PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors

Elad Donsky* and Haim J. Wolfson*

Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

Associate Editor: Burkhard Rost

ABSTRACT

Motivation: Design of protein–protein interaction (PPI) inhibitors is a key challenge in structural bioinformatics and computer-aided drug design. Peptides, which partially mimic the interface area of one of the interacting proteins, are natural candidates to form protein–peptide complexes competing with the original PPI. The prediction of such complexes is especially challenging due to the high flexibility of peptide conformations.

Results: In this article, we present PepCrawler, a new tool for deriving binding peptides from protein–protein complexes and prediction of peptide–protein complexes, by performing high-resolution docking refinement and estimation of binding affinity. By using a fast path planning approach, PepCrawler rapidly generates large amounts of flexible peptide conformations, allowing backbone and side chain flexibility. A newly introduced binding energy funnel 'steepness score' was applied for the evaluation of the protein–peptide complexes binding affinity. PepCrawler simulations predicted high binding affinity for native protein–peptide complexes benchmark and low affinity for low-energy decoy complexes. In three cases, where wet lab data are available, the PepCrawler predictions were consistent with the data. Comparing to other state of the art flexible peptide–protein structure prediction algorithms, our algorithm is very fast, and takes only minutes to run on a single PC.

Availability: http://bioinfo3d.cs.tau.ac.il/PepCrawler/

Contact: eladdons@tau.ac.il; wolfson@tau.ac.il

1 INTRODUCTION

Protein–protein interactions (PPIs) play a major role in cellular function. Thus, development of small molecules that modulate PPIs has become a major challenge in structural bioinformatics and computer-aided drug design. In particular, it is desired to detect small molecules that will block the PPI, without harming the intrinsic catalytic activity of the target (Arkin and Whitty, 2009; Arkin and Wells, 2004). The search for such inhibitors includes drug-like small molecule inhibitors and short peptide inhibitors. The interaction between two proteins is often characterized by a large interface, lacking a significant single binding pocket (Fletcher and Hamilton, 2006; Lo Conte *et al.*, 1999). Therefore, peptides, which are relatively large inhibitors, can be potential leads in the rational design of inhibitory drugs by mimicking the interacting area (Nieddu

and Pasa, 2007; Naider and Anglister, 2009; Mochly-Rosen and Qvit, 2010; Monfregola *et al.*, 2009; Parthasarathi *et al.*, 2008). Both naturally occurring and synthetic peptides have been used, and these are emerging as promising compounds that efficiently modulate signaling pathways *in vitro* and *in vivo*. Some inhibitory peptides were designed by deriving a short linear segment from one of the proteins in a given PPI complex (Barr *et al.*, 2002; Hashemzadeh *et al.*, 2008; Hayouka *et al.*, 2010; Laudet *et al.*, 2007; Phan *et al.*, 2010). These peptides were successfully able to inhibit interactions with the partner protein. In some of the examples, a mutated peptide was used to achieve better results, and in others, the derived peptide was stapled to form a stable cyclic peptide. Generally, a derived inhibitory peptide is target specific and may bind to the target protein with good binding energy. Using computational measures, it was recently shown that linear derived peptides dominate the interaction energy of many PPIs (London *et al.*, 2010b).

Given an input protein–protein complex, we wish to derive a suitable peptide from one of the proteins, with high affinity to bind to the other, as well as predict the structure of the resulting peptide–protein complex. Simply deriving a rigid peptide with high binding energy from the input complex is not enough (Rubinstein and Niv, 2009). First, peptides are highly flexible molecules containing backbone rotatable bonds and movable side chains. Second, the input structure might not be of high enough resolution for an accurate prediction. Finally, it might be hard to determine whether the resulting peptide–protein complex is stable.

Several flexible peptide–protein docking methods have been introduced. Liu *et al.* (2004) applied a Monte Carlo (MC)-based method for flexible peptide–protein docking, which takes into account also protein side chain flexibility. Their suggested new potential, which considers protein–peptide atom pair interactions and atomic solvation contribution, accurately predicted binding affinity on the test set. Recently, an MC-based algorithm for subangstrom peptide docking refinement that utilizes the Rosetta energy function was introduced by Raveh *et al.* (2010). A molecular dynamics-based refinement method was suggested by Antes (2010). Some high-resolution docking algorithms were designed for specific protein receptors (most of them are also MC based)—e.g. MHC, PDZ and SH3 domains (Bordner and Abagyan, 2006; Chaudhury and Gray, 2009; Fernandez-Ballester *et al.*, 2009; Fagerberg *et al.*, 2006, Niv and Weinstein, 2005; Staneva and Wallin, 2009). These previous methods for flexible peptide docking have relatively long running times, mostly due to time consuming side chain energy minimization, which is extensively applied.

