

# Knowledge-Guided Docking of WW Domain Proteins and Flexible Ligands <sup>\*</sup>

Haiyun Lu<sup>1</sup>, Hao Li<sup>1</sup>, Shamima Banu Bte Sm Rashid<sup>1</sup>, Wee Kheng Leow<sup>1</sup>, and Yih-Cherng Liou<sup>2</sup>

<sup>1</sup> Dept. of Computer Science, School of Computing, National University of Singapore, Singapore 117590

{luhaiyun, lihao, shamima, leowwk}@comp.nus.edu.sg

<sup>2</sup> Dept. of Biological Sciences, Faculty of Science, National University of Singapore, Singapore 117543

dbslyc@nus.edu.sg

**Abstract.** Studies of interactions between protein domains and ligands are important in many aspects such as cellular signaling. We present a knowledge-guided approach for docking protein domains and flexible ligands. The approach is applied to the WW domain, a small protein module mediating signaling complexes which have been implicated in diseases such as muscular dystrophy and Liddle’s syndrome. The first stage of the approach employs a substring search for two binding grooves of WW domains and possible binding motifs of peptide ligands based on known features. The second stage aligns the ligand’s peptide backbone to the two binding grooves using a quasi-Newton constrained optimization algorithm. The backbone-aligned ligands produced serve as good starting points to the third stage which uses any flexible docking algorithm to perform the docking. The experimental results demonstrate that the backbone alignment method in the second stage performs better than conventional rigid superposition given two binding constraints. It is also shown that using the backbone-aligned ligands as initial configurations improves the flexible docking in the third stage. The presented approach can also be applied to other protein domains that involve binding of flexible ligand to two or more binding sites.

## 1 Introduction

Protein domains are the fundamental units of tertiary structure of many proteins. One of the most important functions of protein domains is to bind specific ligands to assemble intracellular signaling networks to perform distinct biological functions. The number of defined protein domains has expanded considerably in recent years.

Studies of interactions between protein domains and their ligands are crucial for deeper insight of the binding affinities involved. With this vital understanding target prediction of novel domain-binding ligands would be possible, allowing

---

<sup>\*</sup> This research is supported by NUS R-252-000-293-112.

for subsequent cloning and expression. Determination of possible target ligands by laboratory experimental techniques alone is a known bottleneck requiring intensive consumption of time and resources. Therefore computational techniques are needed to effectively simulate domain bindings.

Many protein docking algorithms have been developed to solve the problem. Two broad classifications [10] of docking algorithms are rigid docking and flexible docking. Rigid docking algorithms solve a simpler version of the protein docking problem termed *bound docking* by reconstruction of a protein complex from the bound structures of the two proteins that constitute the complex [5, 8, 14, 19, 23, 27, 28]. Docking is framed as a rigid alignment problem of two rigid objects with complementary shapes. Flexible docking algorithms solve the general protein docking problem termed *unbound* or *predictive docking* by prediction of binding of two proteins in their free or unbound states [7, 9, 12, 16, 18, 20, 22, 26]. This problem regards one or both proteins as flexible objects to account for significant conformational shape changes which occur during protein interactions. A flexible molecule often presents a very large number of degrees of freedom posing great difficulty for the flexible docking problem.

Flexible docking should be used to predict possible binding and potential novel targets for protein domains as the correct conformations of novel targets are usually unknown. Generally, this remains a very difficult and challenging task. Fortunately, known binding site characteristics of protein domains can be used to help solve the difficult docking problem.

The knowledge of binding sites of protein domains is very useful for predicting possible ligand bindings. General flexible docking algorithms often make use of the binding site information. However, what information to use and how to use it for effective and accurate flexible docking is a challenge. For example, AutoDock [20] requires the user to specify a bounding box around the protein binding site in which an optimal ligand conformation is searched for. The amount of binding site information used in AutoDock is minimal and thus not very effectively used.

