JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

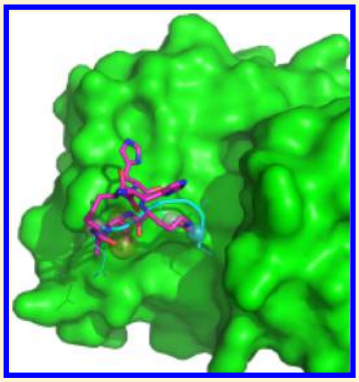
Virtual Screening Using Combinatorial Cyclic Peptide Libraries Reveals Protein Interfaces Readily Targetable by Cyclic Peptides

Fergal J. Duffy,[†,‡,¶] Darragh O'Donovan,[¶,§] Marc Devocelle,[∥] Niamh Moran,[⊥] David J. O'Connell,[¶,§] and Denis C. Shields*[,†,‡,¶]

[†]School of Medicine and Medical Science, [‡]Complex and Adaptive Systems Laboratory, [¶]Conway Institute of Biomolecular and Biomedical Research, and [§]School of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland, and [∥]Department of Chemistry and [⊥]Department of Molecular and Cell Therapeutics, Royal College of Surgeons in Ireland, 123 St. Stephens Green, Dublin 2, Ireland

Supporting Information

ABSTRACT: Protein−protein and protein−peptide interactions are responsible for the vast majority of biological functions *in vivo*, but targeting these interactions with small molecules has historically been difficult. What is required are efficient combined computational and experimental screening methods to choose among a number of potential protein interfaces worthy of targeting lead macrocyclic compounds for further investigation. To achieve this, we have generated combinatorial 3D virtual libraries of short disulfide-bonded peptides and compared them to pharmacophore models of important protein−protein and protein−peptide structures, including short linear motifs (SLiMs), protein-binding peptides, and turn structures at protein−protein interfaces, built from 3D models available in the Protein Data Bank. We prepared a total of 372 reference pharmacophores, which were matched against 108,659 multiconformer cyclic peptides. After normalization to exclude nonspecific cyclic peptides, the top hits notably are enriched for mimetics of turn structures, including a turn at the interaction surface of human α thrombin, and also feature several protein-binding peptides. The top cyclic peptide hits also cover the critical "hot spot" interaction sites predicted from the interaction crystal structure. We have validated our method by testing cyclic peptides predicted to inhibit thrombin, a key protein in the blood coagulation pathway of important therapeutic interest, identifying a cyclic peptide inhibitor with lead-like activity. We conclude that protein interfaces most readily targetable by cyclic peptides and related macrocyclic drugs may be identified computationally among a set of candidate interfaces, accelerating the choice of interfaces against which lead compounds may be screened.



INTRODUCTION

Protein−protein interactions (PPIs) are at the heart of almost all biological processes, but developing drug-like molecules against them has been difficult due to the intrinsic properties of protein−protein interfaces. Protein−protein interfaces generally have large shallow interfaces[1] compared to typical drug binding sites in a deep, well-defined pocket in a protein surface. The large interaction surfaces also tend to easily form weak, transient, nonspecific interactions with small molecule hydrophobic aggregates[2] that can be difficult to distinguish from effective drug activity. It has also been suggested that chemical libraries tailored to common drug targets such as protein kinases and G-coupled protein receptors may not be ideal for mediating PPIs[3] and that current ideas of what constitutes a drug-like molecule may have to be re-examined. Progress has however been made in targeting protein protein interactions[4] by realizing that not all of the surface area of a PPI contributes equally to the strength of the interaction between the protein partners. PPI binding is mediated by "hot spots": small areas of disproportionately high-affinity binding at the protein−protein interface.[5] Drugs targeting PPIs have been found, by random screening, such as cyclosporin[6] and by mimicking important surface epitopes on a protein using a peptidomimetic, such as the RGD mimetics, known to bind integrins.[7] Protein−protein interfaces also tend to be somewhat flexible, with different binding modes for known small molecule inhibitors and for their protein partner. This increases the difficulty of using structural information to predict small molecule binding.[4] Molecular simulation approaches have also been used to search for hot spots and transient pockets which can be then examined using docking.[5]

There is also a possibility that current chemical libraries are biased toward certain classes of targets. Previous analysis has shown that 69.3% of FDA drugs approved as of 2006 act on just 10 families of proteins,[8] with the remainder of drugs working on a further 120 domain families or single proteins. It may be that compound libraries or chemotypes that have been found to be valuable sources of leads for historically important protein classes may not be the same as those that modulate PPIs. It has been observed that molecules inhibiting protein−protein interactions tend to be larger in molecular weight, more three

Received: July 19, 2014
Published: February 10, 2015

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

dimensionally structured, and more rigid than common drugs,[9] possibly to access more of the relatively large PPI surface area. Synthetic macrocyclic molecules intermediate in size between small-molecule drugs and large biological drugs are being commercially explored as PPI inhibitors,[10] which are similar in size to the natural macrocyclic constrained peptides examined in this work.

The lack of success using traditional small molecules to target PPIs has returned focus to peptides as drug candidates.[11] Peptides are traditionally considered to be poor drug candidates due to their low oral bioavailabilty, rapid proteolytic degradation *in vivo*, poor ability to cross biological barriers such as lipid bilayers due to their hydrophilicity, and their high degree of conformational flexibility. However, since proteins have often been designed by evolution to bind with high specificity to polypeptide sequences, peptides can be very effective mediators of biological activity. Peptides can be engineered to represent the smallest functional portion of a protein, offering selectivity and specificity.

Several strategies have been developed to combat the disadvantages of natural linear peptides, including modifying the peptide bond (pseudo peptides), small molecule peptidomimetics, stapled peptides,[12] and cyclic peptides. In this study, we use combinatorial libraries of cyclic peptides to computationally identify protein−protein and protein−peptide interactions which can be inhibited using a short cyclic peptide.

Cyclic peptides are a variety of modified peptide where an extra bond has been added to form the linear peptide into a macrocycle. This bond can be a peptide bond from the N-terminus to the C-terminus, two side-chain amino-acid groups bound together, or some combination of the two. Cyclic peptides exist in nature in the form of disulfide-bonded peptides, with insulin[13] being the best known example. In addition, some natural antimicrobial peptides are also disulfide bonded cyclic peptides.[14]

The main apparent advantages of cyclic peptides are their conformational constraint and proteolytic resistance. Proteases act upon cleavage sites in the peptide sequence, which may be at the N or C-terminus, or sometimes at a particular motif within the peptide. Cyclizing a peptide often changes the cyclic peptide's accessibility to cleavage sites in proteases. A limited set of possible 3-dimensional conformations has benefits for virtual screening, as the difficulty in computationally modeling a flexible molecule increases combinatorially with the number of rotatable bonds in the molecule, meaning that computational problems, such as molecular docking or conformational searches, that are increasingly intractable when dealing with longer linear peptides may be accomplished with cyclic peptides. There is also evidence that cyclic peptides may have improved membrane permeability, compared with linear peptides, in cases where the cyclic peptide can internally satisfy its own hydrogen bonds.[15] Other approaches that have been used to successfully transport linear peptides inside cell membranes, such as palmytoylation,[16] may also be used with cyclic peptides.

Using cyclic peptides does come with disadvantages: their large molecular weight and high number of hydrogen bond donors/acceptors means they generally fare poorly in predicted ADMET (Adsorption, Distribution, Metabolism, Elimination, Toxicology)[17] properties, which is concerning as unfavorable ADMET properties are a primary cause of drugs failing in clinical trials.[18] Despite this, there are several short cyclic peptide based drugs on the market such as cilengitide[19] and eptifibatide.[20]

For this study, we used a library of cyclic peptides against a large panel of protein−protein and protein−peptide interactions to represent a reverse of the standard virtual screening approach - using a tailored library of compounds to find protein targets amenable to modulation. It has the benefit of identifying which targets are best suited to a class of compounds. Three subsets of protein−peptide and protein−protein interactions have been examined: known Short Linear Motif instances (SLiMs or ELMs) as detailed in the ELM database,[21] known protein−peptide interactions from the PepX database,[22] and turn structures at protein−protein interfaces. All examples of the reference structures used in this study have available 3D structures in the Protein Data Bank[23] (PDB), and the complete list is described in Supplementary Table S2. Protein−protein and protein−peptide interactions may be considered as distinct entities; however, it has been shown that peptides bound to proteins adopt structural shapes that are commonly seen in protein folds,[24] allowing them to be treated as a special case of protein−protein interaction. These reference structure types have been chosen to represent small, important portions of a larger protein−protein interface, of a suitable size to potentially allow a short cyclic peptide mimic to be discovered. Using pharmacophore matching, it is possible to break each reference structure into a set of interaction features: hydrogen bond donors and acceptors, hydrophobic and aromatic regions, and positive and negative charges. Pharmacophore features on these reference structures located close to a protein−protein interface can then be taken as the critical interaction features and used to search a multiconformer virtual library of cyclic peptides to identify candidate mimics, along with shape information from the partner protein to exclude cyclic peptides that closely match a reference structure pharmacophore but violate the surface constraints of the partner protein. This then acts as a combination of ligand and structure based drug design, where chemical features from one interactor are combined with structural constraints from the other interactor to search for hits.

The goal of this study is 2-fold - to identify lead cyclic peptides that may be further refined to inhibit protein−protein interactions and to identify known protein−protein interactions that can be considered "druggable" by cyclic peptidomimetics and related compounds.

RESULTS AND DISCUSSION

Preparing Reference Structures and Cyclic Peptides. The purpose of this study was to identify peptide−protein and protein−protein interactions that can be modified by cyclic peptides. This is the reverse of the standard drug discovery approach, where a biological target of interest is identified, and attempts are made to find molecules that will modulate that target, by screening premade libraries of candidate molecules. Our goal is to use a precomputed virtual library with an identical chemical space to natural proteins to explore the interactions of these natural proteins and to find well characterized interactions that appear to be likely to be modifiable and clinically relevant.