In this article, we present PepCrawler—a new method for flexible peptide docking. Our algorithm requires as input an initial peptide

---

*To whom correspondence should be addressed.

docking hypothesis, which is close to the native structure (in a way that will be quantified). Peptides derived from the interface of the PPI are natural input candidates. This method allows backbone flexibility of the peptide, and side chain flexibility for both the peptide and the receptor protein. We generate large amounts of clash-free peptide docking conformations by using Rapidly-exploring Random Trees (RRT) (LaValle and Kuffner, 2001). RRT is a robotics motivated path planning approach, which has been previously used for computing large-amplitude motions of flexible molecules (Cortés *et al.*, 2005, 2007; Enosh *et al.*, 2006; Raveh *et al.*, 2009). The peptide's flexible backbone conformation and its location are generated by the RRT algorithm. Side chain flexibility of both the peptide and the target protein is modeled by a rotamer library. The initial protein–peptide conformation is refined to a low-energy docking solution. Moreover, since the binding energy itself is often not enough for predicting binding affinity (Kastritis and Bonvin, 2010), a dense binding energy/root-mean-square deviation (RMSD) plot is created and scored. This plot assists in evaluating the affinity of the binding solution via energy funnels (Zhang *et al.*, 1999).

2 METHODS

2.1 PepCrawler algorithm phases

The PepCrawler algorithm (Fig. 1) contains the following phases:

- Phase A—given an input protein–protein complex, derives a single, short, low-energy binding peptide from one of the interacting proteins.
- Phase B—given an input protein–peptide complex, explores the conformational space by using our RRT-based algorithm, creating a large amount of flexible peptide conformations from the given complex; scores these conformations; performs clustering and outputs the top five cluster candidates.
- Phase C—similar to Phase B, performs a more refined RRT exploration focused around the conformation of each cluster candidate from the previous phase and allowing less peptide flexibility. It computes the backbone-atom RMSD (bb-RMSD) between the highest scoring conformation of the input peptide and each of the other conformations. It outputs the highest scoring complex, an Energy/bb-RMSD plot containing all conformations generated in this phase and a corresponding funnel 'steepness' score of this plot.
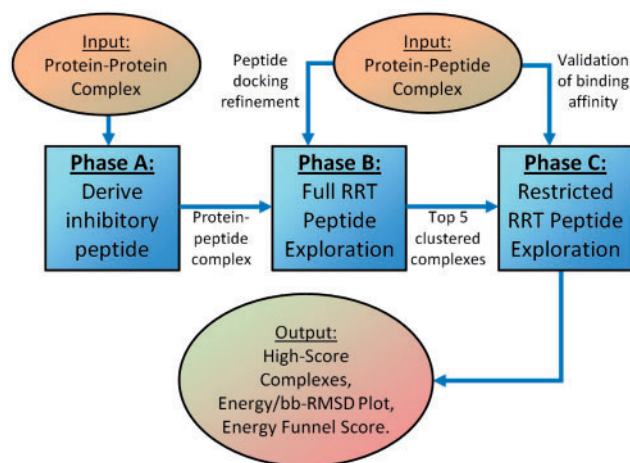

Fig. 1. PepCrawler algorithm phases.

PepCrawler enables three different simulations. First, estimation of peptide binding affinity to a protein by creating and scoring a suitable energy plot, which represents the binding energy funnel (running Phase C). Second, refinement of a given protein–peptide complex (starting from Phase B). Third, derivation of a binding peptide from a protein–protein complex along with refinement and binding affinity estimation of the result (starting from Phase A).

2.2 Phase A—derivation of candidate binding peptides

Given an input protein–protein complex, we wish to derive a small peptide from one of the proteins (P_1) that binds to the other (P_2) with low energy. Assuming that the input complex is given at a sufficient resolution, the algorithm first finds all the amino acids of P_1 that are in the binding interface. Then, it computes the binding energy score (see Section 2.3.6) of all possible derived protein peptide chains of P_1 that satisfy both conditions: the number of amino acids in the chain is between 5 and 12 (inclusive) and there are no three consecutive amino acids that are not in the binding interface with P_2. The lowest binding energy peptide candidate is chosen [at this stage, we do not take into account the free (self) energy of the peptide]. Since peptides use hot spots to bind to proteins (London *et al.*, 2010a), the algorithm tries to trim the candidate peptide, in order to dispose amino acids that are not in the binding site, or contribute non-negative energy (and therefore clearly cannot be hot spots). The peptide is trimmed by removing its tail (from each of the terminal sites), so that the cumulative energy of the peptide tail starting from such amino acid is only slightly negative (it cannot be positive due to the choice of the lowest energy peptide). If such a tail is found, the trimmed peptide is selected as the output-derived peptide. If not, the original peptide is selected.

2.3 Phases B and C—creating and scoring clash-free peptide docking conformations using RRT

Given an initial protein–peptide complex structure, we use the RRT method (Cortés *et al.*, 2005; LaValle and Kuffner, 2001) to rapidly generate a large amount of clash-free peptide conformations in the protein binding site. The main advantages of the RRT method, compared e.g. to other MC methods, is its fast exploration of the feasible conformational space and the fact that a large amount of conformations checked for feasibility are indeed feasible.

2.3.1 Peptide as a moving articulated object   A peptide backbone is highly flexible. This flexibility is caused due to changes in the backbone torsion angles (Φ, Ψ). Therefore, the backbone rotatable bonds form a kinematic backbone chain—where each chain link contains few of the peptide's atoms with no rotatable backbone bond between them. By generating a set of values for the backbone torsion angles, the peptide can be reconstructed, forming a new backbone conformation. Each amino acid contributes two rotatable bonds to the peptide backbone chain (Φ, Ψ).

Apart from its flexibility, we also want to model the peptide global movement in the 3D space. So, given a 3D translation vector and three rotation angles, we can construct a suitable rigid transformation to apply to the peptide.