This paper presents an approach for docking protein domains and flexible ligands using known binding site information as the constraints. Our approach uses known binding site knowledge to first search for the protein domains and the ligand residues recognized by the domains. Then the ligand's peptide backbone is aligned to the domains based on the binding constraints. Finally, existing algorithms can be used to perform flexible docking, with the backbone-aligned ligands as the initial configuration. To be specific, we apply the approach to WW domains as an application example in this paper.

## 2 WW Domains

WW domains are present in signaling proteins found in all living things. They have been implicated in signal mediation of human diseases such as muscular dystrophy, Alzheimer's disease, Huntington's disease, hypertension (Liddle's syndrome) and cancer [3, 11, 24, 25]. WW domains are distinguished by the characteristic presence of two signature tryptophan residues (W) that are spaced 20–

**Table 1.** Residue sequences of sample WW domains and ligands.

Group	PDB	WW domain sequence	Ligand sequence
I	1EG4	HFLSTSVQGPWERAISPKNKVPYYINHETQTTTCWDHPKMTELYQ	KNMTPYRSPPPYPVP
II/III	2DYF	GSWTEHKSPDGRTYYYNTETKQSTWEKPDD	GSTAPPLPR
IV	1PIN	KLPPGWEKRMSRSGRVVYFNHITNASQWERPSGNSSSG	

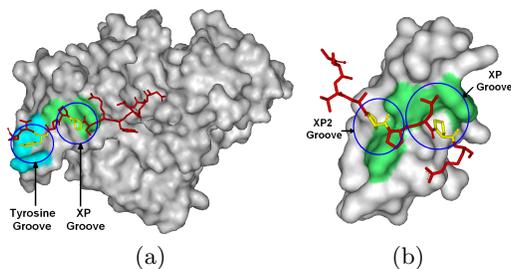
22 amino acids apart (Table 1). They are known to recognize proline-containing peptide ligands and they share similarities with other proline recognition domains such as SH3 and EVH1 domains [17, 29].

WW domains are classified into four groups [11] based on ligand specificity. Group I binds to ligands containing Proline-Proline-‘Any amino acid’-Tyrosine (PPxY) motif (Table 1). Group II binds to ligands containing Proline-Proline-Leucine-Proline (PPLP) motif (Table 1). Group III recognizes Proline-rich segments interspersed with Arginine (R) residues. Group IV binds to short amino acid sequences containing phosphorylated Serine or Threonine followed by proline. Recent studies show that Group II and III WW domains have very similar or almost indistinguishable ligand preferences, suggesting that they should be classified into a single group [15]. Our study focuses on the first three groups of WW domains, as fewer Group IV samples are available in RCSB Protein Data Bank (PDB) [2]. Examples of WW domains and their corresponding ligand amino acid sequences are presented in Table 1.

Group I and II/III WW domains have two binding grooves that recognize ligands [24]. Group I WW domains contain the Tyrosine groove, Group II/III WW domains contain the XP2 groove and both groups contain the XP groove. A Tyrosine groove is formed by three residues, Ile/Leu/Val, His, and Lys/Arg/Gln, and it recognizes Tyrosine (Y) residue of the ligand. An XP groove is formed by Tyr/Phe and Trp residues whereas an XP2 groove is formed by Tyr and Tyr/Trp residues. Both recognize Xaa-Pro (P), including Pro-Pro, segments of the ligand. It is to be noted that the grooves are formed by non consecutive residues in the amino acid sequence because the WW domain protein folds in 3-D to give rise to the grooves (Fig. 1).

### 3 Related Work

Flexible docking algorithms can be classified into three categories. *Rigid docking with refinement* methods perform rigid docking of the proteins followed by refinement of their side chains [7, 9, 12, 16, 26]. By applying side chain refinement, side chain flexibility can be accounted for to improve docking results. The method of [9] performs optimization of both backbone displacement and side chain conformations based on simulated annealing Monte Carlo. The methods of [7, 26] apply biased probability Monte Carlo minimization of the ligand-interacting side chains while [16] uses energy minimization. The algorithm in [12] uses side chain rotamers and rigid body minimization to relax the interfaces of docking results. These methods handle side chain flexibility but not backbone conformational