The decision on which types of protein−peptide interactions would be examined was driven by practical concerns in finding peptides that may represent key portions of protein−protein interfaces but would still be small enough, in terms of the number of pharmacophore features and physical size, to be

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

amenable to high-throughput virtual screening with pharmacophore matching in a computationally feasible manner. Ideally, reference structures should be of comparable size to the virtual library of cyclic peptides they are to be matched against. Protein−peptide interfaces described in the PepX[22] database, experimentally confirmed short linear motifs (SLiMs) representing protein−ligand interactions (annotated as type LIG in the Eukaryotic Linear Motif (ELM)[21] database) with accompanying PDB 3D structures, and UniProt[25] annotated turns located at protein−protein interfaces were selected as promising candidate peptides. Short linear motif and protein-binding peptide structures longer than 6 residues were discarded, and all remaining structures were used in the pharmacophore matching. The final numbers of reference structures are shown in Table 1. The PDB ID, UniProt accession, and UniProt ID of all the reference structures are detailed in Supplementary Table S2. Four of the PepX structures were also short linear motifs, these are contained in PDB structure 1H27, where a protein binding peptide is an example of the LIG_CYCLIN_1 motif, in PDBs 1KYU and 1KYF where protein binding peptides in both structures are also examples of LIG_AP2alpha_2 motifs, and in PDB structure 2FOO where a protein-binding peptide is also the LIG_USP_7 motif. Most of the PepX peptides are novel synthetic peptides, not extracted from known protein sequences. These include protease substrates, recognition sites for chaperone binding, transporter substrates, and post-translation modification sites.

Table 1. Protein−Protein and Protein−Peptide Structures[a]

structure class	number of structures
short linear motifs	57
protein-binding peptides	161
turns at protein−protein interfaces	154
total	**372**

[a]Classes and numbers of protein−protein and protein−peptide structures used as a basis for the pharmacophore models screened against the cyclic peptide libraries.

Combinatorial synthesizable cyclic peptide libraries of the pattern Cys-X-X-X-X-Cys and Cys-X-X-X-Cys were prepared, where X indicates a variable wild-card position, which could be any of the 20 naturally occurring amino acids. Diverse, low energy conformers were calculated for each library, as described in the Methods section. Numbers of the resulting constrained cyclic peptides and conformers are shown in Table 2.

All reference structures and cyclic peptides were protonated at physiological pH (7.4), and pharmacophore features were generated using Pharao.[26] For reference structures located at an interface in the original Protein Data Bank structure, exclusion spheres were added, and pharmacophore features located away from the interface were discarded. Pharao pharmacophore matching works by decomposing a molecular structure into a collection of chemical features representing hydrogen bond donors, hydrogen bond acceptors, positive and negative charges, and lipophilic and aromatic regions. These chemical features are represented mathematically as 3D Gaussian volumes. To screen molecules, Pharao aligns two pharmacophores and calculates the fractional overlap of the Gaussian features. The mathematically continuous nature of Gaussian volumes allows efficient optimization of the alignment

Table 2. Cyclic Peptide Libraries[a]

cyclic peptide library	peptides	conformers	av conformers
Cys-X-X-X-X-Cys	104,174	2,076,203	19.93
Cys-X-X-X-Cys	4,485	89,549	19.97
total	**108,659**	**2,165,752**	**19.93**

[a]Counts of structures and conformers for synthesizable combinatorially generated cyclic peptides used as candidate compounds for virtual screening.

Pharmacophore Matching Results. Each of the 372 reference structure pharmacophore models was matched to every cyclic peptide conformation pharmacophore, resulting in 805,659,744 pharmacophore matching pairs. Pharmacophore matching score distributions are presented in Figure 1.

Each set of structures has a distinct distribution, with score peaks between 0.72−0.78 in the protein-binding peptide distribution (Figure 1C) and two pronounced peaks at 0.60−0.62 and 0.88−0.90 in the protein−protein interface turn score distributions (Figure 1B). Peaks located at high Tverksy_ref scores may indicate enrichment of structure(s) amenable to mimicking with constrained cyclic peptides in the data set. These distributions contrast with the short linear motif score distribution, which is relatively smooth (Figure 1A). Looking at high scoring pharmacophore matches (Tversky_ref >0.90), there are 18 high-scoring matches to short linear motif structures compared with 39,380 matches to protein-binding peptides and 551,002 matches to turns at interfaces.

When examining the amino-acid composition of the top 200 hits (by Tversky_ref score) for each type of structure (Figure 2), cyclic peptides matching protein-binding peptides (Figure 2B) and turns at protein−protein interactions (Figure 2C) have sequences enriched for small amino acids, such as alanine and glycine. These high matching scores may simply reflect nonspecific peptide backbone interactions or may potentially reflect greater flexibility to adopt certain useful structural conformations.

To ensure that pharmacophore matching identified specific cyclic-peptide-PPI interaction similarities, pharmacophore matches were normalized. The normalization was carried out by calculating a Z-score for each pharmacophore matching score, based on mean and standard deviation values for an individual cyclic peptide. The purpose was to identify cyclic peptides that had more specifically matched a PPI structure. If the scores were normally distributed (and we note that the distribution does not appear normal), we would expect $805{,}659{,}744 \times 2.87 \times 10^{-7} \approx 231$ pharmacophore matches with a Z-score over 5.0. After normalization, there were 553,104 pharmacophore matches with a Z-score over 5.0, and 27,554 pharmacophore matches with a Z-score over 6.0. Structures, along with the number of specifically matching cyclic peptides, are described in Table 3. 76.5% of normalized scores with a Z-score >6.0 have a Tversky_ref score >0.90.

Supplementary Figure S1 indicates that the high pharmacophore matching scores were not particularly biased toward structures with few pharmacophore features, as average pharmacophore matching scores and numbers of pharmacophore features correlate poorly.

Cyclic Peptides Effectively Mimic Protein Turns. Table 3 presents the number of top hits for each protein−protein interaction, where there exists a cyclic peptide hit with a normalized score >6.0. While there are seven peptides matched from the set of turns, and seven from the set of protein−

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

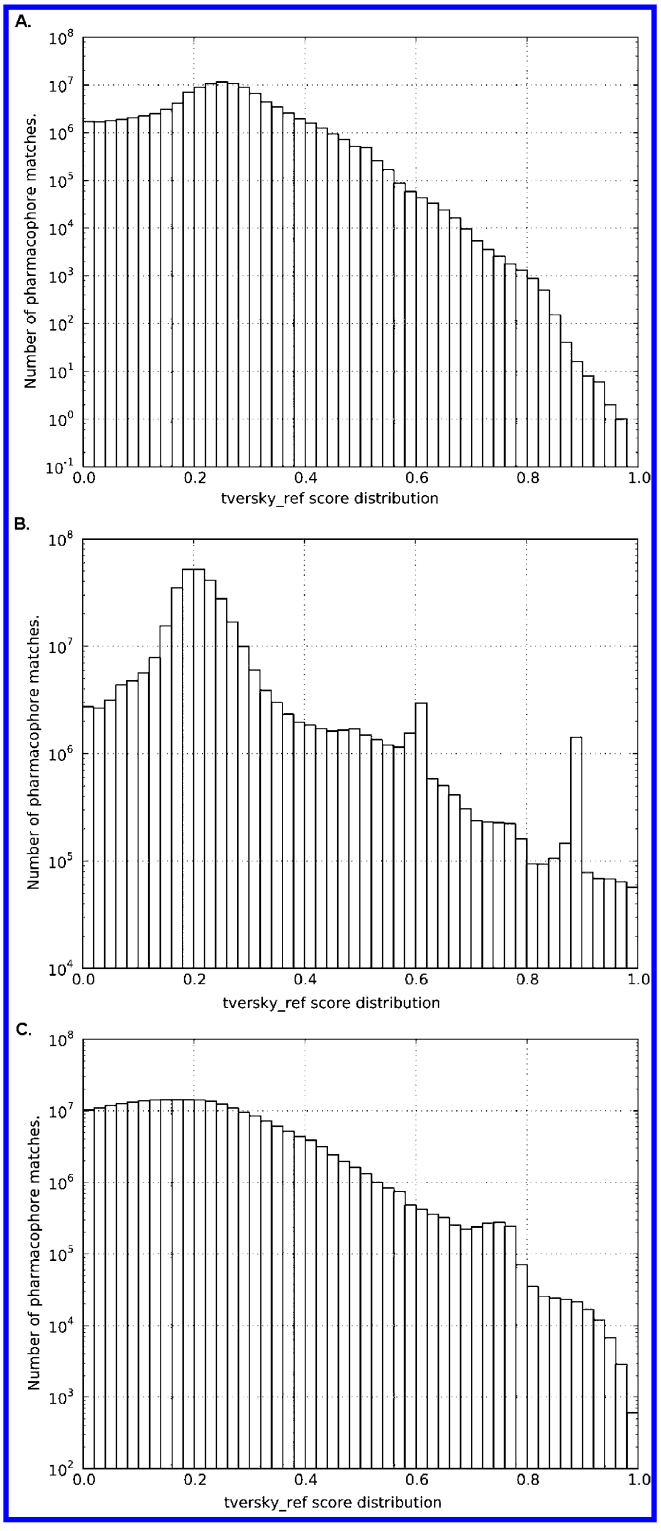


Figure 1. Pharmacophore matching score distributions. Tversky_ref distributions of cyclic peptide pharmacophore matching scores for **A.** short linear motif structures, **B.** turn structures at protein−protein interfaces, and **C.** protein−peptide structures.