To conclude, our method takes into account the following degrees of freedom (DOFs): peptide's backbone torsion angles and a rigid transformation. For a peptide with m backbone rotatable bonds, the number of DOFs will be $D = m + 6$. These DOFs are randomly perturbed to create new conformations.

We define suitable ranges for the possible perturbations in the DOFs values. For Phase B, we use a maximum distance of 3 Å for each translation parameter, and a maximum change of 40° in the global transformation rotation angles and in torsion Φ–Ψ angles from the original conformation. For Phase C, we use a maximum distance of 1 Å for each translation parameter, and a maximum change of 20° in the transformation rotation angles and in the torsion angles relative to the original conformation.
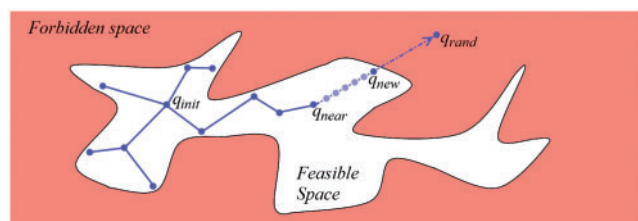
Fig. 2. Expansion of a tree toward a random conformation using an RRT-based algorithm in a 2D DOF space.

2.3.2 The conformational space In robotics motion planning, a basic problem is to produce a continuous motion of an object starting from a given conformation, while avoiding collision with obstacles in the space. The conformational space C is the set of all possible conformations. The set of conformations that avoid collision with obstacles is called the feasible space. The forbidden space is the complement of the feasible space in C. In our method, the number of DOFs G is the dimension of the conformational space, since a suitable G dimensional vector of DOF values (DOF vector) determines the computed 3D backbone conformation. We define the feasible conformation space to be all DOF vectors that generate clash-free conformations. We thus define the forbidden conformational space to be all DOF vectors that generate conformations with steric clashes. RRT allows rapid exploration of the feasible space by expanding a random tree rooted at an initial conformation, using a greedy heuristic (as detailed below).

2.3.3 The RRT-based algorithm Starting from the initial input complex conformation, which is a G-dimensional vector—q_{init}, the algorithm iteratively builds a tree of randomly generated vertices (Cortés *et al.*, 2005). At each iteration, a conformation q_{rand} is sampled from the conformational space, by choosing a random DOF vector (according to the limitations). This sample is used to determine the tree node to be expanded and also the direction of expansion in the conformational space. The nearest node q_{near} in the tree to the sample q_{rand} is found by using a distance metric of DOF vectors in the conformational space. The linear segment $q_{near} - q_{rand}$ is then defined. An attempt is made to expand the tree, progressing by discrete small increments on the selected segment, starting from q_{near}, and checking whether the conformations found are feasible (Fig. 2). This progress stops when a forbidden conformation is reached, or if all conformations that were checked including q_{rand} are feasible. If at least one new feasible conformation was found, the expansion succeeds, resulting in a new node q_{new} being added to the tree as well as the feasible edge $q_{near} - q_{new}$. The feasible conformations that were found on the way are stored for later use.

As a result of this expansion strategy, the probability that a node will be chosen for an expansion is proportional to the volume of its Voronoi region in the conformational space (i.e. the set of points closer to this node than to the other tree nodes). Thus, the strategy has the desired effect of expanding the tree toward unexplored regions of the feasible space. The algorithm output is the final feasible tree, and all the feasible states that were discovered in the tree expansion process.

2.3.4 Collision detection In the tree expansion procedure, the algorithm extensively queries DOF vectors for feasibility. Therefore, a high-performance collision detection mechanism is required. The quadratic cost of enumerating all non-bonded atom pairs will be very slow in our type of complexes. For this purpose, we have implemented a 3D grid, for quick selection of all atoms in the 3D space that are close to a given atom (Halperin and Overmars, 1994).

To determine collision between two non-bonded atoms, we address two different criteria:

(1) Distance criterion—given a reduction factor K, if the centers of the atoms with radii r_1 and r_2 are less than $K(r_1 + r_2)$ apart, they are

defined to be mutually tight. For most complexes we have used $K = 0.8$, if one atom is a peptide atom and the other is a receptor protein atom (intermolecular). For the other possibilities (intramolecular), we have used $K = 0.7$ in order to increase side chain motion space within the molecule.

(2) Repulsive VdW (rVdW) criterion—the complex energy (computed at a later stage) is strongly affected by rVdW forces; thus, we do not want to allow high rVdW forces to exist between atoms in feasible complexes. We define a threshold parameter F—the maximum rVdW force allowed between two atoms. In our simulations, we have used $F = 4$ for Phase B and $F = 1.5$ for Phase C (we thus allow more atom penetration in Phase B than in Phase C). The computation time of the rVdW value is ~ 3 times slower than the computation time of the distance criterion.

In order to achieve fast collision detection, we first compute the distance criterion. If the atoms are not mutually tight—they are considered clash-free. Only if they are mutually tight, the rVdW criterion is computed in order to determine whether there is a collision. Decrement of the K-values or increment of the F-values will enlarge the feasible space, which will make the RRT more probable to explore small feasible cavities and narrow passages, but will also create large amount of feasible conformations with relatively high energy.

