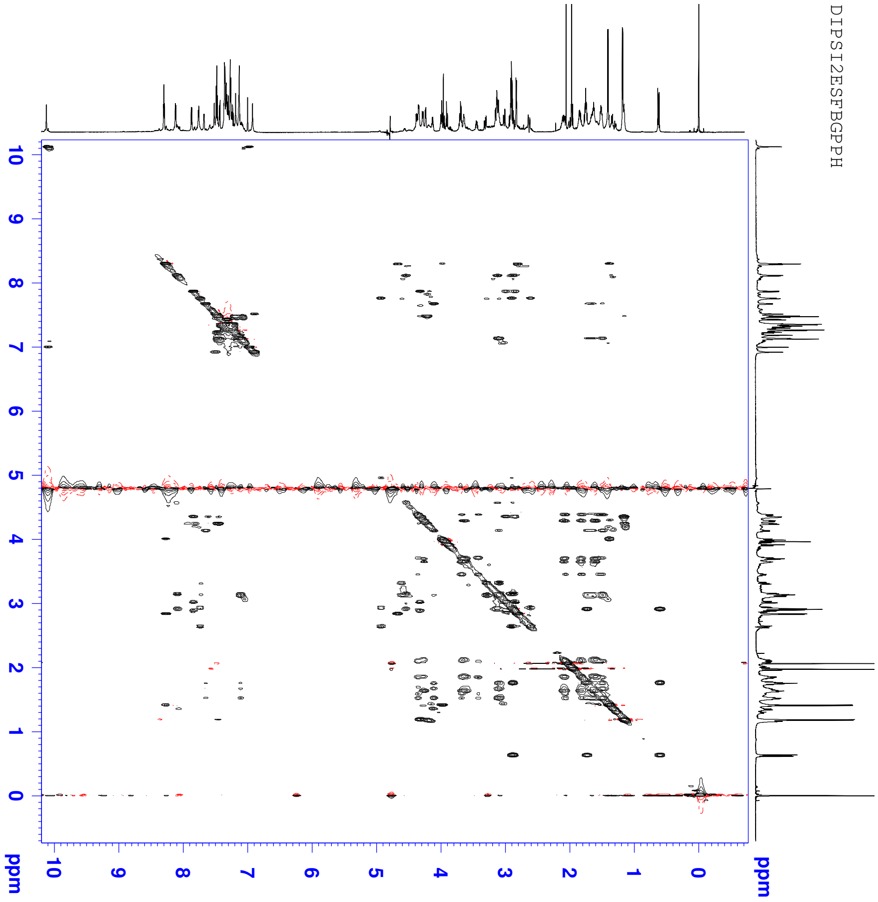


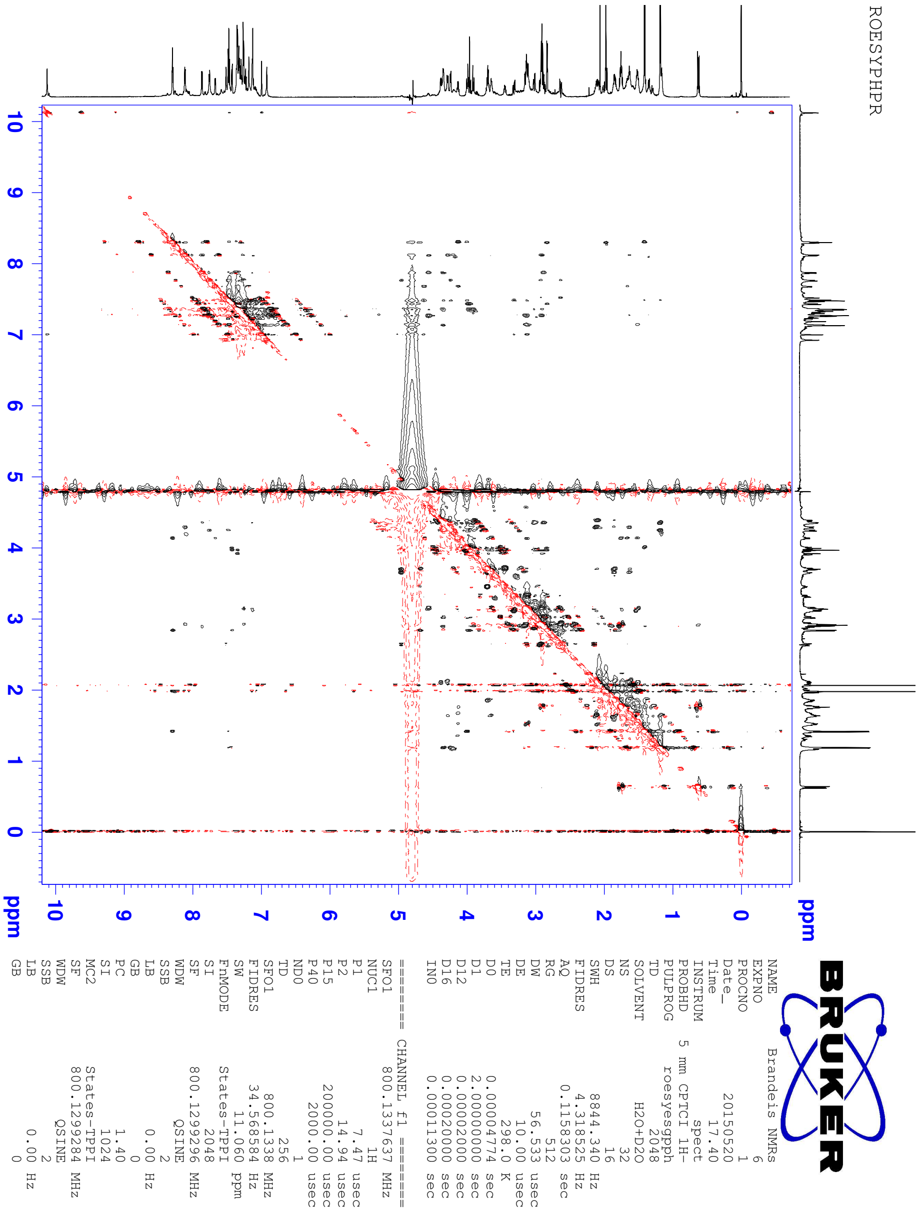
Figure 5.39 A 2D-COSY spectra of **cST2-*ox*** at 25 ºC in 9:1 H_2O:D_2O using the cosydfesgpphpp pulse program on the Brandeis 800 MHz NMR.



```
NAME          Brandeis NMRs
EXPNO                     5
PROCNO                    1
Date_              20150520
Time                  15.07
INSTRUM               spect
PROBHD     5 mm CPTCI 1H-
PULPROG    dipsi2esfbgpph
TD                     2048
SOLVENT             H2O+D2O
NS                       16
DS                       32
SWH                8844.340 Hz
FIDRES             4.318525 Hz
AQ                0.1158303 sec
RG                      512
DW                   56.533 usec
DE                    13.74 usec
TE                    298.0 K
D0               0.00004699 sec
D1               2.00000000 sec
D9               0.06000000 sec
D11              0.03000000 sec
D12              0.00002000 sec
D13              0.00000400 sec
D16              0.00020000 sec
IN0              0.00011300 sec
L1                       22
ZGOPTNS

======== CHANNEL f1 ========
SFO1            800.1337637 MHz
NUC1                     1H
P1                     7.47 usec
P2                    14.94 usec
P6                    24.00 usec
P12                 2000.00 usec
ND0                       1
TD                      256
SFO1               800.1338 MHz
FIDRES            34.568584 Hz
SW                   11.060 ppm
FnMODE          States-TPPI
SI                     1024
SF              800.1299290 MHz
WDW                   QSINE
SSB                       2
LB                     0.00 Hz
GB                        0
PC                     1.40
SI                     1024
MC2             States-TPPI
SF              800.1299535 MHz
WDW                   QSINE
SSB                       2
LB                     0.00 Hz
GB                        0
```