**Fig. 1.** WW domain proteins and ligands. (a) 1EG4: Group I WW domain complexed with  $\beta$ -dystroglycan peptide. (b) 2DYF: Group II WW domain complexed with a peptide ligand with PPLP motif. Gray: proteins containing WW domains, blue: Tyrosine groove, green: XP/XP2 grooves, red: ligands, yellow: groove-binding residues of ligands.

changes. To handle backbone flexibility, HADDOCK [6] performs rigid-body docking followed by refinement of both the backbone and the side chains using simulated annealing based on molecular dynamics (MD). Biochemical and biophysical interaction data such as chemical shift perturbation data resulting from NMR titration experiments are used so it is not a general docking algorithm.

Incremental construction algorithms place ligand fragments one at a time at the binding sites of the binding protein [18, 22]. They require the knowledge of binding sites to place ligand fragments at the sites. Their computation speed while satisfactory for smaller ligands remains unsuitable for large ligands.

Energy minimization methods apply optimization algorithms to search for the ligand conformation with minimum binding energy [13, 20]. The objective is to determine the ligand conformation with minimal binding energy. Various optimization algorithms can be applied such as simulated annealing, Monte Carlo and genetic algorithms. In particular AutoDock [20] uses a hybrid Lamarckian genetic algorithm to optimize an empirical energy function that includes van der Waals potential, hydrogen bonding, Coulombic electrostatics and desolvation. The computational cost of such an energy function is very high. So the number of degrees of freedom is often limited to reduce the search space in practice.

## 4 Knowledge-Guided Protein Docking

The 3 stages in our approach for docking WW domains and flexible ligands are binding groove and motif search, backbone alignment and flexible docking.

### 4.1 Binding Groove and Binding Motif Search

Given a WW domain protein with known group classification (Group I or Group II/III). The two types of binding grooves presented in the WW domain are also known. Each binding groove is formed by residues in a special pattern (Table 2). Residues which form the binding grooves are determined by applying a substring search on the WW domain's amino acid sequence.

**Table 2.** Residue patterns of WW domain binding grooves.

Binding groove	Pattern	Example
Tyrosine groove	I _ H _ _ K	1EG4: ...WERAISP <span style="background-color: #e0e0e0;">NKVPYYIN</span> HETQ <span style="background-color: #e0e0e0;">TTCW</span> ...
	L            R	
	V            Q	
XP groove	Y _ _ _ _ _ _ _ _ W F	1EG4: ...WERAISP <span style="background-color: #e0e0e0;">NKVPYYIN</span> HETQ <span style="background-color: #e0e0e0;">TTCW</span> ...
XP2 groove	Y _ Y W	2DYF: ...WTEHKSPDGRT <span style="background-color: #e0e0e0;">YYY</span> NTETKQSTW...

From the binding grooves, the corresponding ligand motifs (PPxY or PPLP) are also known. Ligand residues forming possible motifs recognized by the binding grooves on WW domains are determined by a substring search applied on the ligand’s amino acid sequence.

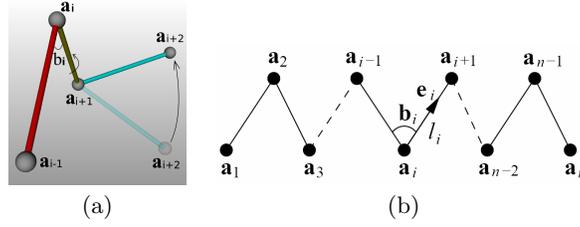
## 4.2 Backbone Alignment

Our backbone alignment method performs flexible alignment of a ligand’s backbone to binding sites given the two binding grooves of a WW domain protein and possible binding motifs of the ligand.

A ligand’s residue sequence is divided into 3 segments according to the binding motifs. As an example the ligand sequence of 1EG4 complex (KNMTPYR-SPPPYVPP) is divided into KNMTPYRS, PPPY and VPP. The middle segment, an instance of the PPxY motif, is flanked by two residues recognized by two binding grooves of the WW domain. The backbone alignment method aligns the backbone of the middle segment to the WW domain such that the flanking residues fit the grooves. The knowledge of relative positions and orientations of the two flanking residues with respect to the (grooves of) WW domain serve as *binding constraints*. The other two segments and ligand side chain atoms are added after backbone alignment.