2.3.5 Side chain flexibility The algorithm takes into account side chain movements of the peptide as well as of the receptor protein. It moves the side chains in order to try to solve clashes before deciding that a DOF vector is in the forbidden space. The side chain movements are modeled by various side chains rotamers, taken from the Dunbrack backbone-dependent rotamer library (Dunbrack and Karplus, 1994). A change in a side chain rotamer affects the 3D location of only a few atoms—as opposed to a change in the RRT DOFs values that might affect the location of the whole peptide. Therefore, the algorithm performs side chain movement only to solve a current clash with the other atoms.

Our clash detection algorithm (CDA) is therefore also responsible to move the relevant side chains. In detail, the input to the CDA is a DOF vector, to be checked for clashes. The algorithm iteratively inserts the atoms of the complex to a 3D grid by their location according to the DOFs, and previously stored rotamer conformations. If a pair of atoms clashes, the algorithm checks whether at least one of them is a moving rotamer atom (MRA), i.e. a side chain atom that changes its location for different rotamers. If so, the algorithm stores these two atoms and continues. If not, a clash in the complex is declared, and CDA is done. After inserting the entire complex, the algorithm randomly chooses a clashing side chain and tries to resolve the clash by replacing its rotamer. The rotamers are selected randomly according to their probability in the rotamer library. For side chains that clash with a non-MRA, all relevant rotamers are checked, until a feasible one is found. For two MRAs in a clash, belonging to different side chains, all possible rotamer pairs are checked. If all clashes were solved, the output is a set of side chain rotamers that makes a feasible conformation.

2.3.6 Energy computation In the process of the RRT construction, we do not compute the energy of the feasible complexes. This method thus has the ability to circumvent energy traps better than other MC algorithms. The RRT exploration results in a tree containing feasible nodes and a list of feasible states found while progressing on line segments. We then compute the energy of all feasible nodes and of every third state on each of the expanded line segment edges (for example, if an edge contains 12 feasible states, 4 will be scored). We use our previously developed binding energy function score of the FireDock method, which was originally implemented for protein–protein binding (Andrusier *et al.*, 2007). This energy score includes a variety of energy terms: desolvation energy (ACE potential), VdW interactions (attractive, repulsive), partial electrostatics, hydrogen and disulfide bonds, π–π interactions, cation–π interactions and aliphatic interactions. The exact formula and further explanations appear

in the FireDock paper (formula no. 17). Since the input peptide undergoes large conformational changes during the simulation, we also compute its self energy—which includes rVdW forces between the peptide's atoms, the probability of side chains rotamers to appear according to the rotamer library and penalties for very low Ramachandran potential torsion angles (Lovell *et al.*, 2003). We are not applying any side chain energy minimization method throughout the algorithm. The final complex energy of a conformation is the sum of the binding energy and the peptide's self energy. Hydrogen atoms were added to the initial conformation for energy computations purposes using REDUCE (Word *et al.*, 1999).

We then compute the bb-RMSD between the lowest energy peptide conformation and all the other feasible-generated conformations, and draw a corresponding energy versus bb-RMSD 2D plot: the protein–peptide complex energy values on the *y*-axis and bb-RMSD values on the *x*-axis.

2.3.7 Clustering The first cluster is created by selecting all conformations that are at most 1.8 Å bb-RMSD from the lowest energy conformation. Then, the lowest energy conformation among all the unclustered conformations is found. All unclustered conformations that are at most 1.8 Å bb-RMSD from this current lowest energy conformation are then assigned to the second cluster. This process is repeated three more times to form five different clusters.

2.3.8 Scoring energy funnels Inspection of protein–protein docking solutions has indicated that near-native results are expected to have not only low-energy values but also visible steep energy funnels (Zhang *et al.*, 1999). Thus, it is plausible that steep energy funnels for protein–peptide docking in the bb-RMSD 2D plot may also indicate good binding affinity (see Section 3.1). We therefore compute a rough funnel score, indicating the steepness of a funnel. In order to compute this score, we first compute the lower envelope of the energy plot (since no energy minimization procedure is performed on PepCrawler conformations, we thus relate only to the lowest energy conformations when identifying the funnel). This computation is done by dividing the bb-RMSD axis into short intervals (0.05 Å wide) and choosing the minimal energy conformation in each one. Then, we select a set of sampling points found on this lower envelope: 30 points in the interval 0–1 Å, 20 in the interval 1–2 Å and 25 in the interval 2–4.5 Å (points are selected by partitioning each bb-RMSD interval into *n* equal segments, where *n* is the number of points). Then, we compute the linear regression line of this set of sampling points, passing through the initial solution conformation (zero bb-RMSD). The output funnel score is defined to be the slope of this regression line. By choosing more sampling points near the origin and less sampling points far from the origin, we favor funnels that are steep in the area near the low-energy conformation (origin) and becoming more and more moderate on funnels that are moderate near the low-energy conformation and becoming steeper. Thus, a high funnel score will be achieved, for example, for peptides that are found inside a binding pocket—creating a relatively steep funnel near the origin.

3 RESULTS

3.1 Energy funnels of native protein–peptide complexes and wrong protein–peptide docking solutions

In order to check the validity of our conjecture that steep energy funnels in the bb-RMSD 2D plot indicate high binding affinity, we have performed Phase C of the PepCrawler algorithm on two benchmark databases—one representing native protein–peptide complexes and the other decoy complexes.