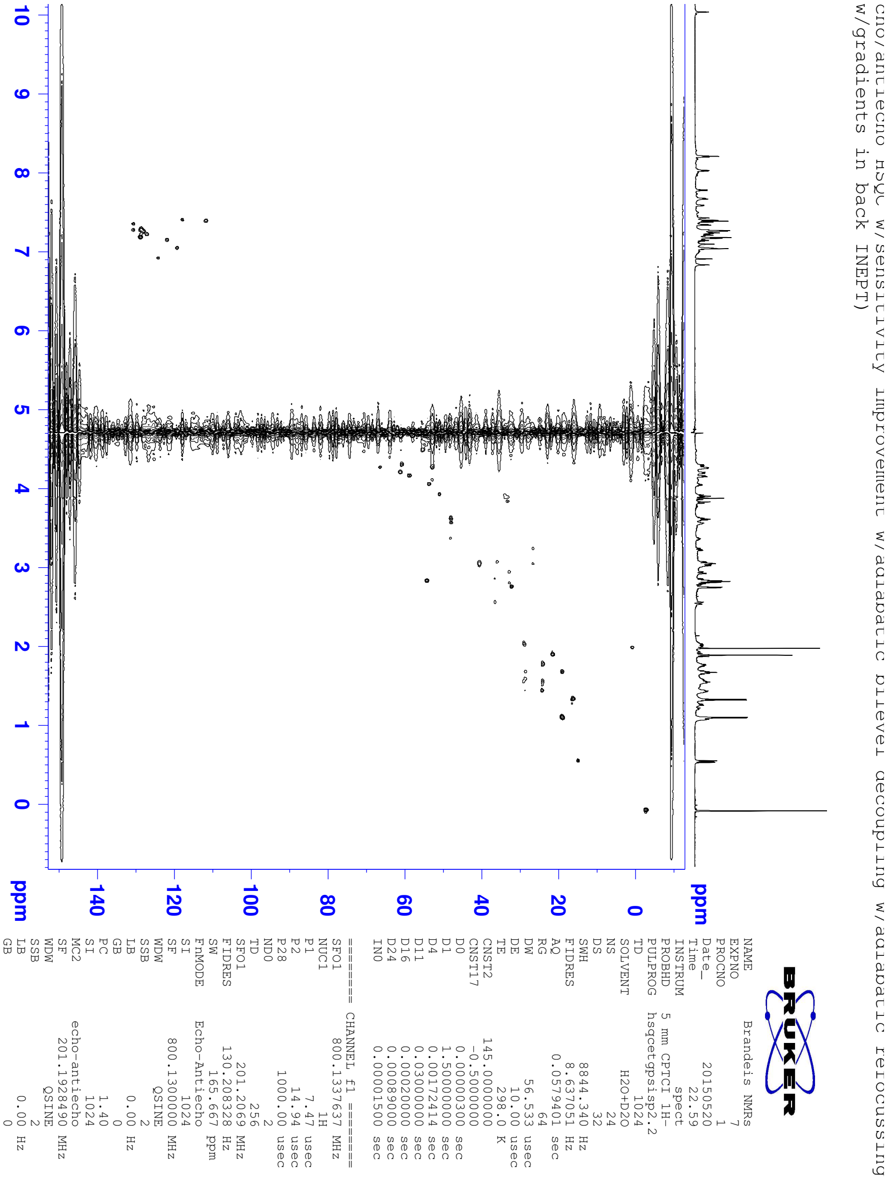
Figure 5.40 A 2D-TOCSY spectra of **cST2-*ox*** at 25 ºC in 9:1 H_2O:D_2O using the dipsi2esfbgpph pulse program on the Brandeis 800 MHz NMR.



```
NAME          Brandeis NMRs
EXPNO                     6
PROCNO                    1
Date_              20150520
Time                  17.40
INSTRUM               spect
PROBHD   5 mm CPTCI 1H-
PULPROG        roesyesgpph
TD                     2048
SOLVENT             H2O+D2O
NS                       32
DS                       16
SWH                8844.340 Hz
FIDRES             4.318525 Hz
AQ                0.1158303 sec
RG                      512
DW                   56.533 usec
DE                    10.00 usec
TE                    298.0 K
D0               0.00004774 sec
D1               2.00000000 sec
D12              0.00002000 sec
D16              0.00020000 sec
IN0              0.00011300 sec

======== CHANNEL f1 ========
SFO1            800.1337637 MHz
NUC1                     1H
P1                     7.47 usec
P2                    14.94 usec
P15                20000.00 usec
P40                 2000.00 usec
ND0                       1
TD                      256
SFO1               800.1338 MHz
FIDRES            34.568584 Hz
SW                   11.060 ppm
FnMODE          States-TPPI
SI                     2048
SF              800.1299296 MHz
WDW                   QSINE
SSB                       2
LB                     0.00 Hz
GB                        0
PC                     1.40
SI                     1024
MC2             States-TPPI
SF              800.1299284 MHz
WDW                   QSINE
SSB                       2
LB                     0.00 Hz
GB                        0
```