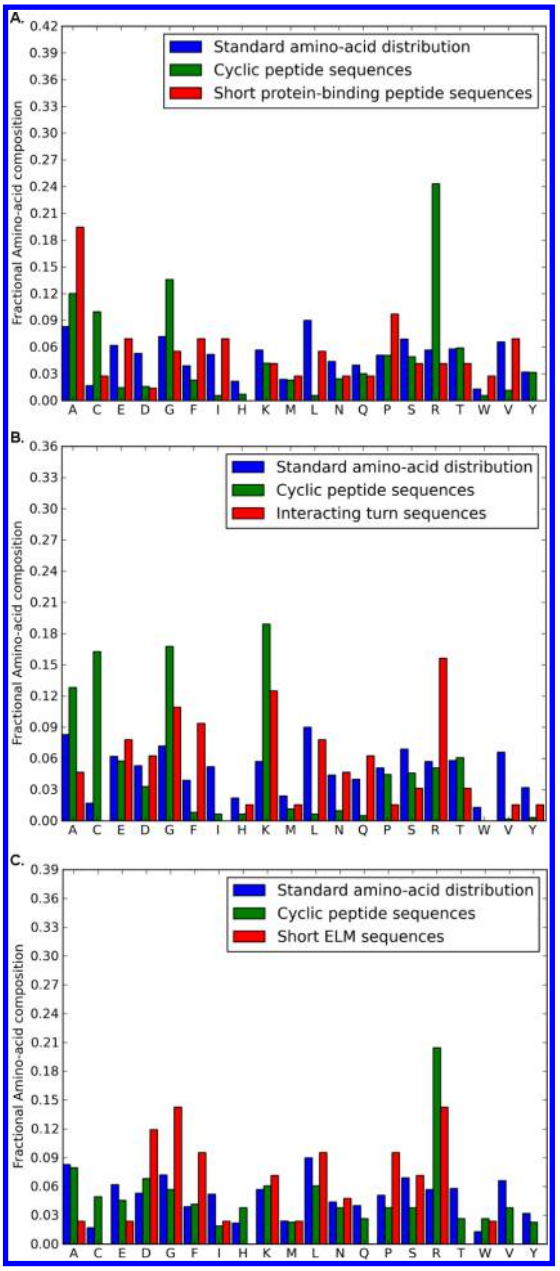


Figure 2. Amino-acid distributions of top reference structures and cyclic peptide hits. Amino-acid frequencies from the best 200 reference structure-cyclic peptide pairs for each class of structure are shown below. All matches had Tversky_ref scores of at least 0.65 and Tanimoto scores of 0.30. **A.** Top protein−peptide interfaces and accompanying cyclic peptide hits. **B.** Top interface turn structures and accompanying cyclic peptide hits. **C.** Top short linear motif structures and accompanying cyclic peptide hits. Standard protein amino-acid distribution taken from McCaldon and Argos.[55]

peptide interactions, the number of cyclic peptides mimicking each structure varied considerably. The results are dominated by cyclic peptide hits against pharmacophore models of turns at protein−protein interfaces, with protein−peptide interactions also well represented. 91.8% (25,304 out of a total of 27,554) of cyclic peptide conformer hits are against seven turn structure interfaces, with the rest being against seven protein−peptide interfaces. In contrast, there are no short linear motif structures represented in Table 3. Short linear motif structures matched by a cyclic peptide hit with a Z-score above 5 are shown in Table 4.

Pharmacophore Feature Spacing Influences Matching Scores. To explore how the geometry of a pharmacophore influences how well it is mimicked by cyclic peptides, the average spread of each pharmacophore model was calculated.

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

Table 3. Normalized Disulfide-Bonded Cyclic Peptide Hits to PPI Structures[a]

PDB ID	peptide chain ID	peptide UniProt accession	peptide residues	description	hits
1DSZ	B	P19793	1281−1286	turn on surface of retinoid X receptor (RXR)- α at interface with with retinoic acid receptor α (RAR)	3
1HAI, 1LHD, 1UCY, 1UVT	L	P00734	6−11	turn on surface of α-thrombin small subunit at interface with α-thrombin large subunit	23,981
1XK4	C	P06702	46−51	turn on surface of calgranulin B at interface with calgranulin A	140
1UKV	Y	P01123	66−71	turn on surface of GTP-binding protein Rab1A at interface with RabGDI	769
1K54	A	P14489	151−156	turn on surface of bacterial beta lactamase OXA-10 homodimer	4
3C4O	A	P0AD64	108−115	turn on surface of bacterial beta-lactamase SHV-1 in complex with beta-lactamase inhibitory protein	6
3HR4	E	P35228	679−685	turn on surface of nitric oxid synthase	403
1JX2	C	N/A	N/A	protein-binding peptide. Synthetic linker peptide in Myosin II heavy chain/dynamin A GTPase complex	1
1KZO	C	P01116	183−188	protein-binding peptide, farnesylated K-Ras4B peptide at interface with farnesyltransferase A and B	2
2V3H	I	P28503	56−65	protein-binding peptide, hirudin IIA α-thrombin inhibitor at interface with thrombin heavy chain	3
1OYT	I	P28506	55−65	protein-binding peptide, hirudin IIB α-thrombin inhibitor at interface with thrombin heavy chain	958
1PIP	B	N/A	N/A	protein-binding peptide, synthetic thiol protease inhibitor in complex with papain	13
1SDX	E	P24627	681−685	protein-binding peptide lactotransferrin in complex with lactotransferrin	1,269
3DRK	B	P0C0P6	70−74	protein-binding peptide neuropeptide S in complex with protein OppA	2

[a]Details of protein−protein and protein−peptide 3D structure and sequences with specific cyclic peptide conformer hits (all structures where a specific hit is a cyclic peptide conformer with a Z-score >6.0).

Table 4. Normalized Cyclic Peptide Hits to Short Linear Motif Structures[a]

PDB ID	short linear motif	hits
2FOO	LIG_USP7_1	1
1KYU	LIG_AP2alpha_2	34
2JXC	LIG_EH_1	31

[a]Counts of cyclic peptide conformer hits against short linear motif structures with a Z-score >5.0.

Figure 4 shows the correlation of average pharmacophore matching score and average pharmacophore feature spread. Tables 3 and 4 show that the largest proportion of high-scoring normalized hits correspond to turns at protein−protein interfaces, followed by protein-binding peptides. Figure 4 indicates that there is a correlation between tighter pharmacophore spacing and higher average pharmacophore scores against cyclic peptides. Furthermore, Figure 4 shows that every pharmacophore structure with an average feature spacing of under 5 Å is either a protein-binding peptide or a turn at a protein−protein interface. This may partly explain the results seen in Tables 3 and 4 where turn structures predominate.

The looping geometry of a turn is geometrically quite similar to a cyclic peptide. Cyclic peptides have previously been successfully used as scaffolds[27] to mimic turns, as compared to protein-binding peptides and short linear motifs, which often bind in an extended conformation[28] by β sheet addition.

Pharmacophore Matching Identifies Key Interaction Features. To validate that the pharmacophore models generated for this study captured the essential binding features of the protein−protein interfaces involved, the HotPOINT[29] PPI hot spot prediction server was used to predict the key amino acids involved in the top scoring interfaces listed in Table 3. Predicted hot spot locations and hydrogen bonds observed in the crystal structure were then mapped to the generated pharmacophore models, to show that pharmacophore features in structures with high-scoring hits occurred in hot spot regions, and that cyclic peptide hits tended to maintain hydrogen bonds observed in the PDB crystal structure.

The LFEKK[SQ] (human/bovine polymorphisms respectively) turn at residues 6−11 of the α-thrombin light chain at the surface of the thrombin light-chain/heavy-chain interface accounts for many of the cyclic peptide matches listed in Table 3. Figure 3A illustrates the pharmacophore model of this turn and the predicted hot spots for the light-chain heavy-chain interaction. The figure shows a pair of hydrogen bond acceptors on the side chain of Glu 343 and a peptide backbone hydrogen-bond donor at Leu 341 as key interaction features, both of which are predicted to be interaction hot spots. Hotspot residues were predicted for PDB structure 1LHD, representing the interaction between the thrombin light and heavy chain. Figure 3B shows the top cyclic peptide hit to this interface, CEPQTC in place of the thrombin light chain, aligned to the light-chain turn pharmacophore model. The cyclic peptide maintains the key hydrogen bond, formed between the heavy-chain Trp-207 and light-chain Glu-8 in the crystal structure.

Supplementary Figures S2 and S4 show a similar analysis carried out for a hit listed in Table 3 against a protein-binding peptide (A K-ras4B peptide in complex with Farnesyltransferase), and for a short linear motif hit listed in Table 4 (a LIG_EH_1 structure), suggesting that automatically generated pharmacophore models are capturing key interaction points in all three of these diverse structures.

Hits at Druggable Interfaces. A number of the *in silico* leads may be of utility in further development of experimental data regarding druggable protein interfaces.

In the 1DSZ complex, the reference structure is a turn on the retinoid X receptor α, which is a nuclear receptor regulating gene expression in various biological processes related to histone acetylation, chromatin condensation, and transcriptional suppression. It is associated with several diseases including cancer.[30] In the case of the K-Ras4B peptide in farnesyltransferase (PDB: 1KZO), farnesylation of Ras

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

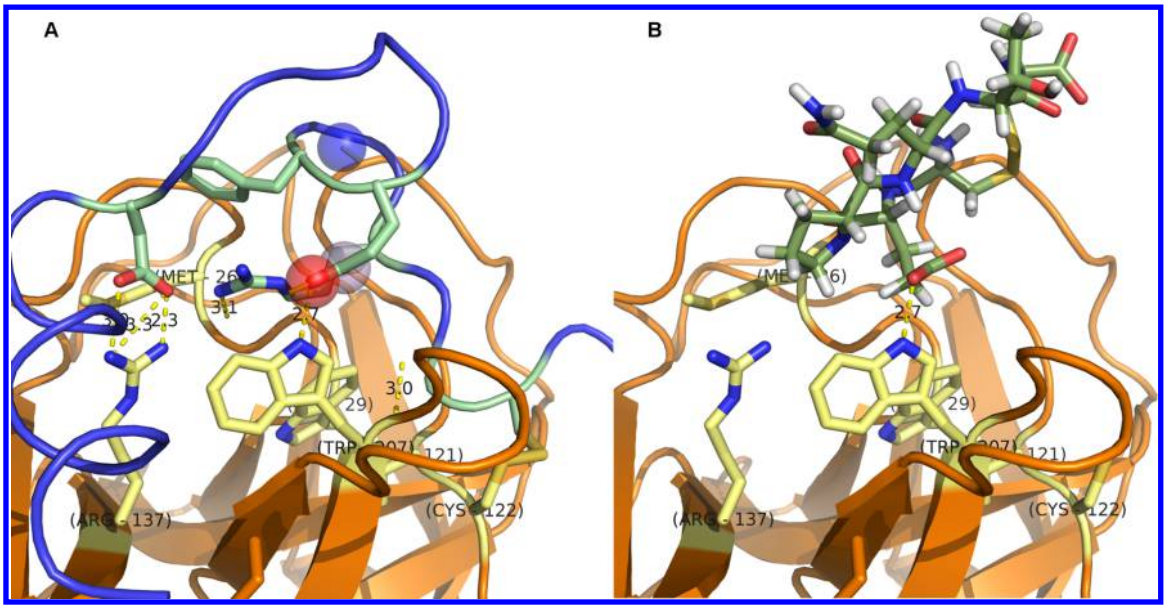


Figure 3. Thrombin light-chain heavy-chain key interactions captured by a cyclic peptide hit thrombin heavy chain in orange and light chain in blue both images. Key residues are labeled with amino-acid code and number. **A.** Shows the pharmacophore model developed for a thrombin turn in the light chain of PDB structure 1LHD. The red transparent sphere represents negative charge, pink represents a hydrogen bond acceptor, and the blue transparent sphere represents a hydrogen bond donor. The predicted hot spots in the thrombin heavy-chain light-chain interaction, as predicted by the HotPOINT server, are shown, with light-chain predicted hot spot residues highlighted in green and predicted heavy-chain hot spots highlighted in yellow. **B.** Shows the top cyclic peptide hit to the thrombin light-chain turn, CEPQTC, aligned to the thrombin light-chain pharmacophore model, and maintaining the hydrogen bonding to heavy-chain Trp-207 that is also shown in part **A.** by light-chain Glu-8.