For native inputs, we have used all PDBs from PeptiDB—a clustered database of protein–peptide complexes (London *et al.*, 2010a). We have then run PepCrawler Phase C, creating the energy plot starting from the lowest energy conformation (Fig. 3).

Table 1. Showing PepCrawler simulations on the native structure of protein–peptide complex 1QKZ (PDB id) and on top 10 solutions (nine decoys and one near-native solution) created using PatchDock and FiberDock given 1QKZ receptor protein and peptide as input (See Section 3.1)

Protein–peptide complex input	PepCrawler-refined complex energy	PepCrawler funnel score
1QKZ—native	−32.69	7.17
Solution-1	−34.88	3.17
Solution-2	−36.42	4.81
Solution-3 (near native)	−31.24	6.93
Solution-4	−23.37	2.77
Solution-5	−16.99	1.77
Solution-6	−41.05	2.98
Solution-7	−18.4	2.67
Solution-8	−11.78	3.3
Solution-9	−12.62	1.5
Solution-10	−19.31	1.87

The docking solutions are numbered according to their final FiberDock score (where Solution-1 is the lowest energy complex). Although Solutions 1, 2 and 6 have better energy than the native complex, their funnel score is significantly lower.

The average funnel score of all 103 input complexes was 8.41, and only 13 complexes had a funnel score < 5.

In order to create decoy protein–peptide docking solutions, we have arbitrarily selected 15 PeptiDB complexes, and used our previously developed rigid-docking PatchDock algorithm (Duhovny *et al.*, 2002) followed by the flexible (both backbone and side chains) docking refinement FiberDock (Mashiach *et al.*, 2010) algorithm, to create new energy-optimized docking solutions (10 solutions for each input). Out of the 150 solutions, we have chosen, in total, 60 low-energy complexes, where the peptide conformation is at least 20 Å apart from the native, forming a decoy docking solution. We have then run Phase C on these 'decoy' docking solutions. The average funnel score of all 60 input complexes was 4.08, and only 11 complexes had a funnel score > 5. Although the refined decoy solutions had low complex energy values, the funnel score of most of them was not steep (Table 1).

Our experiments on the PeptiDB and the decoy benchmarks have shown clear correlation between the funnel steepness score and the binding affinity.

3.2 Peptide refinement

We have tested the PepCrawler refinement mechanism on 25 arbitrarily selected PeptiDB structures. For each structure, we have randomly created few clash-free random peptide conformations, with peptide bb-RMSD of ∼3.5 and 5 Å from the original conformation. We have assigned a random rigid transformation, chosen random Φ–Ψ perturbations (up to 35° shift for each torsion angle), and chosen random rotamer conformations for the side chains of the peptide as well as for the side chains in the binding site of the protein. We have then run Phases B and C of the algorithm on the input conformations. In 88% of the 3.5 Å input structures, one of the top three PepCrawler refinement solutions had < 1.6 Å bb-RMSD from the native conformation. For the 5 Å models, this success rate was 70%.

We present refinement and funnel score results for three specific domain receptor groups. In each group, six arbitrarily selected
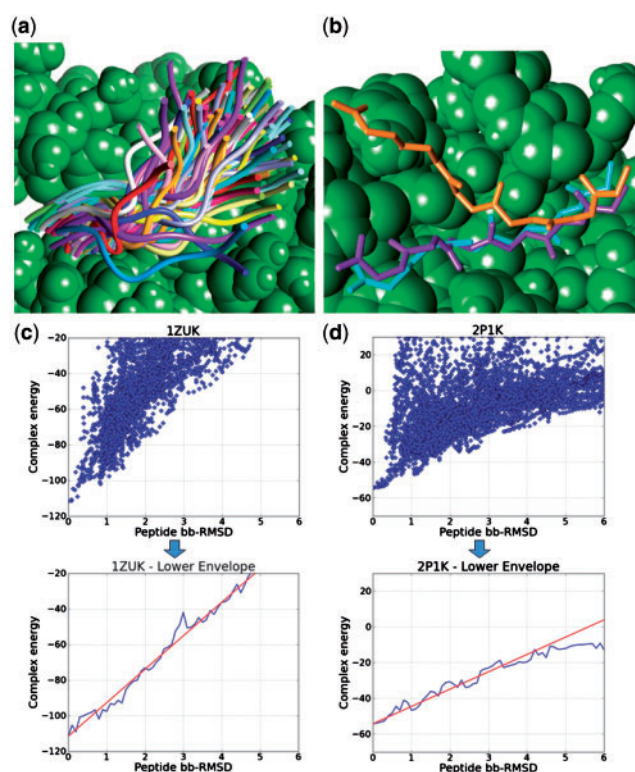
Fig. 3. (**a**) 100 ribbon backbones of feasible peptides generated by PepCrawler (input PDB id: 1JD5). These were randomly chosen out of 17 000 feasible peptides generated in a single Phase C run. (**b**) PepCrawler peptide refinement (PDB id: 1AWU). The initial randomly generated input conformation (orange) forms 5 Å bb-RMSD with the native peptide (purple). Top refinement solution (cyan) is 1.1 Å bb-RMSD away from the native peptide. (**c** and **d**) Energy/bb-RMSD plots and their corresponding lower envelopes and funnel regression lines of input protein–peptide complexes 1ZUK and 2P1K. Plots show bb-RMSD from the minimal energy conformation. Funnel scores (= regression line slopes) are 15.91 (1ZUK) and 8.26 (2P1K).

models from the PDB were used to create six random models with 5 Å peptide bb-RMSD from native.