Figure 5.41 A 2D-ROESY spectra of **cST2-*ox*** at 25 ºC in 9:1 H_2O:D_2O using the roesyesgpph pule program on the Brandeis 800 MHz NMR.



```
cho/antiecho HSQC w/sensitivity improvement w/adiabatic bilevel decoupling w/adiabatic refocussing
w/gradients in back INEPT)

NAME        Brandeis NMRs
EXPNO                   7
PROCNO                  1
Date_            20150520
Time                22.59
INSTRUM             spect
PROBHD     5 mm CPTCI 1H-
PULPROG   hsqcetgpsisp2.2
TD                   1024
SOLVENT           H2O+D2O
NS                     24
DS                     32
SWH              8844.340 Hz
FIDRES           8.637051 Hz
AQ              0.0579401 sec
RG                     64
DW                 56.533 usec
DE                  10.00 usec
TE                  298.0 K
CNST2         145.0000000
CNST17         -0.5000000
D0             0.00000300 sec
D1             1.50000000 sec
D4             0.00172414 sec
D11            0.03000000 sec
D16            0.00020000 sec
D24            0.00089000 sec
IN0            0.00001500 sec

======== CHANNEL f1 ========
SFO1          800.1337637 MHz
NUC1                   1H
P1                   7.47 usec
P2                  14.94 usec
P28               1000.00 usec
ND0                     2
TD                    256
SFO1             201.2069 MHz
FIDRES         130.208328 Hz
SW                165.667 ppm
FnMODE      Echo-Antiecho
SI                   1024
SF            800.1300000 MHz
WDW                 QSINE
SSB                     2
LB                   0.00 Hz
GB                      0
PC                   1.40
SI                   1024
MC2         echo-antiecho
SF            201.1928490 MHz
WDW                 QSINE
SSB                     2
LB                   0.00 Hz
GB                      0
```

Figure 5.42 A 2D-HSQC spectra of **cST2-*ox*** at 25 ºC in 9:1 H_2O:D_2O using the hsqcetgpsisp2.2 pulse program on the Brandeis 800 MHz NMR.

6 References:

(1) Atwell, S.; Ultsch, M.; De Vos, A. M.; Wells, J. A. *Science* **1997**, *278*, 1125.

(2) Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001.

(3) London, N.; Raveh, B.; Schueler-Furman, O. *Curr Opin Chem Biol* **2013**, *17*, 952.

(4) Ivanov, A. A.; Khuri, F. R.; Fu, H. A. *Trends Pharmacol Sci* **2013**, *34*, 393.

(5) Thiel, P.; Kaiser, M.; Ottmann, C. *Angew Chem Int Edit* **2012**, *51*, 2012.

(6) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv Drug Deliver Rev* **2001**, *46*, 3.

(7) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J Med Chem* **2002**, *45*, 2615.

(8) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv Drug Deliver Rev* **1997**, *23*, 3.

(9) Clackson, T.; Wells, J. *Science* **1995**, *267*, 383.

(10) Jones, S.; Thornton, J. M. *Proceedings of the National Academy of Sciences* **1996**, *93*, 13.

(11) Lo Conte, L.; Chothia, C.; Janin, J. *Journal of molecular biology* **1999**, *285*, 2177.

(12) Moll, U. M.; Petrenko, O. *Mol Cancer Res* **2003**, *1*, 1001.

(13) Degterev, A.; Lugovskoy, A.; Cardone, M.; Mulley, B.; Wagner, G.; Mitchison, T.; Yuan, J. Y. *Nat Cell Biol* **2001**, *3*, 173.

(14) Khoo, K. H.; Verma, C. S.; Lane, D. P. *Nat Rev Drug Discov* **2014**, *13*, 217.

(15) Liu, Q.; Wang, H. G. *Communicative & integrative biology* **2012**, *5*, 557.

(16) Guharoy, M.; Chakrabarti, P. *Bioinformatics* **2007**, *23*, 1909.

(17) Bogan, A. A.; Thorn, K. S. *Journal of molecular biology* **1998**, *280*, 1.

(18) DeLano, W. L. *Current opinion in structural biology* **2002**, *12*, 14.

(19) Gohlke, H.; Kiel, C.; Case, D. A. *Journal of molecular biology* **2003**, *330*, 891.

(20) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *Proteins: Structure, Function, and Bioinformatics* **2007**, *68*, 803.

(21) Keskin, O.; Ma, B.; Rogale, K.; Gunasekaran, K.; Nussinov, R. *Physical Biology* **2005**, *2*, S24.

(22) Gonzalez-Ruiz, D.; Gohlke, H. *Curr Med Chem* **2006**, *13*, 2607.

(23) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. *Nucleic Acids Res* **2005**, *33*, W382.

(24) Kortemme, T.; Baker, D. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 14116.

(25) Kortemme, T.; Kim, D.; Baker, D. *Sci STKE* **2004**, *2004*, pl2.