The bond angle and bond length between two neighboring atoms are assumed to be fixed, but the torsion angle of a rotatable bond can change to give rise to various conformations of a protein molecule (Fig. 2(a)). Similar assumptions are made in existing flexible docking algorithms.

Let  $\mathbf{a}_i$ ,  $i \in \{1, 2, \dots, n\}$ , denote the positions of the  $n$  atoms in the middle segment of the ligand backbone (Figure 2(b)). The middle segment has  $n/3$  residues because for each residue 3 backbone atoms N, C $\alpha$ , and C are considered. The two binding constraints specify the atom positions  $a_0, a_1, a_2$  and  $a_{n-2}, a_{n-1}, a_n$ , which represent the two flanking residues. We denote the target positions of these constrained atoms as  $\{a_0^*, a_1^*, a_2^*\}$  and  $\{a_{n-2}^*, a_{n-1}^*, a_n^*\}$ . To satisfy the constraint on the first flanking residue, rigid transformation is applied on the backbone to align  $\{a_0, a_1, a_2\}$  to  $\{a_0^*, a_1^*, a_2^*\}$ . To satisfy the constraint on the



**Fig. 2.** Backbone model. (a) The bond angle  $\mathbf{b}_i$  and bond length  $l_i$  between atoms  $\mathbf{a}_i$  and  $\mathbf{a}_{i+1}$  are fixed. However torsion of the bond (indicated by arrow) can rotate atom  $\mathbf{a}_{i+2}$  to a new position. (b) Model of atoms in the backbone.

second flanking residue, we introduce the cost function

$$C_s = \frac{1}{2} \sum_{j=0}^2 \|\mathbf{a}_{n-j} - \mathbf{a}_{n-j}^*\|_2 \quad (1)$$

Minimizing  $C_s$  minimizes the distances between the atoms in the last residue and their target positions.

Since the backbone can twist but not bend or stretch, constraints on the bond angles and bond lengths should be incorporated in order to correctly deform the backbone to satisfy the constraints on the residues. To represent the stretching and bending constraints we introduce the bond direction  $\mathbf{e}_i$  and bond length  $l_i$  as illustrated in Figure 2(b). Since the first three atoms are fixed by constraints after their rigid transformation we have, for  $i \geq 3$ ,

$$\mathbf{a}_{i+1} = \mathbf{a}_i + l_i \mathbf{e}_i \quad (2)$$

Thus, given  $\{a_0, a_1, a_2\}$ ,  $l_i$  and  $\mathbf{e}_i$  determine the positions of the other atoms. So,  $C_s$  can also be expressed in terms of  $l_i$  and  $\mathbf{e}_i$ . Since a bond cannot stretch  $l_i$  is fixed to ensure this condition. Also,  $\mathbf{e}_i \cdot \mathbf{e}_{i+1}$  corresponds to the bond angle  $b_i$ . So the bending constraint is encoded by the cost function

$$C_b = \frac{1}{2} \sum_{i=1}^{n-2} \|\mathbf{e}_i \cdot \mathbf{e}_{i+1} - \mathbf{e}_i^0 \cdot \mathbf{e}_{i+1}^0\|_2 \quad (3)$$

where  $\mathbf{e}_i^0 \cdot \mathbf{e}_{i+1}^0$  corresponds to the initial value of bond angle  $b_i$ . Minimizing  $C_b$  minimizes the change of the bond angles while keeping the bond lengths fixed.

