Journal of Bioinformatics and Computational Biology
Vol. 10, No. 3 (2012) 1242002 (15 pages)

DOI: 10.1142/S0219720012420024

Imperial College Press
www.icpress.co.uk

AN EVOLUTIONARY CONSERVATION-BASED METHOD FOR REFINING AND RERANKING PROTEIN COMPLEX STRUCTURES

BAHAR AKBAL-DELIBAS[*,‡], IRINA HASHMI[†,§], AMARDA SHEHU[†,¶] and NURIT HASPEL[*,‖]

[*]*Computer Science Department*
University of Massachusetts Boston
100 Morrissey Boulevard
Boston, MA 02125, USA

[†]*Department of Computer Science*
George Mason University, 4400 University Drive
Fairfax, VA 22030, USA
[‡]*abakbal@cs.umb.edu*
[§]*ihashmi@gmu.edu*
[¶]*amarda@gmu.edu*
[‖]*nurit.haspel@umb.edu*

Received 13 February 2012
Accepted 24 March 2012
Published 4 June 2012

Detection of protein complexes and their structures is crucial for understanding their role in the basic biology of organisms. Computational docking methods can provide researchers with a good starting point for the analysis of protein complexes. However, these methods are often not accurate and their results need to be further refined to improve interface packing. In this paper, we introduce a refinement method that incorporates evolutionary information into a novel scoring function by employing Evolutionary Trace (ET)-based scores. Our method also takes Van der Waals interactions into account to avoid atomic clashes in refined structures. We tested our method on docked candidates of eight protein complexes and the results suggest that the proposed scoring function helps bias the search toward complexes with native interactions. We show a strong correlation between evolutionary-conserved residues and correct interface packing. Our refinement method is able to produce structures with better lRMSD (least RMSD) with respect to the known complexes and lower energies than initial docked structures. It also helps to filter out false-positive complexes generated by docking methods, by detecting little or no conserved residues on false interfaces. We believe this method is a step toward better ranking and prediction of protein complexes.

Keywords: Docking refinement; interface conservation; evolutionary trace.

[‖]Corresponding author.

1. Introduction

Protein complexes play a central role in cellular organization and function, ion transport and regulation, signal transduction, protein degradation, and transcriptional regulation.[1] Since the three-dimensional structure and the functionality of proteins are closely related to each other, detection of protein complexes and their structures is crucial for understanding the role of protein complexes in the basic biology of organisms.

Predicting the structure of a complex formed by assembling multiple chains is a difficult problem to solve in wet labs. Computational methods can become very useful where experimental methods fall short and provide researchers a good starting point for the analysis of protein complexes. Computational docking methods use structural and geometric search techniques and physico-chemical filters to model complex binding and rank computed structures according to energetic criteria using scoring functions. These scoring functions typically focus on electrostatic, Van der Waals (VdW), and hydrostatic interactions, similarity to experimental structures; or agreement with other experimental data.[2–6] The results generated by such computational methods are expected to be low-energy structures that are similar to the native complex structures.

Unfortunately, computational docking methods are not complete: low-energy structures often disagree with NMR data.[7] Recent CAPRI (Critical Assessment of PRedicted Interactions) rounds show an important observation[2]: even the most accurate methods predict only about 50% of the targets. Moreover, computational docking methods often produce, in addition to near-native but slightly incorrect complexes, a large number of low-energy false positive complexes, where the two proteins are bound on the wrong interface. Therefore, the results of computational docking methods need to be further refined in order to obtain native-like structures, and their ability to detect the correct binding interface needs to be improved. Usage of refinement methods on protein complexes is not limited to computational docking methods; structures obtained by experimental methods can also be refined.

In this paper we present a novel docking refinement method based on evolutionary information to better discriminate native-like from decoy structures for the protein−protein docking results and improve interface packing. The main idea is that proteins tend to preserve their functionally important amino acids, which play a part in interacting with its partner proteins, throughout the evolution.[8,9] Additionally, functionally important amino acids of different interacting chains are expected to be close to each other on the interface. Recent methods using sequence conservation through evolutionary traces (ET)[9,11,12] allow detecting binding interfaces *in silico*. Our method makes use of this information by employing a novel scoring function based on evolutionary conservation, in addition to the conventional VdW energy term, and drives the search toward conformations which have those functionally important amino acids positioned close to each other on each chain. The scoring function iteratively detects top-scoring transformations at each stage of

the refinement process to improve interface packing. It employs a greedy approach to avoid exponential growth of the putative complexes and speed up the search. The method explained below can be readily extended to run on complexes containing any number of chains and not just dimers.