(26) Jochim, A. L.; Arora, P. S. *ACS Chem. Biol.* **2010**, *5*, 919.

(27) Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R. L.; Sun, X. M.; Orner, B. P. *J Biol Chem* **2010**, *285*, 12078.

(28) Volkov, A. N.; Bashir, O.; Worrall, J. A. R.; Ubbink, M. *Journal of molecular biology* **2009**, *385*, 1003.

(29) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 11287.

(30) Koes, D. R.; Camacho, C. J. *Nucleic Acids Res* **2012**, *40*, W387.

(31) Gao, Y.; Wang, R. X.; Lai, L. H. *Journal of molecular modeling* **2004**, *10*, 44.

(32) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. *Bioinformatics* **2009**, *25*, 621.

(33) London, N.; Raveh, B.; Movshovitz-Attias, D.; Schueler-Furman, O. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 3140.

(34) Gao, M.; London, N.; Cheng, K.; Tamura, R.; Jin, J.; Schueler-Furman, O.; Yin, H. *Peptide Macrocycles* **2014**, *70*, 7664.

(35) Jochim, A. L.; Arora, P. S. *ACS Chem. Biol.* **2010**, *5*, 919.

(36) Bullock, B. N.; Jochim, A. L.; Arora, P. S. *J Am Chem Soc* **2011**, *133*, 14220.

(37) Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; Wagner, G.; Verdine, G. L.; Korsmeyer, S. J. *Science* **2004**, *305*, 1466.

(38) Blackwell, H. E.; Grubbs, R. H. *Angew. Chem. Int. Ed.* **1998**, *37*, 3281.

(39) Jacobsen, Ø.; Maekawa, H.; Ge, N.-H.; Görbitz, C. H.; Rongved, P. l.; Ottersen, O. P.; Amiry-Moghaddam, M.; Klaveness, J. *J. Org. Chem.* **2011**, *76*, 1228.

(40) Patgiri, A.; Jochim, A. L.; Arora, P. S. *Acc Chem Res* **2008**, *41*, 1289.

(41) Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 14751.

(42) Kritzer, J. A.; Luedtke, N. W.; Harker, E. A.; Schepartz, A. *J Am Chem Soc* **2005**, *127*, 14584.

(43) Kritzer, J. A.; Stephens, O. M.; Guarracino, D. A.; Reznik, S. K.; Schepartz, A. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 11.

(44) Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. *Chem Rev* **2001**, *101*, 3219.

(45) Orner, B. P.; Ernst, J. T.; Hamilton, A. D. *J Am Chem Soc* **2001**, *123*, 5382.

(46) Isidro-Llobet, A.; Murillo, T.; Bello, P.; Cilibrizzi, A.; Hodgkinson, J. T.; Galloway, W. R.; Bender, A.; Welch, M.; Spring, D. R. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 6793.

(47) Watkins, A. M.; Arora, P. S. *ACS Chem. Biol.* **2014**, *9*, 1747.

(48) Jayatunga, M. K. P.; Thompson, S.; Hamilton, A. D. *Bioorg Med Chem Lett* **2014**, *24*, 717.

(49) Loughlin, W. A.; Tyndall, J. D. A.; Glenn, M. P.; Fairlie, D. P. *Chem Rev* **2004**, *104*, 6085.

(50) Pelay-Gimeno, M.; Glas, A.; Koch, O.; Grossmann, T. N. *Angew Chem Int Edit* **2015**, *54*, 8896.

(51) North, B.; Lehmann, A.; Dunbrack, R. L. *Journal of molecular biology* **2011**, *406*, 228.

(52) Cunningham, B. C.; Jhurani, P.; Ng, P.; Wells, J. A. *Science* **1988**, *243*, 1330.

(53) Wieland, T.; Faulstich, H. *Experientia* **1991**, *47*, 1186.

(54) Schreiber, S. L.; Crabtree, G. R. *Immunol Today* **1992**, *13*, 136.

(55) Bockus, A. T.; McEwen, C. M.; Lokey, R. S. *Current topics in medicinal chemistry* **2013**, *13*, 821.

(56) Bock, J. E.; Gavenonis, J.; Kritzer, J. A. *ACS Chem Biol* **2013**, *8*, 488.

(57) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. *Nature Chemical Biology* **2014**, *10*, 716.

(58) Chaudhury, S.; Lyskov, S.; Gray, J. J. *Bioinformatics* **2010**, *26*, 689.

(59) Shulman-Peleg, A.; Shatsky, M.; Nussinov, R.; Wolfson, H. J. *BMC biology* **2007**, *5*, 43.

(60) McVey, C. E.; Walsh, M. A.; Dodson, G. G.; Wilson, K. S.; Brannigan, J. A. *Journal of molecular biology* **2001**, *313*, 139.

(61) Harris, B. Z.; Lim, W. A. *Journal of cell science* **2001**, *114*, 3219.

(62) Hao, B.; Zheng, N.; Schulman, B. A.; Wu, G.; Miller, J. J.; Pagano, M.; Pavletich, N. P. *Molecular cell* **2005**, *20*, 9.

(63) DeLano, W. L. DeLabo Scientific LLC.

(64) Golovin, A.; Henrick, K. *BMC bioinformatics* **2008**, *9*, 312.

(65) Bergey, C. M.; Watkins, A. M.; Arora, P. S. *Bioinformatics* **2013**.

(66) Bullock, B. N.; Jochim, A. L.; Arora, P. S. *J Am Chem Soc* **2011**, *133*, 14220.

(67) Robinson, J. A. *Accounts Chem Res* **2008**, *41*, 1278.

(68) Haubner, R.; Gratias, R.; Diefenbach, B.; Goodman, S. L.; Jonczyk, A.; Kessler, H. *J Am Chem Soc* **1996**, *118*, 7461.

(69) Haubner, R.; Schmitt, W.; Holzemann, G.; Goodman, S. L.; Jonczyk, A.; Kessler, H. *J Am Chem Soc* **1996**, *118*, 7881.