The peptide bond nearly always has *trans* configuration since it is more energetically favorable than *cis*. The backbone omega torsion angles are limited to values of  $180^\circ \pm 5^\circ$ , except for proline residues. There is no limitation on the omega torsion angle for proline because it can be in either *trans* or *cis* configuration. Additionally, WW domains often bind to proline-rich ligands and the average distribution of phi, psi torsion angles for polyproline stretches (4 or more consecutive prolines) is  $(-75^\circ, +145^\circ) \pm 10^\circ$  [1]. Let  $t_i$  denote the torsion angles

formed by atoms  $\mathbf{a}_{i-1}$ ,  $\mathbf{a}_i$ ,  $\mathbf{a}_{i+1}$ ,  $\mathbf{a}_{i+2}$ . The torsional constraint is represented by the cost function

$$C_t = \sum_{t_i \text{ is limited}} \|t_i - t_i^0\|_2 \quad (4)$$

where  $t_i = -\text{atan2}(\|\mathbf{e}_i\| \mathbf{e}_{i-1} \cdot (\mathbf{e}_i \times \mathbf{e}_{i+1}), (\mathbf{e}_{i-1} \times \mathbf{e}_i) \cdot (\mathbf{e}_i \times \mathbf{e}_{i+1}))$  and  $t_i^0$  denote the preferred value of  $t_i$ . Minimizing  $C_t$  minimizes the difference between torsion angles and their preferred values.

The total cost function for backbone alignment is then

$$C = k_b C_b + k_s C_s + k_t C_t \quad (5)$$

where  $k_b$ ,  $k_s$  and  $k_t$  are weighting factors. In Eq. 5, the independent variables are the  $\mathbf{e}_i$ 's. Varying  $\mathbf{e}_i$  changes the torsion angles but not the bond angles because of  $C_b$ . Quasi-Newton algorithm [21] is applied to compute the optimal  $\mathbf{e}_i$  that minimizes  $C$  to yield the aligned configuration of the ligand's backbone.

### 4.3 Flexible Docking

In the second stage knowledge of WW domain binding specificity is effectively used to obtain backbone-aligned ligands. In the third stage, such ligands are used as starting configurations for flexible docking. Any existing flexible docking algorithm can be employed in our approach.

## 5 Experiments

Known WW domain binding targets are used in experiments to test the performance of our approach. 14 WW domain proteins complexed with proline-containing ligands were collected from RCSB Protein Data Bank (PDB) [2]. Complexes 1EG4, 1K9R, 1K5R, 1JMQ, 1I5H, 2JO9, 2DJY form WW domain Group I test cases and 2HO2, 2OEL, 2DYF, 1YWI, 2JUP, 2RLY, 2RM0 form Group II/III test cases. The WW domain proteins were separated from their ligands. Molecular Dynamics (MD) simulations were run using the AMBER program [4] to simulate possible unbound ligand conformations.

Firstly, the backbone alignment algorithm was tested against rigid superposition given binding site knowledge. Backbone alignment was performed 30 times for each test run. The results are ranked according to the cost computed by Eq. 5 and only the top ranked alignments are recorded.

Rigid superposition was performed for each test case based on known binding placements of the two ligand residues which bind to two binding grooves of WW domain. Least squares fit is used to compute the rigid transformation of ligand to minimize the distance between the two binding residues and their ideal positions in binding sites.

Results evaluation was performed by comparing the backbone atoms N, C $\alpha$  and C between two binding constraints of the ligand with those in the bound structure and computing the root mean square deviation (RMSD)(Table 3). The