We employed our refinement method to eight docked complexes with known native structures and compared them to the input complexes by means of lRMSD to the native complex and their potential energy. We show that our new scoring scheme exhibits a strong correlation between the conservation score and the native interfaces, and results in refined docked complexes with lower lRMSD with respect to the native structure, as well as lower potential energy, than the input complexes. In other words, we show that our method can guide the search toward more native-like complexes. Furthermore, we show the ability of our method to detect false positives among the docked complexes. Docking methods often result in a large number of putative complexes, many of which dock the two input proteins at the completely wrong interface. When testing our method on such false positive complexes, we detected very few conserved residues on the bound interface, as opposed to near-native structures, which contained a large number of conserved residues. This shows that our method can potentially be used to distinguish falsely docked complexes from near-native complexes and improve the performance of ranking functions.

2. Methods

2.1. *Overview*

The input to our program is a protein complex structure generated by a docking method. The refinement proceeds in cycles where each cycle seeks to improve the conformation of one unit with respect to the other one. A unit corresponds to a single chain in this paper, however, it can also correspond to a combination of chains if this method is extended to run on complexes containing any number of chains and not just dimers.[14] The improvements are done via rigid-body rotations. The search to improve these conformations focuses on their vicinity in order to keep the computational costs low and avoid large changes to the structure. Units are rotated by a random angle within a predefined range around an arbitrary axis passing through the centroid of the unit. Each rotation results in a new conformation and these randomly generated conformations are ranked using a novel ET-based interface scoring function. This process is repeated multiple times to obtain better refinement results. However, only k top-scoring conformations are further refined in the next iteration to avoid the exponential growth of the search space.

After a set of new top conformations is obtained for the selected pair of units, the results are energy minimized for 200 steps to resolve local clashes using NAMD[13] at the end of each cycle. We used a very small number of minimization steps to locally relax the structure without causing large changes to the overall conformation. The output of the program is the top-k conformations generated at the last cycle, which are all refined versions of the input structure. It should be noted that while in this

work we present the results of our method on dimers, it can be easily extended to complexes of more than two chains using an iterative refinement process. This is the subject of ongoing and future work in our group.

2.2. *Creating random conformations with uniform distribution*

In our previous work,[14] we followed a systematic approach when creating conformations. For each input conformation, rotations were performed around the X, Y, and Z coordinate axes from -5° to 5°, resulting in 30 new conformations. In this work we take a probabilistic approach that expands our search space. For each input conformation, 100 random conformations are generated by rotations around an arbitrary axis[15] passing through the centroid of the unit. Both the rotation angle and the rotation axis are selected randomly from a uniformly distributed set.

Since our method assumes a roughly correctly docked conformation as the input and seeks to refine it through small rotations without drastically altering the overall shape, we consider -5° to 5° a reasonable range for rotations.

We select the arbitrary rotation axis from the set of all three-dimensional unit vectors in a unit sphere whose center is at the centroid of the chain. A three-dimensional vector V can be represented by two angles: the angle between V and X-axis (α) and the angle between V and Z-axis (β). Then the x, y, z components of V can be expressed as following:

$$\begin{aligned} V_x &= \cos\alpha \\ V_y &= \sin\alpha \\ V_z &= \cos\beta \end{aligned} \tag{1}$$

The arbitrary rotation axis is then selected among 360×360 three-dimensional unit vectors by randomly selecting α and β values from integers between 1 to 360. Hence, we not only take a probabilistic approach as opposed to a systematic approach but also consider a significantly wider search space for the rotation compared to our previous work.[14]

2.3. *Scoring function*

The scoring function that the search seeks to optimize is computed for the set of interface atoms, which is defined for each chain as the atoms within at most 6 Å distance to the adjacent chain atoms. In our previous work,[13] the scoring function used a term that consists of effective distance restraints[3] and a molecular surface complementarity function[12] based on evolutionary conservation of residues, as well as the usual VdW and electrostatic terms taken from the AMBER ff03 force field.[16] In this paper, we introduce a novel scoring function term for calculating the interface conservation based on ET scores of each interface residue, which we describe below. In addition, we analyze the correlation between lRMSD values and each of VdW, electrostatic, effective distance restraint, and conservation terms.

2.3.1. *Conservation term of the scoring function*

Our quest for this term was based on the finding that "a protein family should conserve its functional sites and have a distinctly lower mutation rate at these sites".[9] In other words, the functionally important surfaces of proteins should consist of highly conserved residues. Through experiments with native structures of different proteins we examined whether atoms of conserved residues create clusters around interfaces. We indeed observed such clusters on interfaces (see Sec. 3). It is worth noting that such clusters were also visible on other parts of these proteins. We believe this may be due to protein allostery and the fact that many proteins have multiple interaction sites.