3.2.1 Group 1—MHC Class I proteins MHC Class I (MHC-I) proteins contain two domains, which create a large groove for peptides to bind. These proteins bind peptides that are 8–11 amino-acid in length and that tightly fit into the groove, exhibiting shape and physicochemical complementarity. Very good results were achieved with MHC-I protein–peptide complexes. In all simulations, the top-scoring result had < 1.2 Å bb-RMSD from the native conformation. The average energy of these top-scoring results was −87.9 and the average funnel 'steepness' score was 11.3. The relatively high funnel score was achieved mainly due to the fact that even minor conformational changes of peptides that are found inside a groove, might cause some clashes with the receptor protein, resulting in a relatively high binding energy.

3.2.2 Group 2—SH3 domain proteins SH3 is a small protein domain of ∼60 amino acids. SH3 domains mediate signal transduction by recognizing peptides that are 8–10 amino acid in

length, where all the amino acids of the peptide are found in the binding interface. In all simulations of the models in this group, one of the top two refinement solutions was < 1.6 Å bb-RMSD from the native peptide conformation. The average energy of the near-native solutions was −58.29 and the average funnel score was 7.26.

3.2.3 Group 3—PDZ domain proteins PDZ domains bind to specific C-terminal sequence motifs that are usually about five amino acids in length. Library screens suggested that two of the peptide residues are most critical for recognition (Harris and Lim, 2001). Each of these two amino acids points into a different pocket of the PDZ domain. In this group's simulations, one of the top four refinement solutions was < 1.8 Å bb-RMSD from the native peptide conformation. The average energy of the near-native solutions was −38.29 and the average funnel score was 7.18. Since PDZ domain binding peptides are relatively short in length and since only two amino acids are critical for binding, we could expect the average energy of this group to be not as low as the other two. However, the average funnel of the PDZ complexes is as steep as in the other two cases.

3.3 Derived binding peptides—experimentation on *in vivo* and *in vitro* examined data

We have examined three cases where *in vivo* and *in vitro* data were published. We have evaluated the ability of PepCrawler to derive binding peptides from the structures of protein–protein complexes for two of these cases. We have run the full PepCrawler algorithm on each of the two complexes, expecting that the derived PepCrawler peptide will be similar to the experimentally detected one, and that the existence of steep energy funnels will be in correlation with the binding affinity shown *in vivo*. In the third case, where three different derived peptides were examined *in vitro*, we checked the ability of PepCrawler to identify the best inhibitor.

3.3.1 Protein kinase CK2 subunit interaction Protein kinase CK2 has a quaternary structure consisting of two catalytic subunits (CK2-α) and two regulatory subunits (CK2-β). Laudet *et al.* (2007) derived an eight-residue peptide from CK2-β chain (residues 186–193), referred to as the Pc peptide. This peptide was chosen by structural identification of CK2-β hot spots, analyzing functional CK2-β mutants, and performing *in vivo* validation of hot spots. A cyclic Pc peptide was created by cyclizing three additional glycine residues via two cysteine residues. The biological activity of Pc was examined, showing that the same amount of CK2-α activity could be found associated with biotinylated CK2-β or biotinylated Pc, indicating that Pc stably interacts with CK2-α. Both cyclic and non-cyclic Pc peptides were shown to be effective. Another binding essay was performed, showing that Pc could antagonize the CK2 subunit interaction.

We have tested the full PepCrawler algorithm (all three phases) on the input CK2 complex (PDB id: 1JWH). The algorithm Phase A first derived the lowest energy peptide (186–197) with binding energy of −19.3. Then, it trimmed that peptide resulting in an 8mer (186–193) with the binding energy of −17.54—(remarkably, this is actually the non-cyclic Pc peptide). The complex energy of this peptide conformation was only −4.5 due to high inner peptide rVdW forces. Running Phases B and C on the conformation, resulted in a best energy peptide with −37.39 complex energy, 12.51
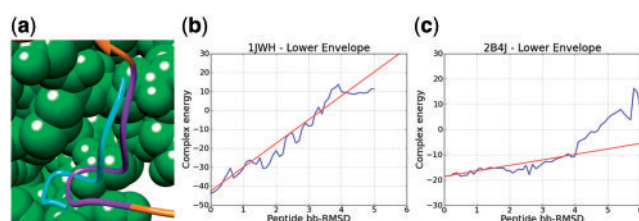
Fig. 4. (**a**) Protein kinase CK2 subunit interaction (PDB id: 1JWH). CK2-α chain (spacefill green), CK2-β chain (orange), the derived Pc peptide (residues 186–193, purple) and top Pc refined solution (cyan). (**b**) Energy plot (lower envelope) of top Pc refined solution, with high funnel score of 12.51. (**c**) Energy plot (lower envelope) of top-refined conformation of LEDGEF-derived peptide (365–369), with low funnel score of 2.16.

funnel score and 2.3 Å bb-RMSD from the starting derived peptide conformation (from the PDB). This top-scoring peptide also had the lowest bb-RMSD to the original peptide conformation, among all other cluster candidates. The relatively high funnel score indicates its high binding affinity (Fig. 4).