**Table 3.** RMSD( $\text{\AA}$ ) of backbone alignment results and rigid superposition results.

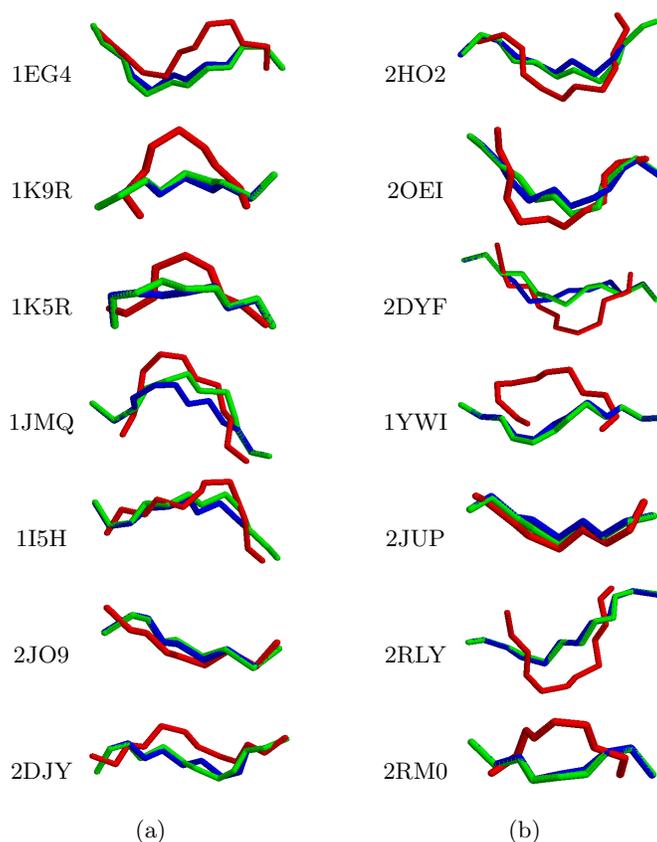
Test Case	Backbone Alignment	Rigid Superposition
1EG4	0.30	1.38
1K9R	0.13	1.77
1K5R	0.28	1.50
1JMQ	0.94	1.71
1I5H	0.42	1.23
2JO9	0.13	0.59
2DJY	0.30	1.02
2HO2	0.27	1.39
2OEI	0.34	1.13
2DYF	0.48	1.99
1YWI	0.19	3.41
2JUP	0.18	0.54
2RLY	0.17	2.19
2RM0	0.12	1.67

average RMSD of results produced by backbone alignment method is  $0.30\text{\AA}$  with a standard deviation of  $0.21\text{\AA}$  and the average RMSD results produced by rigid superposition is  $1.54\text{\AA}$  with a standard deviation of  $0.72\text{\AA}$ . It is obvious that our backbone alignment method produced better results than rigid superposition.

Figure 3 visualizes the results of backbone alignment compared with rigid superposition for 14 test cases. Ligand backbone conformations between two binding constraints resulting from our method are very close to bound structures with RMSD smaller than  $0.5\text{\AA}$  in all cases except 1JMQ. However the placements produced by rigid superposition are far from optimal, producing good results for only two test cases 2JO9 and 2JUP where input ligand shapes are similar to the bound structures. When input ligand shapes differ greatly from native complexes such as in test cases 1K9R, 1YWI, 2RLY and 2RM0 rigid superposition fails to create good ligand placements satisfying the two binding constraints.

To further test our approach’s performance one of the most widely used flexible docking programs, AutoDock was employed at the third stage of our scheme in three experimental settings. In the first setting top-ranked ligand backbone alignments generated by the second stage were used as initial configurations for AutoDock. In the second setting ligand placements resulting from rigid superposition in the previous experiment were used as initial configurations. In the third setting, AutoDock was run using random initial ligand placements.

The configuration files for AutoDock were prepared using AutoDockTools. The grid parameter files specifying the 3D search space were manually defined to surround the two WW domain binding sites. The WW domain protein (receptor) was held rigid and the ligand flexible. As AutoDock limits the number of torsional degrees of freedom to 32 AutoDockTools were used to select up to 32 bonds whose torsional rotations moved the most number of atoms. Several parameters were adjusted for all settings in the same way. For example maximum number of energy evaluations was set to 25,000,000 and maximum number



**Fig. 3.** The backbone alignment results (green) are closer than rigid superposition results (red) to the bound structure (blue). (a) WW domain Group I test cases. (b) Group II/III test cases.

of generations to 27,000. 50 AutoDock runs were performed for each test case in each experimental setting and solutions were ranked according to their scores evaluated by AutoDock.

Usually a solution with RMSD smaller than  $2\text{\AA}$  is classified as successful docking and is considered a very good result. A solution with RMSD smaller than  $3\text{\AA}$  is classified as partially successful. The docking is considered a success if the top-scored solution is successful for each test case. We relaxed the criteria in our evaluation, because most of our test cases possess a large number of degrees of freedom which make the docking problem extremely difficult. We considered the docking successful if one of the top 10 score-ranked solutions has RMSD smaller than  $3\text{\AA}$ .