We obtain the ET rank files for each protein from the Evolutionary Trace Server,[10] which are produced by a sequence analysis on homologous proteins.[9] For each interface atom, we define the evolutionary conservation value as in Eq. (2), where *residueRank* is the ET rank value of the residue that the atom belongs to, μ is the mean of ET rank values of residues in the chain, and σ is the standard deviation of ET rank values of residues in the chain. This is a simplified version of the function we used previously.[12,13] We found that this simplified version creates significantly better correlation with lRMSD values.

$$c_i = (\mu - residueRank)/\sigma \tag{2}$$

For lower ET rank values, which represent lower mutation rates, conservation values will be higher. Similarly, ET rank values larger than the mean will have negative conservation values. Atoms with positive conservation values are considered *conserved*.

The conservation term of our interface scoring function is then defined as in Eq. (3), where f, the conservation value for the interface atom pair i and j, is defined as in Eq. (4). In this manner, each interface atom i on one unit and interface atom j on the other unit are considered in computing the conservation term.

$$E_{\text{conservation}} = \sum_{i,j} f(i,j) \tag{3}$$

$$f(i,j) = \begin{cases} -c_i * c_j & \text{if } c_i < 0 \quad \text{and} \quad c_j < 0 \\ c_i * c_j & \text{otherwise} \end{cases} \tag{4}$$

It is important to make sure that $E_{\text{conservation}}$ is not biased toward conformations with larger interfaces. For example, consider two different hypothetical conformations: one with an interface of 1000 atoms where 300 of them are conserved, and the other with an interface of 300 atoms where 200 of them are conserved. In such cases, the former should not be preferred over the latter by simply considering the number of conserved atoms on the interface. Therefore, *non-conserved* atoms (i.e. atoms with negative conservation values) are set to have negative impact on the calculation.

2.3.2. *Scoring function terms and lRMSD correlation analysis*

In order to improve and validate our scoring function, we performed a detailed correlation analysis between the lRMSD values and each of VdW, electrostatic, effective distance restraint, and conservation terms. For each protein, we created 1000 random conformations from the same initial docked structure as described in Sec. 2.2. Each conformation's lRMSD value with respect to the known native structure is computed along with values of VdW, electrostatic, effective distance restraint, and conservation terms.

Table 1 shows the correlation coefficients for proteins 1C1Y-AB, 1CSE-EI, 1DS6-AB, 1FLT-VY, 1IKN-AC, 1TX4-AB, and 1WWW-WY, where the first four characters represent the PDB ID for each structure (e.g. 1C1Y) and the last two characters after the dash represent the docked chains (e.g. chain A and chain B). Also, Fig. 1 displays the plots of correlation between lRMSD and minimized VdW, electrostatic, effective distance restraint, and conservation terms for protein 1C1Y. Due to limited space, the correlation plots of other proteins are provided in the Supplemental Materials. As seen in Table 1 and Fig. 1(c), the proposed conservation term has a strong negative correlation with lRMSD for all proteins except 1IKN-AC. In other words, higher conservation values result in lower lRMSD results. On the other hand, electrostatic and effective distance restraint terms calculated for interface atoms did not show significant correlation with lRMSD. In particular, we observed almost a random walk relationship between interface electrostatic energy and lRMSD in all proteins.