(70) Olson, G. L.; Voss, M. E.; Hill, D. E.; Kahn, M.; Madison, V. S.; Cook, C. M. *J Am Chem Soc* **1990**, *112*, 323.

(71) Nagai, U.; Sato, K.; Nakamura, R.; Kato, R. *Tetrahedron* **1993**, *49*, 3577.

(72) Souers, A. J.; Ellman, J. A. *Tetrahedron* **2001**, *57*, 7431.

(73) Lo, Y. C.; Ko, T. P.; Su, W. C.; Su, T. L.; Wang, A. H. J. *Journal of inorganic biochemistry* **2009**, *103*, 1082.

(74) Samal, A.; Schormann, N.; Cook, W. J.; DeLucas, L. J.; Chattopadhyay, D. *Acta Crystallographica Section D* **2007**, *63*, 571.

(75) Goto, M.; Hayashi, H.; Miyahara, I.; Hirotsu, K.; Yoshida, M.; Oikawa, T. *To be Published.*

(76) Ho, M. C.; Menetret, J. F.; Tsuruta, H.; Allen, K. N. *Nature* **2009**, *459*, 393.

(77) Tirupati, B.; Vey, J. L.; Drennan, C. L.; Bollinger, J. M. *Biochemistry-Us* **2004**, *43*, 12210.

(78) Chang, W.-r.; Jiang, T.; Wan, Z.-l.; Zhang, J.-p.; Yang, Z.-x.; Liang, D.-c. *Journal of molecular biology* **1996**, *262*, 721.

(79) Robien, M. A.; Hol, W. G. J. *To be Published.*

(80) Mackenzie, J.; Pedersen, L.; Arent, S.; Henriksen, A. *J Biol Chem* **2006**, *281*, 31012.

(81) Öster, L. M.; Lester, D. R.; Terwisscha van Scheltinga, A.; Svenda, M.; van Lun, M.; Généreux, C.; Andersson, I. *Journal of molecular biology* **2006**, *358*, 546.

(82) Gu, M.; Rajashankar, K. R.; Lima, C. D. *Structure* **2010**, *18*, 216.

(83) Cordes, F. S.; Roversi, P.; Kraiczy, P.; Simon, M. M.; Brade, V.; Jahraus, O.; Wallis, R.; Skerka, C.; Zipfel, P. F.; Wallich, R.; Lea, S. M. *Nature Structural & Molecular Biology* **2005**, *12*, 276.

(84) Miallau, L.; Faller, M.; Chiang, J.; Arbing, M.; Guo, F.; Cascio, D.; Eisenberg, D. *J Biol Chem* **2009**, *284*, 276.

(85) Kilmartin, J. *Unpublished Results* **2010**.

(86) Boumis, G.; Giardina, G.; Angelucci, F.; Bellelli, A.; Brunori, M.; Dimastrogiovanni, D.; Saccoccia, F.; Miele, A. E. *Biochem Bioph Res Co* **2012**, *425*, 806.

(87) Pike, A. C. *Unpublished Results* **2013**.

(88) Chen, P. H.; Chen, X. Y.; Lin, Z. H.; Fang, D. Y.; He, X. L. *Gene Dev* **2013**, *27*, 1345.

(89) Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. *Journal of molecular biology* **1987**, *196*, 641.

(90) Lins, L.; Thomas, A.; Brasseur, R. *Protein Science* **2003**, *12*, 1406.

(91) Tsai, C.-J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *Protein Science* **1997**, *6*, 53.

(92) Janin, J.; Miller, S.; Chothia, C. *Journal of molecular biology* **1988**, *204*, 155.

(93) Lee, K. H., Xie, D., Freire, E., Amzel, L.M., *Proteins Struct. Funct. Bioinforma.* **1994**, *20*, 68.

(94) Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S.-G. *Plos One*, *7*.

(95) Schlessinger, A.; Rost, B. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 115.

(96) Kodan, A.; Yamaguchi, T.; Nakatsu, T.; Sakiyama, K.; Hipolito, C. J.; Fujioka, A.; Hirokane, R.; Ikeguchi, K.; Watanabe, B.; Hiratake, J.; Kimura, Y.; Suga, H.; Ueda, K.; Kato, H. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 4049.

(97) Chen, S. Y.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. *Angew Chem Int Edit* **2014**, *53*, 1602.

(98) Yamagata, K.; Goto, Y.; Nishimasu, H.; Morimoto, J.; Ishitani, R.; Dohmae, N.; Takeda, N.; Nagai, R.; Komuro, I.; Suga, H.; Nureki, O. *Structure* **2014**, *22*, 345.

(99) Glas, A.; Bier, D.; Hahne, G.; Rademacher, C.; Ottmann, C.; Grossmann, T. N. *Angew Chem Int Edit* **2014**, *53*, 2489.

(100) Filippakopoulos, P.; Low, A.; Sharpe, T. D.; Uppenberg, J.; Yao, S. G.; Kuang, Z. H.; Savitsky, P.; Lewis, R. S.; Nicholson, S. E.; Norton, R. S.; Bullock, A. N. *Journal of molecular biology* **2010**, *401*, 389.

(101) Sundström, M.; Lundqvist, T.; Rödin, J.; Giebel, L. B.; Milligan, D.; Norstedt, G. *J Biol Chem* **1996**, *271*, 32197.

(102) Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081.

(103) Cunningham, B. C.; Wells, J. A. *Proceedings of the National Academy of Sciences* **1991**, *88*, 3407.

(104) Cunningham, B. C.; Wells, J. A. *Journal of molecular biology* **1993**, *234*, 554.

(105) Kensler, T. W.; Wakabayash, N.; Biswal, S. *Annu Rev Pharmacol* **2007**, *47*, 89.

(106) Hancock, R.; Bertrand, H. C.; Tsujita, T.; Naz, S.; El-Bakry, A.; Laoruchupong, J.; Hayes, J. D.; Wells, G. *Free Radical Biology and Medicine* **2012**, *52*, 444.