Table 4 lists RMSD results for all three experimental settings. Only 3 test cases (1K9R, 2OEI and 1YWI) are successful in the second and third settings.

**Table 4.** RMSD analysis of AutoDock results in three experimental settings. The minimum, maximum and average RMSD (Å) of the top 10 ranked solutions are listed.

Test Cases	Torsion DoF	Backbone Alignment			Rigid Superposition			Random Placement		
		min	max	avg	min	max	avg	min	max	avg
1EG4	41	4.05	8.73	6.79	4.36	11.56	7.22	4.62	9.66	6.52
1K9R	12	2.80	5.07	3.71	2.86	5.93	4.71	2.85	6.38	4.65
1K5R	24	3.33	6.20	4.43	3.73	7.60	5.69	3.81	7.30	5.29
1JMQ	24	3.46	6.08	4.84	3.80	5.71	4.78	3.14	8.00	4.91
1I5H	50	4.68	6.71	5.62	4.76	8.92	7.42	4.96	10.89	7.73
2JO9	28	3.52	6.86	5.08	4.30	7.62	5.92	4.16	10.53	7.06
2DJY	65	4.51	7.87	6.47	5.41	8.24	7.26	5.35	9.79	7.20
2HO2	13	3.73	6.45	5.15	3.03	5.75	4.57	3.74	7.70	5.38
2OEI	12	2.93	5.00	4.16	2.18	4.35	3.27	2.93	5.48	4.16
2DYF	25	2.68	6.72	4.41	3.74	7.85	5.43	3.14	5.32	4.04
1YWI	9	1.90	4.94	3.86	2.48	4.34	3.15	2.82	4.26	3.38
2JUP	16	2.59	5.93	4.39	4.10	6.06	5.36	3.41	6.87	4.88
2RLY	14	2.57	5.86	4.31	3.06	4.13	3.58	3.18	5.25	4.15
2RMO	15	3.43	5.49	4.51	3.19	5.55	4.31	3.06	5.53	4.09

They all have short ligands comprising of at most 9 amino acids and 12 torsional degrees of freedom. The difficulty of docking small flexible ligands is relatively low and AutoDock is able to produce good docking results. However for more difficult test cases AutoDock failed to give successful docking solutions using initial ligand structures produced by rigid superposition or using random initial ligand placement. Unsurprisingly AutoDock results in the second and third settings are fairly close to each other. The reason is that rigid superposition is unable to produce optimal ligand conformations that satisfy the two binding constraints and thus initial ligand structures used by AutoDock in the second setting are no better than the random ligand conformations used in the third setting.

In the first setting 6 test cases (1K9R, 2OEI, 2DYF, 1YWI, 2JUP and 2RLY) are successful. Besides the three simple test cases AutoDock in our approach succeeded in three more cases with larger numbers of torsional degrees of freedom. In particular 2DYF has 25 torsional degrees of freedom which is difficult for flexible docking. Among the 8 failed test cases the results in the first setting are still better than those for the other two settings. In 5 out of the 8 failed cases the average RMSD of top 10 ranked solutions are better than the second setting and in 6 out of 8 cases better than the third setting. Clearly using our backbone alignment method to create initial ligand structures improves the overall performance of AutoDock.

## 6 Conclusions

This paper presents a three-stage approach for docking of WW domains and flexible ligands. The first stage searches for possible binding motifs of ligands

using a substring search. The second stage aligns the ligand's peptide backbone to binding grooves in WW domains using a quasi-Newton constrained optimization algorithm. The cost function used in the optimization represents multiple constraints on the alignment including positional constraints of ligand residues at the binding grooves, bond angle constraints of backbone atoms and torsion constraints of selected phi, psi as well as omega torsion angles of the backbone atoms. Knowledge of WW domain binding grooves and ligand residues bound to the grooves is used to set the cost function. As shown from the experimental results, the backbone alignment method in stage two works better than conventional rigid superposition. The backbone-aligned ligands produced in this stage serve as good starting structures to the third stage which uses any flexible docking algorithm to perform docking. In the experiments AutoDock in our approach yields better results than using rigid superposition to create initial structures or using random initial ligands. The presented approach can also be applied to other protein domains that involve binding of flexible ligands to two or more binding sites. The optimal placement of ligands near binding sites produced by our backbone alignment stage can be used as good initial structures for subsequent stages.