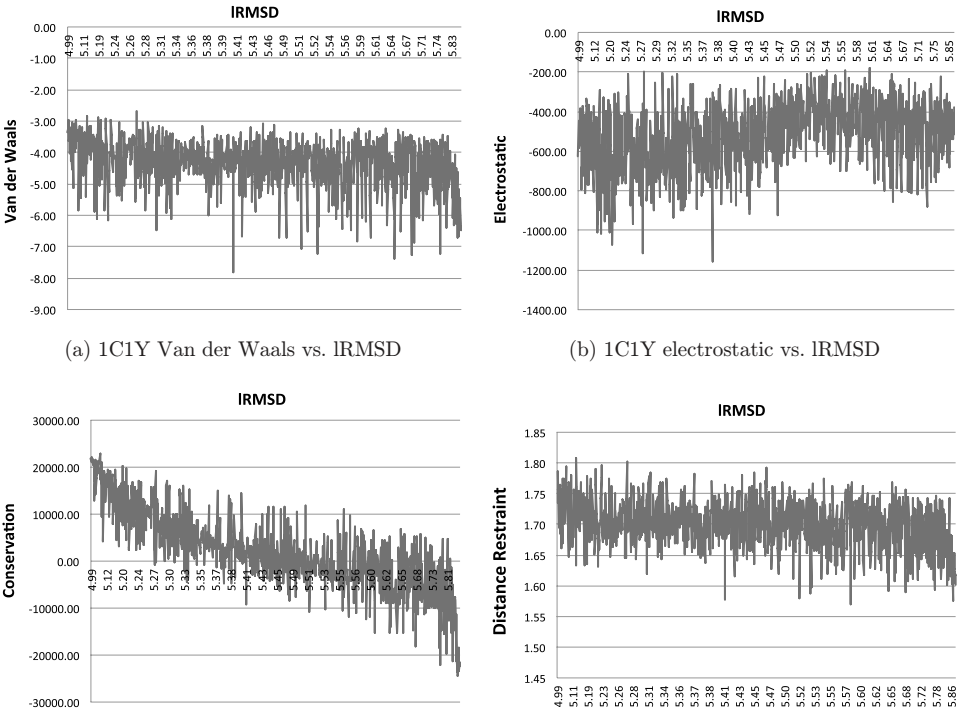
2.3.3. *Putting it all together*

As discussed above, the proposed conservation term shows strong negative correlation with lRMSD values. Therefore, we build our scoring function based on this term to discriminate native-like structures from decoy ones. In addition to the conservation term, we use minimized VdW term to eliminate structures with clashing atoms. We calculate "soft" Van der Waals term, E_{VdW}, as in Eq. (5), which

Table 1. Correlation coefficients between lRMSD and different energy terms.

Protein	VdW	Electrostatic	Distance Restraint	Conservation
1C1Y-AB	−0.32	0.30	−0.30	−0.84
1CSE-EI	−0.56	−0.11	−0.79	−0.57
1DS6-AB	−0.19	−0.35	0.04	−0.81
1FLT-VY	−0.56	−0.13	−0.55	−0.52
1IKN-AC	−0.56	−0.45	−0.60	0.25
1TX4-AB	−0.78	−0.03	−0.57	−0.56
1WWW-WY	−0.76	0.24	−0.74	−0.70

Note: The conservation term shows a strong negative correlation with lRMSD. Electrostatic and effective distance restraint terms calculated for interface atoms do not show significant correlation with lRMSD.



(a) 1C1Y Van der Waals vs. lRMSD

(b) 1C1Y electrostatic vs. lRMSD

(c) 1C1Y conservation vs. lRMSD

(d) 1C1Y distance restraint vs. lRMSD

Fig. 1. Correlations between lRMSD values and Van der Waals, electrostatic, effective distance restraint and conservation terms for 1C1Y.

is explained in our previous work.[13]

$$E_{VdW} = \sum_{\text{atom pairs}} \varepsilon \left[\left(\frac{r_{ij}}{d_{ij}} \right)^9 - \left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] \tag{5}$$

$$E_{\text{total}} = E_{VdW} - E_{\text{conservation}} \tag{6}$$

Electrostatic and effective distance restraint terms are not incorporated into the scoring function to avoid impairing the correlation provided by the conservation term. Especially the electrostatic term would have large enough values to suppress the other terms and could bias the search in wrong directions with its random-walk-like behavior. We did, however, include the electrostatic term when minimizing the refined structures.

3. Results

We tested our refinement method on structures that are docked *in silico* by Shehu Lab for modeling the protein complexes with the following PDB IDs: 1C1Y, 1CSE, 1DS6, 1FLT, 1G4U, 1IKN, 1TX4, and 1WWW. Shehu Lab uses a method based on geometric hashing to produce dimeric structures from two disjoint molecules.[17,18] We compare the lRMSD values of those docked structures and their refined versions by our method against the

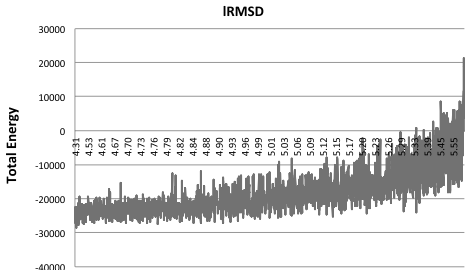
native structures. Most of the refined structures produced by our method have better lRMSD values than the initial structures with respect to the native structures.

For each docked structure, the refinement is performed iteratively in two steps. In the first step, 100 random conformations are generated from the input docked structure as described in Sec. 2.2. These 100 conformations are ranked using our scoring function and the best 10 conformations with lowest interface energies are fed into the second step for further refinement. In the second step, 100 new random conformations are created for each of the 10 conformations produced in the first step. Then, these 1000 new conformations, created in the second step, are ranked via the scoring function and the best 10 conformations with lowest interface energies are returned. We found that further iterations do not significantly improve the quality of the results.