3.3.2 HIV-1 integrase and LEDGF protein complex The human transcriptional coactivator p75 (which is also referred to as LEDGF—lens epithelium-derived growth factor), is the dominant cellular binding partner of HIV-1 IN (Cherepanov *et al.*, 2005). Hayouka *et al.* (2010) designed a short LEDGF-derived peptide containing residues 365–369, which were identified as the main residues that participate in IN binding. Using a combination of alanine scanning, fluorescence anisotropy, NMR and IN enzymatic assay, they showed that this peptide is a weak IN inhibitor. Although the residues 365–369 are important for IN binding at the protein level, they are not sufficient for inhibition at the peptide level.

We run the full PepCrawler algorithm on the IN-LEDGF complex (PDB id: 2B4J) input. Phase A of the algorithm derived the peptides 365–369 from LEDGF (similar to the one chosen manually) with binding energy of −6.1. Phases B and C generated five output complexes where the best energy solution had an energy of −18.59, a relatively low funnel score of 2.16, and bb-RMSD of 2.7 Å from the starting derived peptide conformation (from the PDB). This top-score solution also had the best funnel score out of the other 4. Although few solutions with good complex energies were achieved, the lack of significant energy funnels indicates low binding affinity of this peptide (Fig. 4).

3.3.3 p53C and 53BP2 complex Friedler *et al.* (2002) tested the feasibility of stabilizing p53C protein by trying to find small peptides that bind its native state. They examined the p53C-53BP2 complex structure (PDB id: 1YCS). 53BP2 contains three loops which make the contacts with p53C. They tested three derived peptides corresponding to these three loops: residues 422–428 (CDB1), residues 469–477 (CDB2) and residues 490–498 (CDB3). *In vitro* tests showed no significant binding of CDB1 to p53C. CDB2 and CDB3 did bind, where CDB3 had much tighter binding. CDB3 was chosen as the lead peptide for further experiments in p53 research.

We run PepCrawler on the three different CDB-p53C complexes. Results are shown on Table 2.

Table 2. PepCrawler simulation results on three input-derived peptides (CDB1-CDB3) binding to p53C

Peptide	Initial input complex energy	PepCrawler-refined complex energy	PepCrawler funnel score
CDB1	−14.48	−16.1	2.96
CDB2	−3.25	−21.06	5.34
CDB3	−15.21	−58.46	9.53

PepCrawler simulations predict that CDB3 has significantly tighter binding to p53C than the other two peptides, with better complex energy score and funnel score, in full agreement with the experimental results.

3.4 Running time

We have measured PepCrawler performance on a PC with a single core 2.4 GHz Intel processor, using a single-threaded version of our algorithm—where all the algorithm operations are done sequentially. For all input complexes checked, Phase A took <5 s to run. Phase C runs on average for 1:20 min for each input, performing 600 RRT exploration rounds with an average of ∼10 000 conformations checked for feasibility and ∼2500 energy computations of some of the feasible conformations. Phase B runs on average for 14:15 min, performing 7000 RRT rounds with average of ∼143 000 conformations checked for feasibility and ∼44 600 energy computations. The running times presented here can be significantly reduced by parallelizing clash detections and energy computations of complexes, or by running the program on a cluster.

4 SUMMARY

We have presented a generic method for high-resolution peptide docking in cases where an approximately binding peptide candidate is provided. By efficiently filtering conformations, performing fast energy computations and avoiding the need of running side chain minimization procedure on the result conformations, we drastically reduce running time compared with other MC type protein–peptide docking methods. The algorithm allows distinguishing between near-native docking solutions and decoy conformations of peptide–protein complexes. We have also presented an application of the algorithm for the highly challenging and useful task of detection of peptide inhibitors to PPIs. The latter application was demonstrated to be consistent with the published experimental data.

REFERENCES

Andrusier,N. *et al*. (2007) FireDock: fast interaction refinement in molecular docking. *Proteins*, **69**, 139–159.

Antes,I. (2010) DynaDock: a new molecular dynamics-based algorithm for protein–peptide docking including receptor flexibility. *Proteins*, **78**, 1084–1104.

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

Arkin,M.R. and Whitty,A. (2009) The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr. Opin. Chem. Biol.*, **13**, 284–290.

Barr,R.K. *et al*. (2002) Identification of the critical features of a small peptide inhibitor of JNK activity. *J. Biol. Chem.*, **277**, 10987–10997.

Bordner,A.J. and Abagyan,R. (2006) Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins*, **63**, 512–526.

Chaudhury,S. and Gray,J.J. (2009) Identification of structural mechanisms of HIV-1 protease specificity using computational peptide docking: implications for drug resistance. *Structure*, **17**, 1636–1648.

Cherepanov,P. *et al*. (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl Acad. Sci. USA*, **102**, 17308–17313.

Cortés,J. *et al*. (2005) A path planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics*, **21**, 116–125.

Cortés,J. *et al*. (2007) Molecular disassembly with RRT-like algorithms. In *Proceedings of the International Conference Robotics and Automation*, Computer Society Press, Washington DC, pp. 3301–3306.

Duhovny,D. *et al*. (2002) Efficient unbound docking of rigid molecules. In *Workshop on Algorithms in Bioinformatics*. Vol. 2452 of *LNCS*, Springer, Berlin, pp. 185–200.

Dunbrack,R.L. and Karplus,M. (1994) Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.*, **1**, 334–340.