(107) Sun, H. P.; Jiang, Z. Y.; Zhang, M. Y.; Lu, M. C.; Yang, T. T.; Pan, Y.; Huang, H. Z.; Zhang, X. J.; You, Q. D. *Medchemcomm* **2014**, *5*, 93.

(108) Hong, D. S.; Kurzrock, R.; Supko, J. G.; He, X. Y.; Naing, A.; Wheler, J.; Lawrence, D.; Eder, J. P.; Meyer, C. J.; Ferguson, D. A.; Mier, J.; Konopleva, M.; Konoplev, S.; Andreeff, M.; Kufe, D.; Lazarus, H.; Shapiro, G. I.; Dezube, B. J. *Clin Cancer Res* **2012**, *18*, 3396.

(109) Liby, K. T.; Yore, M. M.; Sporn, M. B. *Nat Rev Cancer* **2007**, *7*, 357.

(110) Kobayashi, M.; Li, L.; Iwamoto, N.; Nakajima-Takagi, Y.; Kaneko, H.; Nakayama, Y.; Eguchi, M.; Wada, Y.; Kumagai, Y.; Yamamoto, M. *Molecular and Cellular Biology* **2009**, *29*, 493.

(111) Dinkova-Kostova, A. T.; Talalay, P.; Sharkey, J.; Zhang, Y.; Holtzclaw, W. D.; Wang, X. J.; David, E.; Schiavoni, K. H.; Finlayson, S.; Mierke, D. F.; Honda, T. *J Biol Chem* **2010**, *285*, 33747.

(112) Lo, S. C.; Li, X. C.; Henzl, M. T.; Beamer, L. J.; Hannink, M. *Embo J* **2006**, *25*, 3605.

(113) Tong, K. I.; Katoh, Y.; Kusunoki, H.; Itoh, K.; Tanaka, T.; Yamamoto, M. *Molecular and Cellular Biology* **2006**, *26*, 2887.

(114) Hancock, R.; Schaap, M.; Pfister, H.; Wells, G. *Org Biomol Chem* **2013**, *11*, 3553.

(115) Tanimura, R., Toray Industries Inc.

(116) Mazar, A. P.; Ahn, R. W.; O'Halloran, T. V. *Curr Pharm Design* **2011**, *17*, 1970.

(117) Huai, Q.; Mazar, A. P.; Kuo, A.; Parry, G. C.; Shaw, D. E.; Callahan, J.; Li, Y.; Yuan, C.; Bian, C.; Chen, L.; Furie, B.; Furie, B. C.; Cines, D. B.; Huang, M. *Science* **2006**, *311*, 656.

(118) Sato, S.; Kopitz, C.; Schmalix, W. A.; Muehlenweg, B.; Kessler, H.; Schmitt, M.; Kruger, A.; Magdolen, V. *Febs Lett* **2002**, *528*, 212.

(119) Skerra, A. *Journal of Molecular Recognition* **2000**, *13*, 167.

(120) Schilling, J.; Schoppe, J.; Pluckthun, A. *Journal of molecular biology* **2014**, *426*, 691.

(121) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* **2000**, *28*, 235.

(122) Smith, M. C.; Gestwicki, J. E. *Expert Rev Mol Med* **2012**, *14*.

(123) Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. *Nature Chemical Biology* **2014**, *10*, 723.

(124) Hopkins, A. L.; Groom, C. R.; Alex, A. *Drug Discov Today* **2004**, *9*, 430.

(125) de Lau, W.; Peng, W. C.; Gros, P.; Clevers, H. *Gene Dev* **2014**, *28*, 305.

(126) Hao, H.-X.; Xie, Y.; Zhang, Y.; Charlat, O.; Oster, E.; Avello, M.; Lei, H.; Mickanin, C.; Liu, D.; Ruffner, H.; Mao, X.; Ma, Q.; Zamponi, R.; Bouwmeester, T.; Finan, P. M.; Kirschner, M. W.; Porter, J. A.; Serluca, F. C.; Cong, F. *Nature* **2012**, *485*, 195.

(127) Wang, D.; Huang, B.; Zhang, S.; Yu, X.; Wu, W.; Wang, X. *Gene Dev* **2013**, *27*, 1339.

(128) Zebisch, M.; Xu, Y.; Krastev, C.; MacDonald, B. T.; Chen, M.; Gilbert, R. J. C.; He, X.; Jones, E. Y. *Nat Commun* **2013**, *4*.

(129) Moad, H. E.; Pioszak, A. A. *Biochemistry-Us* **2013**, *52*, 7295.

(130) Wisniewska, M.; Goettig, P.; Maskos, K.; Belouski, E.; Winters, D.; Hecht, R.; Black, R.; Bode, W. *Journal of molecular biology* **2008**, *381*, 1307.

(131) Moss, M. L.; Bartsch, J. W. *Biochemistry-Us* **2004**, *43*, 7227.

(132) Black, R. A.; Rauch, C. T.; Kozlosky, C. J.; Peschon, J. J.; Slack, J. L.; Wolfson, M. F.; Castner, B. J.; Stocking, K. L.; Reddy, P.; Srinivasan, S.; Nelson, N.; Boiani, N.; Schooley, K. A.; Gerhart, M.; Davis, R.; Fitzner, J. N.; Johnson, R. S.; Paxton, R. J.; March, C. J.; Cerretti, D. P. *Nature* **1997**, *385*, 729.

(133) Idriss, H. T.; Naismith, J. H. *Microsc. Res. Tech.* **2000**, *50*, 184.

(134) Lee, M.-H.; Maskos, K.; Knäuper, V.; Dodds, P.; Murphy, G. *Protein Science* **2009**, *11*, 2493.

(135) Amour, A.; Slocombe, P. M.; Webster, A.; Butler, M.; Knight, C. G.; Smith, B. J.; Stephens, P. E.; Shelley, C.; Hutton, M.; Knäuper, V.; Docherty, A. J. P.; Murphy, G. *Febs Lett* **1998**, *435*, 39.

(136) Qiu, Z.; Yan, M.; Li, Q.; Liu, D.; Van den Steen, P. E.; Wang, M.; Opdenakker, G.; Hu, J. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2012**, *27*, 533.