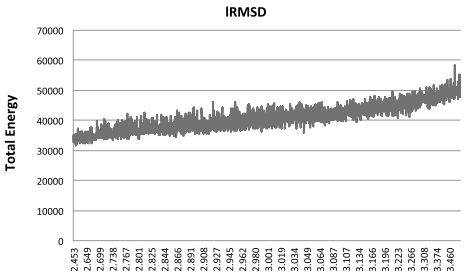
Refinement results of our program for 1C1Y-AB, 1CSE-EI, 1DS6-AB, 1FLT-VY, 1IKN-AC, 1TX4-AB, and 1WWW-WY, are shown in Table 2. In addition, a correlation chart between total interface energy and lRMSD for each protein docked structure is depicted in Fig. 2.

Table 2. Refinement results. lRMSD values in Å with respect to the native structure, total number of interface atoms, and number of conserved interface atoms are shown for the initial docked structures and the top-10 solutions generated by our method for each initial docked structure.

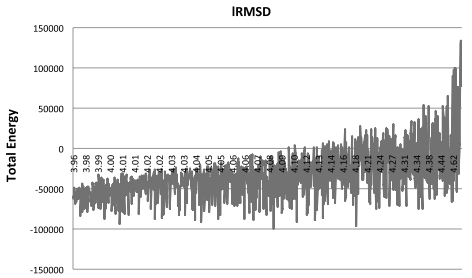
	1C1Y	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	5.46	4.35	4.42	4.77	4.50	4.32	4.42	4.91	4.49	4.73	4.69
Total int. size	813	620	680	665	689	646	662	665	658	683	682
Cons. int. size	613	505	541	537	547	516	528	537	528	548	546
	1CSE	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	3.35	2.58	2.58	2.64	2.56	2.67	2.65	2.66	2.66	2.58	2.62
Total int. size	879	709	725	744	722	735	732	713	724	716	729
Cons. int. size	530	450	467	482	466	476	474	451	466	455	471
	1DS6	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	4.32	4.08	4.18	4.00	4.11	4.04	4.07	4.12	4.04	4.02	4.06
Total int. size	1523	1310	1234	1326	1246	1284	1287	1248	1363	969	912
Cons. int. size	1006	953	902	952	904	921	924	902	972	1361	1270
	1FLT	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	6.00	5.55	5.76	4.90	5.05	5.25	5.36	5.06	5.32	5.89	5.18
Total int. size	711	526	549	514	522	524	539	528	532	556	527
Cons. int. size	326	226	241	219	222	224	238	231	231	244	227
	IKN	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	5.19	5.47	5.90	5.85	5.86	5.82	5.48	5.43	5.81	5.83	6.00
Total int. size	81	110	107	105	107	107	104	105	106	106	103
Cons. int. size	72	91	91	90	92	92	89	89	91	90	90
	ITX4	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	5.19	4.94	5.00	4.88	5.19	5.15	5.13	5.06	5.16	5.12	5.11
Total int. size	2909	2695	2761	2693	2865	2857	2863	2779	2868	2774	2872
Cons. int. size	2151	2035	2074	2030	2139	2133	2142	2079	2135	2078	2142
	IWWW	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	Sol7	Sol8	Sol9	Sol10
lRMSD	5.56	4.91	4.87	4.89	4.85	4.95	4.93	4.83	4.83	5.05	4.97
Total int. size	834	437	402	426	378	441	450	391	393	462	420
Cons. int. size	485	261	234	251	215	262	269	227	227	280	246