Enosh,A. *et al*. (2006) Prediction and simulation of motion in pairs of transmembrane α-helices. *Bioinformatics*, **23**, 212–218.

Fagerberg,T. *et al*. (2006) Structural prediction of peptides bound to MHC class I. *J. Mol. Biol.*, **356**, 521–546.

Fernandez-Ballester,G. *et al*. (2009) Structure-based prediction of the Saccharomyces cerevisiae SH3-ligand interactions. *J. Mol. Biol.*, **388**, 902–916.

Fletcher,S. and Hamilton,A.D. (2006) Targeting protein-protein interactions by rational design: mimicry of protein surfaces. *J. R. Soc. Interface*, **3**, 215–233.

Friedler,A. *et al*. (2002) A peptide that binds and stabilizes p53 core domain: chaperone strategy for rescue of oncogenic mutants. *Proc. Natl Acad. Sci. USA*, **99**, 937–942.

Halperin,D. and Overmars,M.H. (1994) Spheres, molecules, and hidden surface removal. In *Symposium on Computational Geometry*, ACM, New York, pp. 113–122.

Harris,B.Z. and Lim,W.A. (2001) Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.*, **114**, 3219–3231.

Hashemzadeh,M. *et al*. (2008) Chemical structures and mode of action of intravenous glycoprotein IIb/IIIa receptor blockers: a review. *Exp. Clin. Cardiol.*, **13**, 192–197.

Hayouka,Z. *et al*. (2010) Mechanism of action of the HIV-1 integrase inhibitory peptide LEDGF 361–370. *Biochem. Biophys. Res. Commun.*, **394**, 260–265.

Kastritis,P.L. and Bonvin,A.M.J.J. (2010) Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res.*, **9**, 2216–2225.

Laudet,B. *et al*. (2007) Structure-based design of small peptide inhibitors of protein kinase CK2 subunit interaction. *Biochem. J.*, **408**, 363–373.

LaValle,S.M. and Kuffner,J.J. (2001) Rapidly-exploring random trees: progress and prospects. In *Algorithmic and Computational Robotics: New Directions*. Peters, Dartmouth, pp. 293–308.

Liu,Z. *et al*. (2004) Structural mining: self-consistent design on flexible protein-peptide docking and transferable binding affinity potential. *J. Am. Chem. Soc.*, **126**, 8515–8528.

Lo Conte,L. *et al*. (1999) The atomic structure of proteinprotein recognition sites. *J. Mol. Biol.*, **285**, 2177–2198.

London,N. *et al*. (2010a) The structural basis of peptide-protein binding strategies. *Structure*, **18**, 188–199.

London,N. *et al*. (2010b) Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? *Proteins*, **78**, 3140–3149.

Lovell,S.C. *et al*. (2003) Structure validation by Cα geometry: Φ, Ψ and Cβ deviation. *Proteins*, **50**, 437–450.

Mashiach,E. *et al*. (2010) FiberDock: flexible induced-fit backbone refinement in molecular docking. *Proteins*, **78**, 1503–1519.

Mochly-Rosen,D. and Qvit,N. (2010) Peptide inhibitors of protein-protein interactions: from rational design to the clinic. *Chim. oggi*, **28**, 14–16.

Monfregola,L. *et al*. (2009) A SPR strategy for high-throughput ligand screenings based on synthetic peptides mimicking a selected subdomain of the target protein: a proof of concept on HER2 receptor. *Bioorg. Med. Chem.*, **17**, 7015–7020.

Naider,F. and Anglister,J. (2009) Peptides in the treatment of AIDS. *Curr. Opin. Struct. Biol.*, **19**, 473–482.

Nieddu,E. and Pasa,S. (2007) Interfering with protein-protein contact: molecular interaction maps and peptide modulators. *Curr. Top. Med. Chem.*, **7**, 21–32.

Niv,M.Y. and Weinstein,H. (2005) A flexible docking procedure for the exploration of peptide binding selectivity to known structures and homology models of PDZ domains. *J. Am. Chem. Soc.*, **127**, 14072–14079.

Parthasarathi,L. *et al*. (2008) Approved drug mimics of short peptide ligands from protein interaction motifs. *J. Chem. Inf. Model*, **48**, 1943–1948.

Pettersen,E.F. *et al*. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *Comput. Chem.*, **25**, 1605–1612.

Phan,J. *et al*. (2010) Structure-based design of high affinity peptides inhibiting the interaction of p53 with MDM2 and MDMX. *J. Biol. Chem.*, **285**, 2174–2183

Raveh,B. *et al*. (2009) Rapid sampling of molecular motions with prior information constraints. *Plos. Comp. Biol.*, **5**, 2.

Raveh,B. *et al*. (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins*, **78**, 2029–2040.

Rubinstein,M. and Niv,M.Y. (2009) Peptidic modulators of protein-protein interactions: progress and challenges in computational design. *Biopolymers*, **91**, 505–513.

Staneva,I. and Wallin,S. (2009) All-atom Monte Carlo approach to protein-peptide binding. *J. Mol. Biol.*, **393**, 1118–1128.

Word,J.M. *et al*. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of sidechain amide orientation. *Mol. Biol.*, **285**, 1735–1747.

Zhang,C. *et al*. (1999) Protein-protein recognition: exploring the energy funnels near the binding sites. *Proteins*, **34**, 255–267.