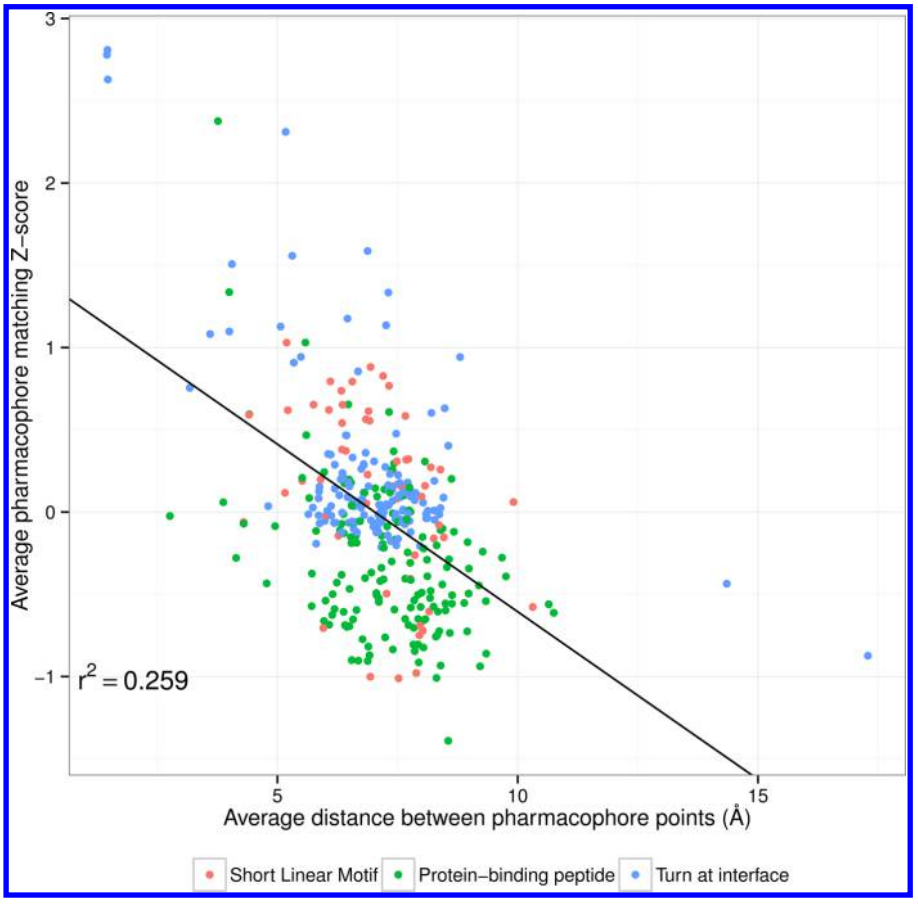


Figure 4. Most specifically binding cyclic peptides tend to target protein regions whose pharmacophore points are more tightly clustered. The spread of pharmacophore feature points within the target negatively correlates (r^2 = 0.259) with average matching normalized score (Z-score) against cyclic peptides, including turns at protein−protein interfaces (blue), protein-binding peptides (green), and short linear motifs (red).

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

oncoproteins is associated with 30% of human cancers, and blocking the peptide substrate site is a key mode of action in clinically active anticancer farnesyltransferase inhibitors.[31] PDB structure 1PIP contains a thiol protease inhibitor complexed with papain, a cysteine protease enzyme present in papaya. Papain-like lysosomal cysteine proteases are known drug-discovery targets.[32]

Two peptide-binding thrombin structures are also present among the top hits in Table 3, in structures 2V3H and 1OYT, with thrombin binding to hirudin peptides. Recombinant hirudin was the first ever therapeutic anticoagulant,[33] and hirudin peptidomimetics are in common use today.[34]

Experimentally Validating a Cyclic Peptide Thrombin Inhibitor. To see if our pharmacophore search method has revealed novel cyclic peptides with therapeutic potential, we focused on cyclic peptides predicted to mimic the hirudin C-terminus in the interaction between hirudin and thrombin (Table 3, PDB ID: 1OYT). This interface is illustrated in Figure 5. Thrombin is a serine protease which plays a critical role in the blood coagulation pathway,[35] including activation of platelets to form clots (thrombi) that reduce blood loss. When the layer of endothelial cells lining a blood vessel are damaged, thrombin becomes activated, which promotes platelet activation by activating several protease-activated receptors (PARs), such as PAR-1, -3, and -4 on the platelet surface.[36]

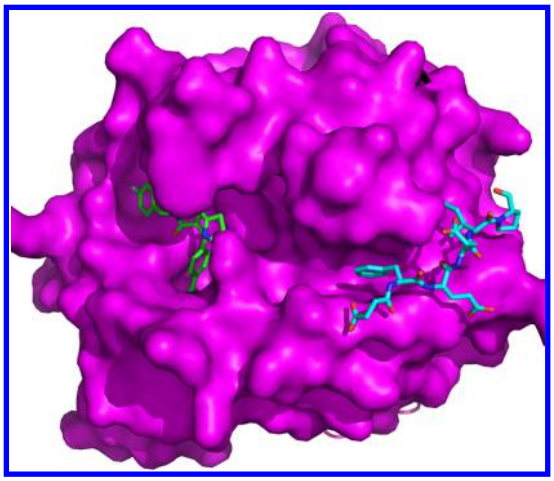


Figure 5. A hirudin C-terminal peptide shown to inhibit thrombin. Representation of protein data bank structure 1OYT.[56] The thrombin protein (magenta surface) is depicted with a modified hirudin terminal peptide (stick representation with blue carbons) bound to the PAR-1 recognition exosite-1, along with a small molecule inhibitor (stick representation with green carbons), at the protease active site. The conformation of the hirudin peptide bound to the PAR-1 recognition site is the template structure for the pharmacophore model used to screen cyclic peptides in this study.

Hirudin is a naturally occurring peptide in the salivary glands of medicinal leeches, that is an extremely potent inhibitor of thrombin.[37] PDB structure 1OYT contains a hirudin C-terminal peptide bound to the thrombin exosite-1, where thrombin binds to the platelet PAR-1 receptor. The full hirudin protein blocks both the thrombin exosite and active site,[38] but exosite-1 inhibitors alone have been shown to somewhat decrease the clotting activity of thrombin.[39] Pharmacophore matches to this region of hirudin are therefore candidate anti-thrombin compounds. Table 3 lists 958 cyclic peptides with a high normalized pharmacophore matching score against this interaction. To select a smaller number of cyclic peptides for testing, a second virtual screening step was used, employing the USRCAT[40] package, which combines shape matching with pharmacophoric features.

To test the potency of these peptides as potential thrombin inhibitors, a platelet aggregation assay was used. Platelet aggregation can be measured by the amount of light transmitted through a stirred volume of platelets. Unactivated platelets will appear as a milky, opaque suspension. When thrombin is added to the stirred mixture, platelets rapidly coagulate and drop out of the suspension, leaving a clear liquid. When thrombin is inhibited with hirudin, no activation effect is seen. Seventeen cyclic peptides identified with a Z-score of >6.0 and a USRCAT score of >0.55 were synthesized, and these were tested in a thrombin-activated platelet aggregation assay, using 4 healthy platelet donors who had not ingested NSAIDs in the 10 days prior to blood drawing. The full list of cyclic peptide sequences tested is shown in Table 5.

Table 5. Cyclic PeptIdes Tested for anti-Thrombin Activity[a]

sequence	Pharao Tversky_ ref Score	USRCAT score
CFGKQC	0.917	0.761
CFGVEC	0.921	0.750
CFNQDC	0.958	0.749
CFPKEC	0.974	0.812
CFQLDC	0.933	0.766
CFVKQC	0.960	0.766
CEPKFC	0.905	0.794
CEQTWC	0.926	0.750
CFDKQC	0.942	0.776
CFDMQC	0.922	0.779
CDQNFC	0.914	0.762
CDDKFC	0.904	0.768
CNENWC	0.930	0.751
CKPEYC	0.927	0.548
CYNDRC	0.912	0.567
CYRGQC	0.933	0.554
CRWTKC	0.932	0.551

[a]This table shows the sequences of all peptides tested for anti-thrombin activity and their similarity scores to the hirudin peptide in PDB: 1OYT calculated with both Pharao and USRCAT.

From this set of cyclic peptides, we were able to identify one peptide with the sequence CEPKFC as having anti-thrombin activity. While only one peptide out of 17 has been observed as showing antithrombin activity, showing activity in this assay is a high bar to clear, as it requires binding to thrombin strongly enough to effectively compete with the hirudin-thrombin binding interaction which possesses subpicomolar[41] (K_d of 0.2 pM) affinity. Therefore, more than one of the peptides could be binding thrombin, but only one peptide binds thrombin effectively enough to significantly inhibit hirudin binding to thrombin.