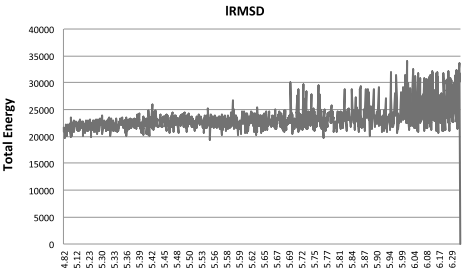
(a) 1C1Y Total energy vs. lRMSD



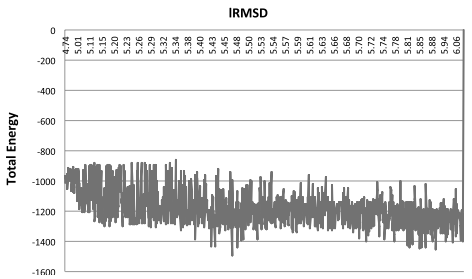
(b) 1CSE Total energy vs. lRMSD



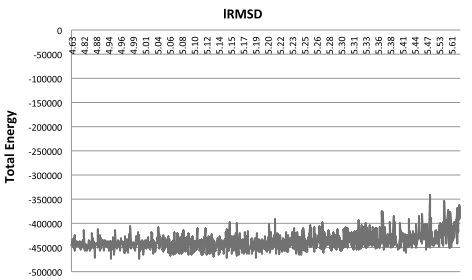
(c) 1DS6 Total energy vs. lRMSD



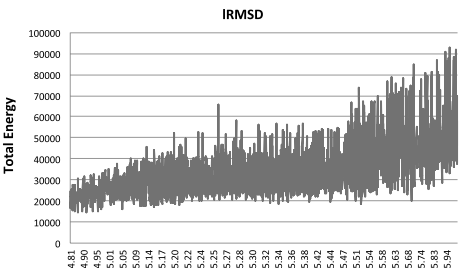
(d) 1FLT Total energy vs. lRMSD



(e) 1IKN Total energy vs. lRMSD



(f) 1TX4 Total energy vs. lRMSD

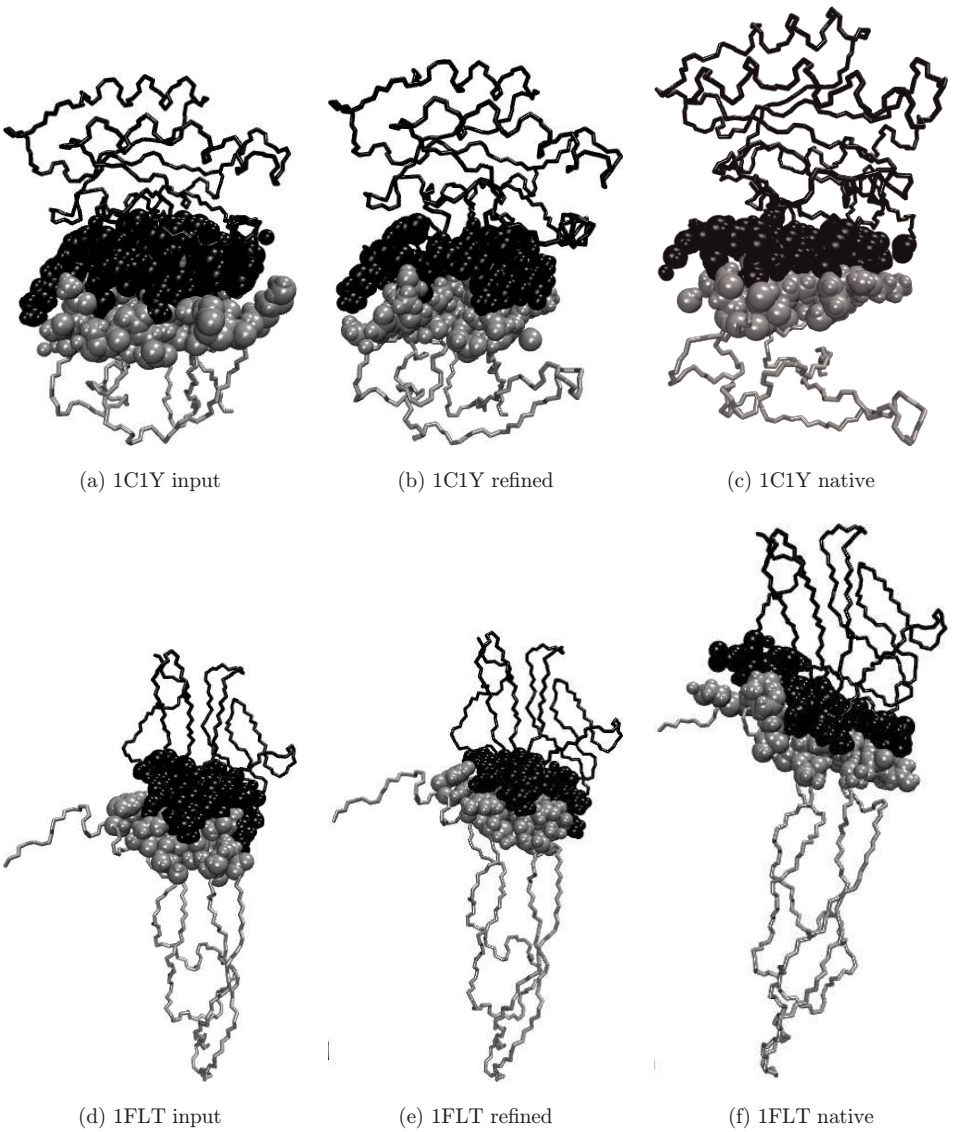


(g) 1WWW Total energy vs. lRMSD

Fig. 2. Correlations between lRMSD values and total energy values computed by our scoring function. For 1C1Y-AB, 1CSE-EI, 1DS6-AB, 1FLT-VY, 1TX4-AB, and 1WWW-WY, conformations with lower interface energy values tend to have lower lRMSD values and are closer to the native conformation. Notice the change of scale in the y-axis due to the difference in the interface size.