(137) Murumkar, P. R.; DasGupta, S.; Chandani, S. R.; Giridhar, R.; Yadav, M. R. *Expert Opinion on Therapeutic Patents* **2010**, *20*, 31.

(138) Fisher, J. F.; Mobashery, S. *Cancer Metast Rev* **2006**, *25*, 115.

(139) Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. *Chembiochem* **2005**, *6*, 821.

(140) Wisniewska, M.; Goettig, P.; Maskos, K.; Belouski, E.; Winters, D.; Hecht, R.; Black, R.; Bode, W. *Journal of molecular biology* **2008**, *381*, 1307.

(141) Yu, W.; Guo, Z.; Orth, P.; Madison, V.; Chen, L.; Dai, C.; Feltz, R. J.; Girijavallabhan, V. M.; Kim, S. H.; Kozlowski, J. A.; Lavey, B. J.; Li, D.; Lundell, D.; Niu, X.; Piwinski, J. J.; Popovici-Muller, J.; Rizvi, R.; Rosner, K. E.; Shankar, B. B.; Shih, N.-Y.; Arshad Siddiqui, M.; Sun, J.; Tong, L.; Umland, S.; Wong, M. K. C.; Yang, D.-y.; Zhou, G. *Bioorg Med Chem Lett* **2010**, *20*, 1877.

(142) Hamilton, C. J.; Saravanamuthu, A.; Eggleston, I. M.; Fairlamb, A. H. *Biochemical Journal* **2003**, *369*, 529.

(143) Walsh, C.; Bradley, M.; Nadeau, K. *Trends in biochemical sciences* **1991**, *16*, 305.

(144) WHO In *Fact sheets*; centre, W. M., Ed.; WHO: http://www.who.int/mediacentre/factsheets/fs259/en/, 2016; Vol. 2016.

(145) Khan, M. O. *Drug target insights* **2007**, *2*, 129.

(146) Austin, S. E.; Khan, M. O.; Douglas, K. T. *Drug design and discovery* **1999**, *16*, 5.

(147) Dixon, M. J.; Maurer, R. I.; Biggi, C.; Oyarzabal, J.; Essex, J. W.; Bradley, M. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 4513.

(148) Khan, M. O.; Austin, S. E.; Chan, C.; Yin, H.; Marks, D.; Vaghjiani, S. N.; Kendrick, H.; Yardley, V.; Croft, S. L.; Douglas, K. T. *J Med Chem* **2000**, *43*, 3148.

(149) Parveen, S.; Khan, M. O.; Austin, S. E.; Croft, S. L.; Yardley, V.; Rock, P.; Douglas, K. T. *J Med Chem* **2005**, *48*, 8087.

(150) Toro, M. A.; Sánchez-Murcia, P. A.; Moreno, D.; Ruiz-Santaquiteria, M.; Alzate, J. F.; Negri, A.; Camarasa, M.-J.; Gago, F.; Velázquez, S.; Jiménez-Ruiz, A. *Chembiochem* **2013**, *14*, 1212.

(151) Cavallo, L.; Kleinjung, J.; Fraternali, F. *Nucleic Acids Res* **2003**, *31*, 3364.

(152) Klett, J.; Nunez-Salgado, A.; Dos Santos, H. G.; Cortes-Cabrera, A.; Perona, A.; Gil-Redondo, R.; Abia, D.; Gago, F.; Morreale, A. *J Chem Theory Comput* **2012**, *8*, 3395.

(153) Baiocco, P.; Colotti, G.; Franceschini, S.; Ilari, A. *J Med Chem* **2009**, *52*, 2603.

(154) Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. *Chembiochem* **2005**, *6*, 821.

(155) Confalonieri, S.; Di Fiore, P. P. *Febs Lett* **2002**, *513*, 24.

(156) Henegouwen, P. M. P. V. E. *Cell Communication and Signaling* **2009**, *7*.

(157) Suzuki, R.; Toshima, J. Y.; Toshima, J. *Mol Biol Cell* **2012**, *23*, 687.

(158) vanDelft, S.; Schumacher, C.; Hage, W.; Verkleij, A. J.; Henegouwen, P. M. P. V. E. *J Cell Biol* **1997**, *136*, 811.

(159) Benmerah, A.; Bayrou, M.; Cerf-Bensussan, N.; Dautry-Varsat, A. *Journal of Cell Science* **1999**, *112*, 1303.

(160) Benmerah, A.; Poupon, V.; Cerf-Bensussan, N.; Dautry-Varsat, A. *J Biol Chem* **2000**, *275*, 3288.

(161) Carbone, R.; Fre, S.; Iannolo, G.; Belleudi, F.; Mancini, P.; Pelicci, P. G.; Torrisi, M. R.; Di Fiore, P. P. *Cancer Research* **1997**, *57*, 5498.

(162) Haffner, C.; Takei, K.; Chen, H.; Ringstad, N.; Hudson, A.; Butler, M. H.; Salcini, A. E.; Di Fiore, P. P.; De Camilli, P. *Febs Lett* **1997**, *419*, 175.

(163) Rumpf, J.; Simon, B.; Jung, N.; Maritzen, T.; Haucke, V.; Sattler, M.; Groemping, Y. *Embo J* **2008**, *27*, 558.

(164) Tsushima, H.; Malabarba, M. G.; Confalonieri, S.; Senic-Matuglia, F.; Verhoef, L. G. G. C.; Bartocci, C.; D'Ario, G.; Cocito, A.; Di Fiore, P. P.; Salcini, A. E. *Plos One* **2013**, *8*.

(165) Koh, T. W.; Korolchuk, V. I.; Wairkar, Y. P.; Jiao, W.; Evergren, E.; Pan, H. L.; Zhou, Y.; Venken, K. J. T.; Shupliakov, O.; Robinson, I. M.; O'Kane, C. J.; Bellen, H. J. *J Cell Biol* **2007**, *178*, 309.