Figure 6 shows the effects of thrombin, thrombin plus hirudin, and thrombin plus the cyclic peptide CEPKFC on the aggregation of human platelets. Addition of CEPKFC results in a significant lowering of the total amount of thrombin aggregation ($P = 0.05$ in a one tailed t-test), as well as delaying aggregation compared to thrombin alone at every time point shown in Figure 6. It has a much smaller effect than the very potent hirudin peptide. This is consistent with our pharmacophore screen, as these peptides match only the

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

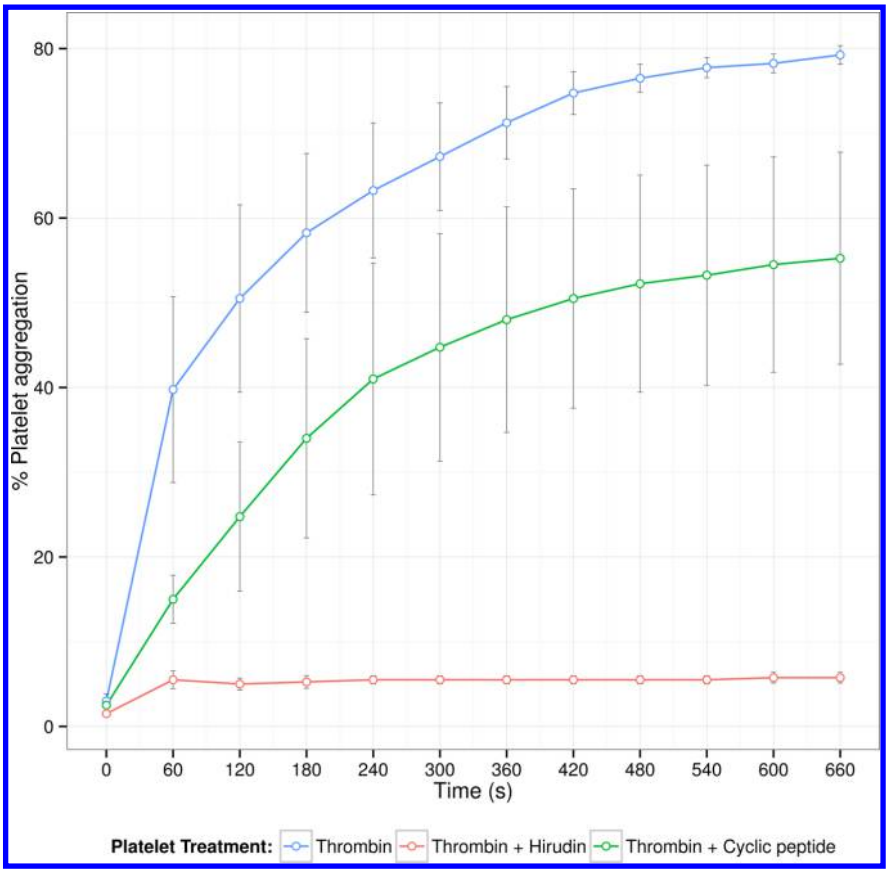


Figure 6. Thrombin induced platelet activation. The effect of adding hirudin and the cyclic peptide CEPKFC on thrombin-induced platelet aggregations over 4 donors. Thrombin alone is in blue (positive control), thrombin mixed with hirudin is in red (negative control), and thrombin with the cyclic peptide is in green. Error bars indicate standard error.

portion of hirudin binding to the exosite-1 thrombin pocket which, unlike the full hirudin protein, does not block the proteolytic active site.

Thrombin is of critical clinical importance for the central role it plays in many heart conditions, such as stroke[42] and atherosclerosis,[43] and there exist a large class of anticoagulant drugs[44] developed to modulate its activity. From Figure 6, it can be seen that the cyclic peptide is able to both delay and reduce the final amount of thrombin-mediated aggregation. The effects of the cyclic peptide are quite variable between donors; however, this is likely due to the high variability in platelet activity between donors. There is less variability in the positive and negative controls due to the very strong activation and inhibition effects of thrombin and hirudin, respectively. However, weak platelet inhibition activity may be a therapeutically desirable effect - it has been been previously observed that antagonists of the glycoprotein GPIIb/IIIa receptor, part of the platelet activation pathway, have failed clinical trials due to their excessive inhibition of platelets, which can lead to bleeding complications.[45,46] A 15-mer single stranded DNA aptamer has previously been identified that binds the thrombin exosite described here,[47] but a cyclic peptide lead may be an easier lead compound from which to generate more potent and more "drug-like" compounds.

Surface Plasmon Resonance (SPR) experiments were performed to provide a quantitive measure of the potency of this peptide. Figure 8 shows the results of SPR studies on the behavior of the CEPKFC peptide binding to thrombin at a range of concentrations. Supplementary Table S1 lists the data underlying this plot. The data indicates an approximate dissociation constant (K_d) of 545 μM, which means that this is a relatively low affinity binder to thrombin. However, despite this low binding affinity, the peptide does have a significant influence on the coagulation pathway *in vitro* (Figure 6) and demonstrates the ability of computational screening to identify a cyclic peptide that act against protein−protein interactions.

Other Cyclic Peptide Structures. The peptides examined in this study have been disulfide-bonded cyclic peptides of the form Cys-X-X-X-X-Cys and Cys-X-X-X-Cys. Disulfide-bonded peptides were chosen as the most cost-effective to chemically synthesize and experimentally test, and screening combinatorial multiconformer libraries of peptides with 3 or 4 variable positions results in large but feasible numbers of structures. However, as mentioned in the Introduction, many other types of constrained cyclic peptide exist, and it would be desirable to explore the potential for identifying PPI-inhibitor cyclic peptides from a wide range of peptide backbone lengths and constraint types.

To explore the potential of various types and sizes of cyclic peptides to act as PPI inhibitors, a cyclic peptide diversity library was generated using CycloPs. This library consisted of 10,000 unique randomly selected peptides with sequence lengths of 6 to 9 amino acids and a randomly chosen appropriate cyclization method (head−tail bonded, disulfide bonded, side-chain to side-chain bonded, or side-chain to N- or C-terminal bonded). This diversity library was then subjected to a conformation generation step identical to that applied to the Cys-X-X-X-X-Cys and Cys-X-X-X-X-Cys libraries. Pharma-

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

cophore screening was then carried out on the multiconformer diversity library against each reference pharmacophore structure referred to in Table 1.

Figure 7 shows the distribution of the pharmacophore matching scores across all reference structures matched to each cyclic peptide in the diverse cyclization library. Comparing the distributions seen in Figure 7 to those in Figure 1, it can be seen that the shape distribution of scores for each class of reference structure (short linear motifs, loops, and protein-binding peptides) has a very similar shape in both the original disulfide bonded libraries and the diverse cyclization library. This suggests that there may be some commonality in the type of protein−protein interactions that various cyclic peptide structures are suited to inhibit.

Table 6 identifies the reference structures that have large numbers of normalized cyclic peptide matches from the diverse cyclization library. It is very noticeable that there is a much larger number of reference structures in Table 6 than in Table 3, despite the diverse cyclization library containing far fewer structures than the disulfide bonded libraries tested above. The diverse cyclic peptide library hits also include several structures of potential clinical interest, such as monoclonal antibody loops that bind HIV proteins. This suggests that cyclic peptide libraries that include a broad array of sequence lengths and cyclization methods have the potential to reveal screening "hits" against a much broader set of target structures.

Conclusions. We have presented for the first time a high throughput method of identifying both protein−protein interactions suitable for inhibition by cyclic peptides and the accompanying cyclic peptides that represent promising lead compounds. Our method represents a simple approach for very rapidly screening a large virtual library against a wide variety of protein−protein interactions. Since it can quickly identify protein−protein interactions that seem to be druggable by a particular virtual library, it is an easy way of reducing the search space of proteins when looking to modify a particular biological pathway. This broader kind of screening approach offers an alternative to focusing on a single important PPI. For example, a broad screening approach could be used to identify all proteins involved in a biological pathway of interest: first identify proteins that have been experimentally shown to interact and have available structural data; then develop pharmacophore models of key interface hotspots; and finally screen against a diverse compound library to identify both druggable PPIs and the most common chemotypes they bind. This kind of analysis can be done computationally on thousands of possible PPIs and millions of compounds using moderate computational resources.

Pharmacophore screening represents a simplification of the complex processes and molecular properties that guide target-ligand binding, and a key challenge for any such screening is to demonstrate that it has captured the most essential parts of the interaction, giving the study the power to accurately identify bioactive compounds and reject unsuitable compounds. We have shown that our method is able to capture essential elements of the protein−peptide and protein−protein interactions, by comparing our pharmacophore models to predicted interaction hotspots, followed by further validation with experimental work on thrombin inhibition. Our method has also highlighted known and previously unidentified druggable protein−protein and protein−peptide interactions as candidates for modulation by cyclic peptides. Thus, developing interaction pharmacophore models to identify peptidomimetics of surfaces can identify relevant interfaces that have proven to be druggable, as well as being high throughput enough to search for PPIs that have not been previously considered drug targets.