For 1C1Y-AB, 1CSE-EI, 1DS6-AB, 1FLT-VY, 1TX4-AB, and 1WWW-WY, we see clearly in Fig. 2 that conformations with lower interface energy values tend to have lower lRMSD values and are closer to the native conformation:

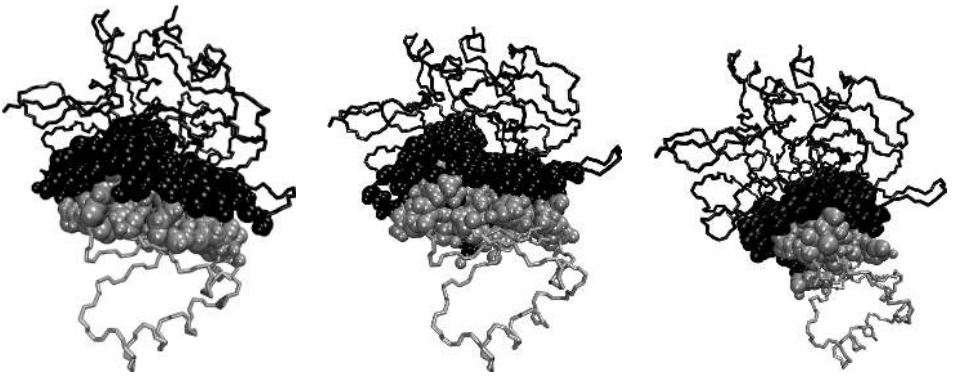
- for 1C1Y-AB, our solutions improve lRMSD values up to 21% and all 10 solutions are better than the input docked structure,



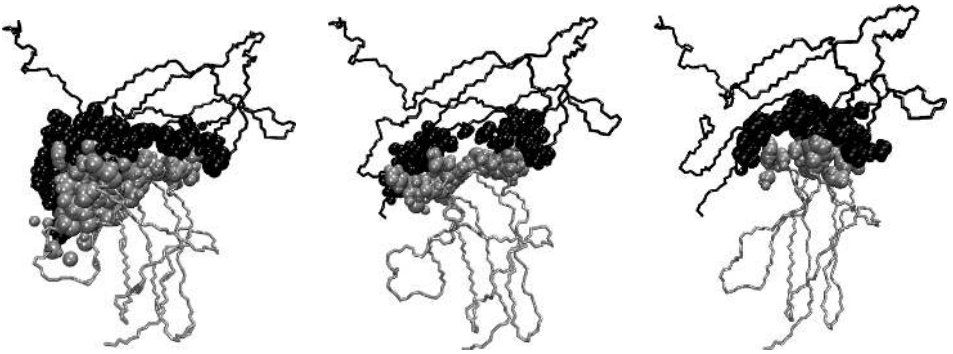
(a) 1C1Y input (b) 1C1Y refined (c) 1C1Y native

(d) 1FLT input (e) 1FLT refined (f) 1FLT native

Fig. 3. Interface atoms are represented as spheres for the initial docked structures (left column), the refined structures (middle column) and the native structures (right column).



(g) 1CSE input (h) 1CSE refined (i) 1CSE native



(j) 1WWW input (k) 1WWW refined (l) 1WWW native

Fig. 3. (*Continued*)

- for 1CSE-EI, our solutions improve lRMSD values up to 23% and all 10 solutions are better than the input docked structure,
- for 1DS6-AB, our solutions improve lRMSD values up to 7% and all 10 solutions are better than the input docked structure,
- for 1FLT-VY, our solutions improve lRMSD values up to 19% and all 10 solutions are better than the input docked structure,
- for 1TX4-AB, our solutions improve lRMSD values up to 6% and 9 out of 10 solutions are better than the input docked structure,
- for 1WWW-WY, our solutions improve lRMSD values up to 14% and all 10 solutions are better than the input docked structure.

On the other hand, for 1IKN-AC, none of our solutions has lower lRMSD value than the input docked structure. We believe this could be because the interface between chain A and chain C is very small (only 81 atoms) compared to other proteins. In fact, as shown in Table 1, the correlation coefficient between conservation term and

lRMSD for this structure shows weaker correlation in the opposite direction. We plan to investigate this in the future.