(166) Benmerah, A.; Lamaze, C.; Begue, B.; Schmid, S. L.; Dautry-Varsat, A.; Cerf-Bensussan, N. *J Cell Biol* **1998**, *140*, 1055.

(167) Bhattacharyya, S.; Warfield, K. L.; Ruthel, G.; Bavari, S.; Aman, M. J.; Hope, T. J. *Virology* **2010**, *401*, 18.

(168) Dutta, D.; Donaldson, J. G. *Cellular logistics* **2012**, *2*, 203.

(169) Gonzalez-Jamett, A. M.; Momboisse, F.; Haro-Acuna, V.; Bevilacqua, J. A.; Caviedes, P.; Cardenas, A. M. *Frontiers in endocrinology* **2013**, *4*, 126.

(170) Macia, E.; Ehrlich, M.; Massol, R.; Boucrot, E.; Brunner, C.; Kirchhausen, T. *Developmental Cell* **2006**, *10*, 839.

(171) Vercauteren, D.; Vandenbroucke, R. E.; Jones, A. T.; Rejman, J.; Demeester, J.; De Smedt, S. C.; Sanders, N. N.; Braeckmans, K. *Mol Ther* **2010**, *18*, 561.

(172) Liu, S. H.; Marks, M. S.; Brodsky, F. M. *J Cell Biol* **1998**, *140*, 1023.

(173) Zhao, X. H.; Greener, T.; Al-Hasani, H.; Cushman, S. W.; Eisenberg, E.; Greene, L. E. *Journal of Cell Science* **2001**, *114*, 353.

(174) Confalonieri, S.; Di Fiore, P. P. *Febs Lett* **2002**, *513*, 24.

(175) Doria, M.; Salcini, A. E.; Colombo, E.; Parslow, T. G.; Pelicci, P. G.; Di Fiore, P. P. *J Cell Biol* **1999**, *147*, 1379.

(176) Maldonado-Baez, L.; Wendland, B. *Trends in Cell Biology* **2006**, *16*, 505.

(177) de Beer, T.; Carter, R. E.; Lobel-Rice, K. E.; Sorkin, A.; Overduin, M. *Science* **1998**, *281*, 1357.

(178) de Beer, T.; Hoofnagle, A. N.; Enmon, J. L.; Bowers, R. C.; Yamabhai, M.; Kay, B. K.; Overduin, M. *Nature Structural Biology* **2000**, *7*, 1018.

(179) Enmon, J. L.; de Beer, T.; Overduin, M. *Biochemistry-Us* **2000**, *39*, 4309.

(180) Kim, S.; Cullis, D. N.; Feig, L. A.; Baleja, J. D. *Biochemistry-Us* **2001**, *40*, 6776.

(181) Kieken, F.; Jovic, M.; Tonelli, M.; Naslavsky, N.; Caplan, S.; Sorgen, P. L. *Protein science : a publication of the Protein Society* **2009**, *18*, 2471.

(182) Kamens, A. J.; Eisert, R. J.; Corlin, T.; Baleja, J. D.; Kritzer, J. A. *Biochemistry-Us* **2014**, *53*, 4758.

(183) Kieken, F.; Sharma, M.; Jovic, M.; Giridharan, S. S. P.; Naslavsky, N.; Caplan, S.; Sorgen, P. L. *J Biol Chem* **2010**, *285*, 8687.

(184) Henry, G. D.; Corrigan, D. J.; Dineen, J. V.; Baleja, J. D. *Biochemistry-Us* **2010**, *49*, 3381.

(185) Whitehead, B.; Tessari, M.; Carotenuto, A.; Henegouwen, P. M. P. V. E.; Vuister, G. W. *Biochemistry-Us* **1999**, *38*, 11271.

(186) Kim, S.; Cullis, D. N.; Feig, L. A.; Baleja, J. D. *Biochemistry-Us* **2001**, *40*, 6776.

(187) Cullis, D. N.; Philip, B.; Baleja, J. D.; Feig, L. A. *J Biol Chem* **2002**, *277*, 49158.

(188) Yamabhai, M.; Hoffman, N. G.; Hardison, N. L.; McPherson, P. S.; Castagnoli, L.; Cesareni, G.; Kay, B. K. *J Biol Chem* **1998**, *273*, 31401.

(189) Nowick, J. S. *Accounts Chem Res* **2008**, *41*, 1319.

(190) Shepherd, N. E.; Hoang, H. N.; Abbenante, G.; Fairlie, D. P. *J Am Chem Soc* **2005**, *127*, 2974.

(191) Fasan, R.; Dias, R. L. A.; Moehle, K.; Zerbe, O.; Obrecht, D.; Mittl, P. R. E.; Grütter, M. G.; Robinson, J. A. *Chembiochem* **2006**, *7*, 515.

(192) Robinson, J. A. *Accounts Chem Res* **2008**, *41*, 1278.

(193) Chen, S.; Morales-Sanfrutos, J.; Angelini, A.; Cutting, B.; Heinis, C. *J Pept Sci* **2012**, *18*, S169.

(194) Schlippe, Y. V. G.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *J Am Chem Soc* **2012**, *134*, 10469.

(195) Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. *Nat Chem Biol* **2009**, *5*, 888.

(196) Jo, H.; Meinhardt, N.; Wu, Y. B.; Kulkarni, S.; Hu, X. Z.; Low, K. E.; Davies, P. L.; DeGrado, W. F.; Greenbaum, D. C. *J Am Chem Soc* **2012**, *134*, 17704.

(197) Bechara, C.; Sagan, S. *Febs Lett* **2013**, *587*, 1693.

(198) Schmidt, N.; Mishra, A.; Lai, G. H.; Wong, G. C. L. *Febs Lett* **2010**, *584*, 1806.

(199) Madani, F.; Lindberg, S.; Langel, U.; Futaki, S.; Graslund, A. *J Biophys* **2011**, *2011*, 414729.

(200) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, P.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* **2005**, *59*, 687.

(201) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr D* **1998**, *54*, 905.