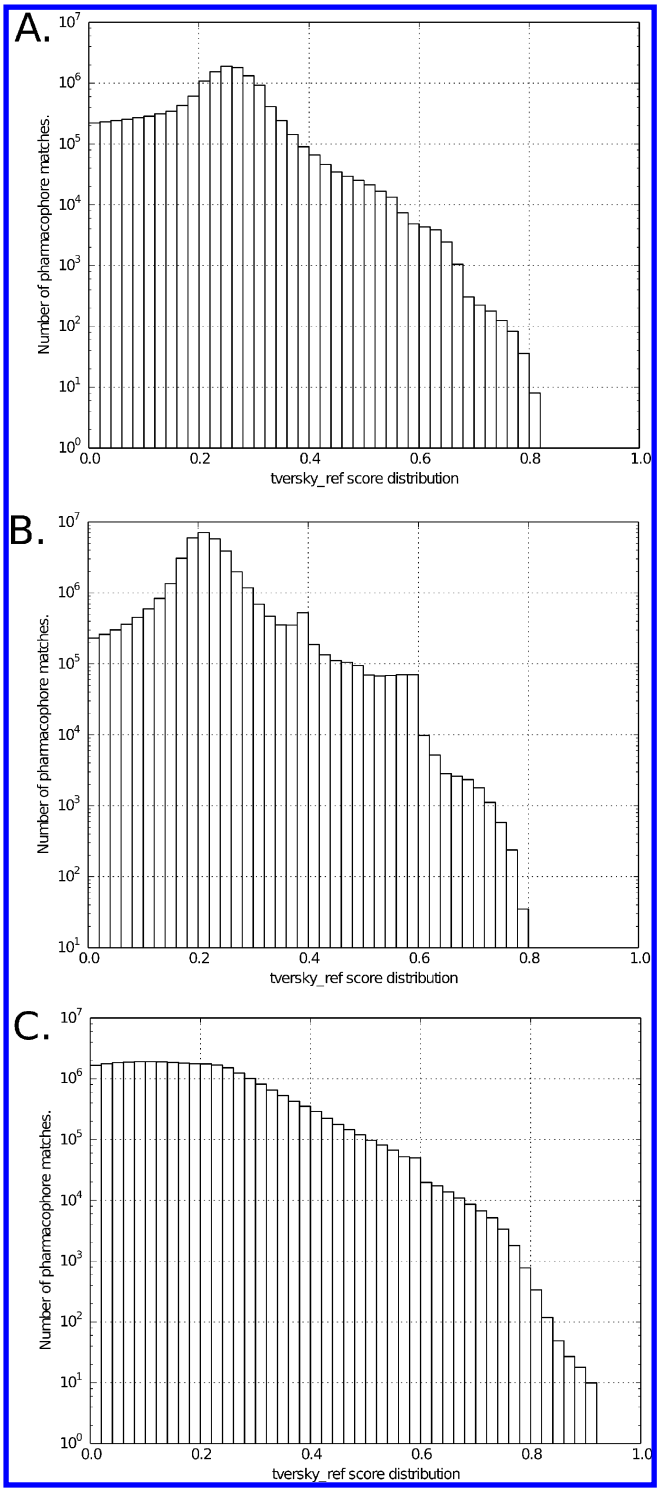


Figure 7. Pharmacophore score distribution of diverse cyclic peptide library. A diverse multiconformer library of 10,000 cyclic peptides with various forms of cyclization and sequence lengths of between 6 and 9 residues were pharmacophore matched to each reference structure prepared for this study (Table 1). Tversky_ref distributions of cyclic peptide pharmacophore matching scores for **A.** short linear motif structures, **B.** turn structures at protein−protein interfaces, and **C.** protein−peptide structures are shown here.

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

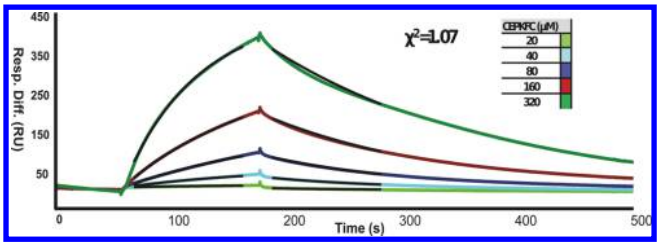


Figure 8. SPR kinetic analysis of cyclic peptide CEPKFC to immobilized thrombin. Sensorgram of molar series of CEPKFC peptide (20−320 M) binding to thrombin coated SPR biosensor chip surface. Curves represent the response difference in binding of experimental flow cell (FC2) minus control uncoated flow cell (FC1), with (FC1-FC2) represented in response units (RU). Association (ka) and dissociation (kd) rates were calculated using the Langmuir model of 1:1 analyte:ligand binding. Accuracy of model fitting was calculated using Pearsons chi-squared test, 2 = 1.07.

A key advantage of using simple cyclic peptides based on disulfide bonds is their low cost. This enables progression from virtual screening, to medium throughput biological screening of cyclic peptides to identify lead compounds, at micromolar or better potencies.

However, it seems clear from testing a more diverse set of structures, that only testing short cyclic peptides is a limiting factor in discovering protein−protein interactions amenable to modulation. To reveal a larger number of PPI structures targetable by cyclic peptides and to further improve the physical properties and binding affinity of the lead compounds, the use of different cyclization strategies, such as head−tail binding to avoid oxidation of disulfide bonds, should also be employed. In addition the use of non-natural amino acids to add rigidity to the macrocycle, improve side-chain binding, or to improve bioavailability can also be investigated. There exist limits to this approach: it is known that protein−protein interfaces are not rigid but constantly shifting and forming transient pockets, with

Table 6. Normalized Hits from a Diverse Cyclic Peptide Library against PPI Structures[a]

PDB ID	peptide chain ID	peptide UniProt accession	peptide residues	description	hits
2MFN, 1MFN	A	P11276	2181−2183	short linear motif LIG_RGD which binds to several types of Integrin membrane proteins	45
2JQ8, 1S4G, 1SSU	A	P04004	64−66	short linear motif LIG_RGD which binds to several types of Integrin membrane proteins	28
1TTF, 1TTG	A	P02751	1524−1526	short linear motif LIG_RGD which binds to several types of Integrin membrane proteins	26
1R0P, 3A4P, 3C1X, 3F82, 1R1W, 3CTJ, 3CE3	A	P08581	1356−1359	short linear motif LIG_SH2_GRB2 - this binds to proteins containing a Src homology 2 domain	24
1QG1, 1WCP	I	P29353	349−352	short linear motif LIG_SH2_GRB2 - this binds to proteins containing a Src homology 2 domain	3
1FMK, 2SRC	A	P12931	530−533	short linear motif LIG_SH2_SRC, this binds to proteins containing a Src homology 2 domain	3
2XRA, 1Q1J, 1TZG, 1RZ8	H	N/A	N/A	loop on monoclonal antibody binding to HIV gp41 protein	10
1M1N, 1QH1, 1QGU, 1QH8	A	P07328	207−212	loop on nitrogenase molybdenum−iron protein alpha chain involved in interaction with the beta chain	5
1JEQ, 1JEY	A	P12956	352−356	short linear motif LIG_HP1_1: motif in proteins that interact with the chromoshadow domain of heterochromatin-binding protein 1	4
1LF8	D	Q9NZ52	100−105	loop on ADP-ribosylation factor binding protein GGA3 binding cation-independent mannose-6-phosphate receptor	2
1Q1J	I	N/A	N/A	loop on human monoclonal antibody bound to HIV gp120 V3 protein	5
2JBG	B	Q47112	463−469	loop on *E. coli* colicin E7 protein binding to colicin E7 immunity protein	1
1K55	D	P14489	151−156	loop involved in beta lactamase OXA-10 homointeraction	1
1CDF	A	P25942	250−253	short linear motif LIG_TRAF2_1: binding motif for tumor necrosis factor receptor	6
1SM3	H	P15941	61−66	loop on antibreast cancer antibody	1
2MIP	E	N/A	N/A	synthetic HIV-2 protease inhibitor protein-binding peptide	7
3M4W	C	P0AFX7	120−125	loop on sigma-E factor regulatory protein rseB bound to sigma-E factor negative regulatory protein	1
3H8D	C, A	P18429	1200−1205	loop on xylanase inhibitor protein bound to Endo-1,4-beta-xylanase A	3
2BS2	C	P17413	247−252	loop on quinol-funmarate reductase diheme cytochrome *b* subunit C	2
2WEV, 3BHU, 1OIY, 2G9X	B	P20248	197−202	loop on cyclin A-2 bound to cell division protein kinase 2	5
2AKA	A	P08799	19−25	loop on myosin II heavy chain in complex with dynamin	1
1LVM	C	P04517	302−310	peptide substrate for tobacco etch virus protease	1
2G38	B	Q79FE1	50−55	loop on PPE family protein bound to PE family protein	1
3H0L	A	O66610	361−367	loop glutamyl-tRNA(Gln) amidotransferase subunit A	1
1AXT	H	P01865	61−66	mouse antibody loop in homointeraction	2
1OY3	B	Q04207	296−301	loop on transcription factor p65 bound to transcription factor inhibitor I-kappa-B-beta	8
3GZE	X	Q9FPQ6	1−10	peptide substrate for algal prolyl 4-hydroxylase	2
2H64	C	P27040	17−22	loop on Acvr2b protein in the BMP-2 complex	1

[a]This table describes all the reference structures that are pharmacophore matched to one or more cyclic peptides from a 10,000 member diversity library with a normalized Z-score of 6.0 or greater.

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

the regions amenable to binding an inhibitor even more likely to show such flexibility.[48] Ultimately, a more complete approach would involve incorporating this surface flexibility into pharmacophore models, with molecular dynamics simulations or NMR measured protein surfaces being used to generate an ensemble of surfaces which can be translated into a pharmacophore model.

In conclusion, while disulfide bonded cyclic peptide are less likely to constitute the final therapeutically useful macrocyclic compound, the advantage of rapid computational screening with low cost initial lead generation is an attractive first step in a screening process.

METHODS

Ethics Statement. Human platelets used in this study were obtained from healthy volunteers that were free of medication. Patients were provided with a detailed information sheet and gave written consent under RCSI Ethical Review Protocol approved by the RCSI Research Ethics Committee.

Reference Preparation. The portions of known 3D structures extracted from the Protein Data Bank which represent portions of protein−protein/peptide interactions were referred to as reference peptides. These comprised short linear motif instances, protein-binding peptides, and protein turns located at protein−protein interfaces.

Short linear motif (SLiM) definitions were extracted from the Eukaryotic Linear Motif[21] (ELM) database using the available SOAP Web service. Only motifs of class "LIG", representing ligand binding motifs, were used. For each "LIG" motif definition, ELM instances corresponding to UniProt protein IDs were examined and searched for the existence of PDB structures which covered the portion of the protein containing the short linear motif. ELM structures found in PDB structures were extracted using the PDB module from the BioPython[49] project.

The PepX[22] database was used to identify short protein-binding peptides. All PDB structures containing a protein-binding peptide chain of 5 residues in length or shorter were downloaded, and the protein-binding peptide chain was extracted using BioPython.