## References

1. A. A. Adzhubei and M. J. E. Sternberg. Left-handed polyproline II helices commonly occur in globular proteins. *Journal of Molecular Biology*, 229:472–493, 1993.
2. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.
3. P. Bork and M. Sudol. The WW domain: a protein module that binds proline-rich or proline-containing ligands, 2000.
4. D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Jr Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26:1668–1688, 2005.
5. R. Chen, L. Li, and Z. Weng. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, 52:80–87, 2003.
6. C. Dominguez, R. Boelens, and A. M. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125(7):1731–1737, 2003.
7. J. Fernández-Recio, M. Totrov, and R. Abagyan. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins*, 52:113–117, 2003.
8. H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg. Modelling protein docking using shape complementarity, electrostatics, and biochemical information. *Journal of Molecular Biology*, 272:106–120, 1997.
9. J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology*, 331:281–299, 2003.
10. I. Halperin, B. Ma, H. Wolfson, and R. Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47:409–443, 2002.

11. J. L. Ilesya, M. Sudolb, and S. J. Winder. The WW domain: Linking cell signalling to the membrane cytoskeleton. *Cellular Signalling*, 14:183–189, 2002.
12. R. M. Jackson, H. A. Gabb, and M. J. Sternberg. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *Journal of Molecular Biology*, 276:265–285, 1998.
13. G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267:727–748, 1997.
14. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. Friesem, C. Aflalo, and I. Vakser. Molecular surface recognition: Determination of geometric fit between protein and their ligands by correlation techniques. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 89, pages 2195–2199, 1992.
15. Y. Kato, K. Nagata, M. Takahashi, L. Lian, J. J. Herrero, M. Sudol, and M. Tanokura. Common mechanism of ligand recognition by group II/III WW domains. *Journal of Biological Chemistry*, 279(30):31833–31841, 2004.
16. L. Li, R. Chen, and Z. Weng. RDOCK: refinement of rigid-body protein docking predictions. *Proteins*, 53:693–707, 2003.
17. M. J. Macias, S. Wiesner, and M. Sudol. Ww and sh3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters*, 53(1):30–37, 2002.
18. S. Makino and I. D. Kuntz. Automated flexible ligand docking method and its application for database search. *Journal of Computational Chemistry*, 18:1812–1825, 1997.
19. J. G. Mandell, V. A. Roberts, M. E. Pique, V. Kotlovyi, J. C. Mitchell, E. Nelson, I. Tsigelny, and L. F. Ten Eyck. Protein docking using continuum electrostatics and geometric fit. *Protein Engineering*, 14:105–113, 2001.
20. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *Journal of Computational Chemistry*, 19:1639–1662, 1998.
21. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press, 2002.
22. M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261:470–489, 1996.
23. D. Ritchie and G. Kemp. Protein docking using spherical polar Fourier correlations. *Proteins*, 39(2):178–194, 2000.
24. M. Sudol. Structure and function of the WW domain. *Progress in Biophysics and Molecular Biology*, 65(1-2):113–132, 1996.
25. M. Sudol. From src homology domains to other signaling modules: proposal of the ‘protein recognition code’. *Oncogene*, 17:1469–1474, 1998.
26. M. Totrov and R. Abagyan. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*, S1:215–220, 1997.
27. A. Tovchigrechko and I. A. Vakser. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research*, 34:W310–W314, 2006.
28. I. A. Vakser. Protein docking for low-resolution structures. *Protein Engineering*, 8:371–377, 1995.
29. A. Zarrinpar, R. P. Bhattacharyya, and W. A. Lim. The structure and function of proline recognition domains. *Science’s STKE*, 179:re8, 2003.