To identify turn structures at protein−protein interfaces, protein−protein interaction structures were extracted from the PDB. The PDB was searched for all structures with a resolution of 2.5 Å or better, containing at least two protein chains in the biological assembly. The DSSP[50] secondary structure assignment software was then used to identify each turn structure in each protein−protein complex. PyMol[51] was then used to test if each identified turn was located at a protein−protein interface region. A turn was considered part of an interface if any part of the turn was within 5 Å of another protein. Each interface turn was then extracted from the overall PDB structure using PyMol.

Reference peptide structures were protonated at physiological pH (7.4) using OpenBabel before pharmacophore feature identification was carried out using Pharao[26] with default settings. For reference structures taken from a protein−protein interface (all structures, excepting certain short linear motifs in single-protein PDB structures), exclusion spheres were added using the pharao PyMol plugin, which creates an exclusion volume around each atom from a PDB chain other than the chain containing the reference structure that is within 5 Å of the reference structure. Examples of high and low-scoring Pharao pharmacophore matches are shown in Supplementary Figure S3. A PyMol script was used to identify and remove pharmacophore features that fell outside a distance cutoff of 3.5 Å from a protein−protein or protein−peptide interface.

Virtual Library Preparation. Single conformation 3D cyclic peptide libraries were prepared using the CycloPs[52] virtual peptide library generation software. CycloPs generates virtual cyclic peptide libraries combinatorially from amino-acid SMILES string building blocks. It is capable of applying empirical rules to peptide structures to predict peptide synthesizability and water solubility, as well as using the UFF force field to produce 3D peptide libraries, although it does not perform an exhaustive exploration of potential peptide conformations. Cyclic peptides were generated combinatorially using the "Filter out peptides that break synthesis rules" option, to remove virtual peptides that may be difficult to chemically synthesize. After library generation, libraries were protonated at physiological pH (7.4) using OpenBabel. Conformer generation was performed with the RDKit[53] Python module. The RDKit is an open-source cheminformatics toolkit which implements a large set of methods for reading, writing, depicting, modeling, searching, and screening chemical libraries. A modified version of the method described by Ebejer[54] et al. was used, where 20 random conformers were generated for each peptide, before minimization using the UFF force field. The energy of each cyclized peptide was then calculated, and the peptides sorted from lowest energy to highest energy. Taking the lowest energy conformer first, the root mean squared deviation (RMSD) to every other conformer was calculated, and any conformer with an RMSD within 0.5 Å of the low energy conformer was deleted. This procedure was repeated for each conformer, from lowest to highest energy, to create a panel of diverse, low energy conformers. The figure of 20 conformations to be tested per cyclic peptide was chosen to strike a balance between the requirement to cover a large amount of the available conformational space and the limits of available computational resources. It has been shown[54] that applying this method to molecules with 13 rotational bonds will result in an average RMSD of about 2 Å between conformers. While each peptide may not have its conformational space exhaustively explored, the method is designed to produce a broad and diverse sampling of the available conformational space.

Pharmacophore Matching. Pharmacophore matching was carried out using Pharao and the Irish Centre for High-End Computing (ICHEC) Stokes cluster. Pharmacophore matching scores were based on the overlapping feature volume of pharmacophore features using the Tversky_ref and Tanimoto scores where

$$\text{Tversky\_ref} = \frac{V_{\text{shared}}}{V_{\text{reference}}}$$

and

$$\text{Tanimoto} = \frac{V_{\text{shared}}}{V_{\text{reference}} + V_{\text{cyclic\_peptide}} - V_{\text{shared}}}$$

Normalization. Pharmacophore matching score normalization was carried out by calculating the average and the standard deviation of the set of Tversky_ref scores for each individual cyclic peptide, including all conformers, and producing a normalization statistic

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

$$Z_{score} = \frac{\text{Tversky\_ref} - \mu_{\text{Tversky\_ref}}}{\sigma_{\text{Tversky\_ref}}}$$

Calculating Pharmacophore Spread. The average pharmacophore spread distance was calculated as the average of the Euclidean distance over all pairs of pharmacophore features, where the Euclidean distance between two 3-dimensional points was defined as

$$d(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Platelet Aggregations. *Platelet Preparation.* Blood was obtained from healthy donors who had not taken medications known to affect platelet function for at least 10 days. Blood was collected by venipuncture through a 19-gauge butterfly needle without a tourniquet, to avoid platelet activation. For the preparation of washed platelets, blood was collected into Acid-Citrate-Dextrose (ACD: 38 mM citric acid, 75 mM sodium citrate, 124 mM D-glucose) as anticoagulant (15% vol/vol). Blood was centrifuged at 150 g for 10 min at room temperature, to separate platelet-rich plasma. Platelet rich plasma was extracted and acidified to pH 6.5 with ACD, and prostaglandin E_1 (PGE_1:1 μM) was added to avoid platelet activation during centrifugation. Platelets were pelleted by centrifugation at 750 g for 10 min. The supernatant was removed, and the platelet pellet was resuspended in JNL buffer [130 mM NaCl, 10 mM sodium citrate, 9 mM $NaHCO_3$, 6 mM D-glucose, and 0.9 mM $MgCl_2$, 0.81 mM KH_2PO_4, and 10 mM Tris, pH 7.4] to a concentration of 300 × 10^3/mL. The platelets were allowed to sit for 45 min to allow the effects of PGE_1 to wear off and supplemented with 1.8 mM $CaCl_2$ 10 min prior to aggregation.

Platelet Aggregation Assay. Platelet aggregations were performed using the BioData PAP-8E Platelet Aggregation Profiler. Aggregations were carried out stirred at 37 °C, with a total aggregation volume of 250 μM. Bovine thrombin was obtained from Sigma-Aldrich (Product T6634). Disulfide bonded cyclic peptides were ordered from ChinaPeptides and prepared as 1 mM stock solutions in DI water.

Both hirudin and peptides were preincubated at room temperature with thrombin for 30 min to allow time to bind to and inactivate thrombin. Washed platelets were added to the aggregometer and stirred for 2 min prior to adding the thrombin/hirudin or thrombin/cyclic peptide mixture. The aggregometer was calibrated with plain JNL buffer representing 100% aggregation. Thrombin alone was used as a positive control. Final aggregation well concentrations were 0.05 Units/mL thrombin (in all wells), 20 Units/mL hirudin (in the negative control well), and 100 μM cyclic peptide (in experimental wells) in a total aggregation volume of 250 μL.

Surface Plasmon Resonance Experiments. The SPR experiments were performed using a Biacore T200 system (GE Healthcare) with four flow cells. Lyophilized thrombin was dissolved in flow buffer to a concentration of 10 g mL^{-1} and covalently coupled onto the carboxymethylated dextran matrix of CM5 sensor chips using standard amine coupling. The flow buffer contained 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% surfactant P20. The chip was activated with an injection of 70 μL of a mixture of 0.05 M EDC and 0.2 M NHS, followed by the injection of approximately 70 μL of 10 μg mL^{-1} thrombin in 10 mM sodium acetate, pH 4.0 in one flow cell to yield a response of 8500 RU. Remaining reactive groups in the channel were blocked with an injection of 100 μL of ethanolamine. Binding experiments were performed by injecting peptide for 120 s at a flow rate of 10 μL min^{-1}, at concentrations from 20 to 320 μM in the flow buffer over the chip surface to monitor association kinetics, followed by buffer flow for 300 s to measure dissociation kinetics. All binding experiments were performed in duplicate. Between experiments, the surface was regenerated by injecting a 60 s pulse of 10 mM glycine pH 1.7.

■ ASSOCIATED CONTENT

Supporting Information

Supporting Figures S1 (correlation between Z scores and number of pharmacophore features), S2 (farnesyltransferase protein−peptide interaction with K-Ras4B peptide and a cyclic peptide hit), S3 (example of a good and a bad pharmacophore match), and S4 (short linear motif LIG EH 1 from PDB structure 2JXC aligned with cyclic peptide hit CELLSC) and Tables S1 (SPR kinetic analysis of cyclic peptide CEPKFC to immobilized thrombin) and S2 (list of structures containing protein−peptide interfaces, short linear motifs, and turns at protein-protein interfaces used in pharmacophore matching studies). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: denis.shields@ucd.ie.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Science Foundation Ireland (grant 08 IN.1 B1864) for funding this work, along with the Irish Centre for High-End Computing for providing compute time on their cluster.

■ REFERENCES

(1) Lo Conte, L.; Chothia, C.; Janin, J. The Atomic Structure of Protein-Protein Recognition Sites. *J. Mol. Biol.* **1999**, *285*, 2177−2198.
(2) McGovern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K. A Specific Mechanism of Nonspecific Inhibition. *J. Med. Chem.* **2003**, *46*, 4265−4272.
(3) Mullard, A. Protein-Protein Interaction Inhibitors Get into the Groove. *Nat. Rev. Drug Discovery* **2012**, *11*, 173−175.
(4) Wells, J. A.; McClendon, C. L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, *450*, 1001−1009.
(5) Metz, A.; Pfleger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K.-H.; Gohlke, H. Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein-Protein Interface. *J. Chem. Inf. Model.* **2012**, *52*, 120−133.
(6) Schreiber, S. L.; Crabtree, G. R. The Mechanism of Action of Cyclosporin A and FK506. *Immunol. Today* **1992**, *13*, 136−142.
(7) Sulyok, G. A.; Gibson, C.; Goodman, S. L.; Holzemann, G.; Wiesner, M.; Kessler, H. Solid-Phase Synthesis of a Nonpeptide RGD Mimetic Library: New Selective alphavbeta3 Integrin Antagonists. *J. Med. Chem.* **2001**, *44*, 1938−1950.
(8) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discovery* **2006**, *5*, 993−996.
(9) Fry, D. C. Protein-Protein Interactions As Targets for Small Molecule Drug Discovery. *Biopolymers* **2006**, *84*, 535−552.
(10) Therapeutics, E. Our Science and Technology. 2012. http://www.ensemblediscovery.com/sciTech/index.html (accessed Jan 30, 2015).

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

(11) Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. Synthetic Therapeutic Peptides: Science and Market. *Drug Discovery Today* **2010**, *15*, 40−56.
(12) Madden, M. M.; Muppidi, A.; Li, Z.; Li, X.; Chen, J.; Lin, Q. Synthesis of Cell-Permeable Stapled Peptide Dual Inhibitors of the p53-Mdm2/Mdmx Interactions via Photoinduced Cycloaddition. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1472−1475.
(13) Sanger, F.; Tuppy, H. The Amino-Acid Sequence in the Phenylalanyl Chain of Insulin. I. The Identification of Lower Peptides from Partial Hydrolysates. *Biochem. J.* **1951**, *49*, 463−481.
(14) Romeo, D.; Skerlavaj, B.; Bolognesi, M.; Gennaro, R. Structure and Bactericidal Activity of an Antibiotic Dodecapeptide Purified from Bovine Neutrophils. *J. Biol. Chem.* **1988**, *263*, 9573−9575.
(15) Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. Testing the Conformational Hypothesis of Passive Membrane Permeability Using Synthetic Cyclic Peptide Diastereomers. *J. Am. Chem. Soc.* **2006**, *128*, 2510−2511.
(16) Stephens, G.; O'Luanaigh, N.; Reilly, D.; Harriott, P.; Walker, B.; Fitzgerald, D.; Moran, N. A Sequence within the Cytoplasmic Tail of GpIIb Independently Activates Platelet Aggregation and Thromboxane Synthesis. *J. Biol. Chem.* **1998**, *273*, 20317−20322.
(17) van de Waterbeemd, H.; Gifford, E. ADMET in Silico Modelling: Towards Prediction Paradise? *Nat. Rev. Drug Discovery* **2003**, *2*, 192−204.
(18) Kennedy, T. Managing the Drug Discovery/Development Interface. *Drug Discovery Today* **1997**, *2*, 436−444.
(19) Burke, P. A.; DeNardo, S. J.; Miers, L. A.; Lamborn, K. R.; Matzku, S.; DeNardo, G. L. Cilengitide Targeting of alpha(v)beta(3) Integrin Receptor Synergizes with Radioimmunotherapy To Increase Efficacy and Apoptosis in Breast Cancer Xenografts. *Cancer Res.* **2002**, *62*, 4263−4272.
(20) Nguyen, C. M.; Harrington, R. A. Glycoprotein IIb/IIIa Receptor Antagonists: A Comparative Review of Their Use in Percutaneous Coronary Intervention. *Am. J. Cardiovasc. Drugs* **2003**, *3*, 423−436.
(21) Gould, C. M.; et al. ELM: The Status of the 2010 Eukaryotic Linear Motif Resource. *Nucleic Acids Res.* **2010**, *38*, D167−80.
(22) Vanhee, P.; Reumers, J.; Stricher, F.; Baeten, L.; Serrano, L.; Schymkowitz, J.; Rousseau, F. PepX: A Structural Database of Non-Redundant Protein-Peptide Complexes. *Nucleic Acids Res.* **2010**, *38*, D545−551.
(23) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(24) Vanhee, P.; Stricher, F.; Baeten, L.; Verschueren, E.; Lenaerts, T.; Serrano, L.; Rousseau, F.; Schymkowitz, J. Protein-Peptide Interactions Adopt the Same Structural Motifs As Monomeric Protein Folds. *Structure* **2009**, *17*, 1128−1136.
(25) Reorganizing the Protein Space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2012**, *40*, D71−75.
(26) Taminau, J.; Thijs, G.; De Winter, H. Pharao: Pharmacophore Alignment and Optimization. *J. Mol. Graphics Modell.* **2008**, *27*, 161−169.
(27) Arbor, S.; Kao, J.; Wu, Y.; Marshall, G. R. c[D-pro-Pro-D-pro-N-methyl-Ala] Adopts a Rigid Conformation That Serves As a Scaffold to Mimic Reverse-Turns. *Biopolymers* **2008**, *90*, 384−393.
(28) London, N.; Movshovitz-Attias, D.; Schueler-Furman, O. The Structural Basis of Peptide-Protein Binding Strategies. *Structure* **2010**, *18*, 188−199.
(29) Tuncbag, N.; Keskin, O.; Gursoy, A. HotPoint: Hot Spot Prediction Server for Protein Interfaces. *Nucleic Acids Res.* **2010**, *38*, W402−406.
(30) Mallipeddi, R.; Wessagowit, V.; South, A. P.; Robson, A. M.; Orchard, G. E.; Eady, R. A.; McGrath, J. A. Reduced Expression of Insulin-like Growth Factor-Binding Protein-3 (IGFBP-3) in Squamous Cell Carcinoma Complicating Recessive Dystrophic Epidermolysis Bullosa. *J. Invest. Dermatol.* **2004**, *122*, 1302−1309.
(31) Long, S. B.; Casey, P. J.; Beese, L. S. Reaction Path of Protein Farnesyltransferase at Atomic Resolution. *Nature* **2002**, *419*, 645−650.
(32) Turk, D.; Turk, B.; Turk, V. Papain-like Lysosomal Cysteine Proteases and Their Inhibitors: Drug Discovery Targets? *Biochem. Soc. Symp.* **2003**, 15−30.
(33) Greinacher, A.; Warkentin, T. E. The Direct Thrombin Inhibitor Hirudin. *Thromb. Haemostasis* **2008**, *99*, 819−829.
(34) Kikelj, D. Peptidomimetic Thrombin Inhibitors. *Pathophysiol. Haemostasis Thromb.* **2003**, *33*, 487−491.
(35) Davie, E. W.; Fujikawa, K.; Kisiel, W. The Coagulation Cascade: Initiation, Maintenance, and Regulation. *Biochemistry* **1991**, *30*, 10363−10370.
(36) Coughlin, S. R. Thrombin Signalling and Protease-Activated Receptors. *Nature* **2000**, *407*, 258−264.
(37) Markwardt, F. Development of Hirudin As an Antithrombotic Agent. *Semin. Thromb. Hemostasis* **1989**, *15*, 269−282.
(38) Vitali, J.; Martin, P. D.; Malkowski, M. G.; Robertson, W. D.; Lazar, J. B.; Winant, R. C.; Johnson, P. H.; Edwards, B. F. The Structure of a Complex of Bovine alpha-Thrombin and Recombinant Hirudin at 2.8-A Resolution. *J. Biol. Chem.* **1992**, *267*, 17670−17678.
(39) Wu, C.-C.; Wang, W.-Y.; Wei, C.-K.; Teng, C.-M. Combined Blockade of Thrombin Anion Binding Exosite-1 and PAR4 Produces Synergistic Antiplatelet Effect in Human Platelets. *Thromb. Haemostasis* **2011**, *105*, 88−95.
(40) Schreyer, A. M.; Blundell, T. USRCAT: Real-Time Ultrafast Shape Recognition with Pharmacophoric Constraints. *J. Cheminf.* **2012**, *4*, 27.
(41) Zoldhelyi, P.; Chesebro, J. H.; Owen, W. G. Hirudin as a Molecular Probe for Thrombin in Vitro and during Systemic Coagulation in the Pig. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 1819−1823.
(42) Xi, G.; Reiser, G.; Keep, R. F. The Role of Thrombin and Thrombin Receptors in Ischemic, Hemorrhagic and Traumatic Brain Injury: Deleterious or Protective? *J. Neurochem.* **2002**, *84*, 3−9.
(43) Maruyama, I.; Shigeta, K.; Miyahara, H.; Nakajima, T.; Shin, H.; Ide, S.; Kitajima, I. Thrombin Activates NF-kappaB through Thrombin Receptor and Results in Proliferation of Vascular Smooth Muscle Cells: Role of Thrombin in Atherosclerosis and Restenosis. *Ann. N.Y. Acad. Sci.* **1997**, *811*, 429−436.
(44) Fareed, J.; Lewis, B. E.; Callas, D. D.; Hoppensteadt, D. A.; Walenga, J. M.; Bick, R. L. Antithrombin Agents: The New Class of Anticoagulant and Antithrombotic Drugs. *Clin. Appl. Thromb./Hemostasis* **1999**, *5*, S45−S55.
(45) Chew, D. P.; Bhatt, D. L.; Sapp, S.; Topol, E. J. Increased Mortality with Oral Platelet Glycoprotein IIb/IIIa Antagonists: A Meta-Analysis of Phase III Multicenter Randomized Trials. *Circulation* **2001**, *103*, 201−206.
(46) Doggrell, S. Clinical Trials with Glycoprotein IIb/IIIa Antagonists - No Benefit without Bleeding? *Drugs Today* **2001**, *37*, 509.
(47) Padmanabhan, K.; Padmanabhan, K. P.; Ferrara, J. D.; Sadler, J. E.; Tulinsky, A. The Structure of alpha-Thrombin Inhibited by a 15-mer Single-Stranded DNA Aptamer. *J. Biol. Chem.* **1993**, *268*, 17651−17654.
(48) Johnson, D. K.; Karanicolas, J. Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface. *PloS Comput. Biol.* **2013**, *9*, e1002951.
(49) Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422−1423.
(50) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(51) Schrodinger LLC, The PyMOL Molecular Graphics System, Version 1.3r1. 2010. http://www.pymol.org/ (accessed Jan 30, 2015).
(52) Duffy, F. J.; Verniere, M.; Devocelle, M.; Bernard, E.; Shields, D. C.; Chubb, A. J. CycloPs: Generating Virtual Libraries of Cyclized and Constrained Peptides Including Nonnatural Amino Acids. *J. Chem. Inf. Model.* **2011**, *51*, 829−836.

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613

(53) Landrum, G. RDKit: Open-Source Cheminformatics. http://www.rdkit.org (accessed Jan 30, 2015).
(54) Ebejer, J. P.; Morris, G. M.; Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **2012**, *52*, 1146−1158.
(55) McCaldon, P.; Argos, P. Oligopeptide Biases in Protein Sequences and Their Use in Predicting Protein Coding Regions in Nucleotide Sequences. *Proteins* **1988**, *4*, 99−122.
(56) Olsen, J. A.; Banner, D. W.; Seiler, P.; Obst Sander, U.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an Enzyme Active Site: Evidence for C-F···C═O Interactions. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 2507−2511.

DOI: 10.1021/ci500431q
J. Chem. Inf. Model. 2015, 55, 600−613