The initial docked structure for 1G4U-RS is another interesting test case, with a 16.52 Å lRMSD value. It was docked on the wrong surface and it was clearly a bad refinement candidate to start with because such big errors cannot be corrected with small scale rotations. Thus when we ran our program on this structure, the lRMSD values did not improve. Nevertheless, we observed that the $E_{\text{conservation}}$ values fluctuated from the order of 10^5 to 0. Especially the worst lRMSD values had $E_{\text{conservation}}$ values close to 0. Normally, the $E_{\text{conservation}}$ values stay in the range of 10^4 to 10^5 for each structure we tested. We believe this is a promising hint for using our method to distinguish falsely docked complexes from near-native complexes and improve the performance of ranking functions.

4. Discussion

In this work we present a docking refinement method that takes as input a docked complex and uses evolutionary conservation information to generate a set of refined structures with better interface packing. Our method conducts a local greedy geometry-based search in the configuration space for improved binding configurations, picking the top-k configurations at each iteration to avoid exponential growth in the number of tested configurations. We rank the resulting complexes using a novel scoring function based on evolutionary conservation, in addition to the conventional VdW energy term. These terms bias the search toward complexes with evolutionarily conserved residues on the binding interface, and rewards configurations where such conserved residues on the different monomers face each other on the binding site. Such a scoring function can help us refine and re-rank putative docked complexes such that complexes with more conserved residues on the interface will get higher score. In our experiments, we found a strong negative correlation between the evolutionary conservation score and the lRMSD from the native structure. These encouraging results show that near-native complexes indeed tend to have more conserved residues on the interface than false-positive docked complexes. Moreover, we tested our method on several false-positive docking results, where the two monomers were docked at the completely wrong interface. Our results show that these false interfaces contain little or no conserved residues, which makes our method a useful tool to filter out these false-positives and improve the ranking of docking methods. This is especially important due to the fact that even the state-of-the-art geometry-based docking methods are still lacking when it comes to correctly predicting the binding interface.

Current and future work includes extending our method to complexes containing more than two monomers and testing more powerful and sophisticated ways to search in the vicinity of the docked complex, adding local flexibility and physicochemical interactions on the binding interface to the current geometry-based search.

Acknowledgments

We thank the group of Dr. Olivier Lichtarge for their help in using the ET server. We also thank members of Haspel's and Shehu's research groups for useful discussions.

References

1. Goodsell DS, Olson AJ, Structural symmetry and protein function, *Annu Rev Biophys Biomolec Struct* **29**:105−153, 2000.
2. Vajda S, Kozakov D, Convergence and combination of methods in protein-protein docking, *Curr Opinion Struct Biol* **19**(2):164−170, 2009.
3. Dominguez C, Boelens R, Bonvin A, Haddock: A protein-protein docking approach based on biochemical or biophysical information, *J Am Chem Soc* **125**(1):1731−1737, 2003.
4. Schneidman-Duchovny D, Inbar Y, Nussinov R, Wolfson HJ, Geometry based flexible and symmetric protein docking, *Proteins: Struct Funct Bioinf* **60**(2):224−231, 2005.
5. Camacho CJ, Vajda S, Protein-protein association kinetics and protein docking, *Curr Opinion Struct Biol* **12**(1):36−40, 2005.
6. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Eyck LFT, Protein docking using continuum electrostatic and geometric fit, *Protein Eng* **14**(2):105−113, 2001.
7. Potluri S, Yan AK, Chou JJ, Donald BR, Bailey-Kellog C, Structure determination of symmetric homo-oligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing, *Proteins: Struct Funct Bioinf* **65**(1):203−219, 2006.
8. Madaoui H, Guerois R, Coevolution at protein complex interfaces can be detected by the complementarity trace with important impact for predictive docking, *Proc Natl Acad Sci* **105**(22):7708−7713, 2008.
9. Lichtarge O, Bourne HR, Cohen FE, An evolutionary trace method defines binding surfaces common to protein families, *J Mol Biol* **257**(2):342−358, 1996.
10. Baylor College of Medicine Lichtarge Computational Lab, Evolutionary Trace Server. Available: http://mammoth.bcm.tmc.edu/ETserver.html.
11. Engelen S, Ladislas AT, Sacquin-More S, Lavery R, Carbone A, Joint Evolutionary Trees: A large-scale method to predict protein interfaces based on sequence sampling, *PLoS Comp Bio* **5**(1):e1000267, 2009.
12. Kanamori E, Murakami Y, Tsuchiya Y, Standley D, Nakamura H, Kinoshita K, Docking of protein molecular surfaces with evolutionary trace analysis, *Proteins: Struct Funct Bioinf* **69**(4):832−838, 2007.
13. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K, Scalable molecular dynamics with NAMD, *J Comput Chem* **26**:1781−1802, 2005.
14. Akbal- Delibas B, Hashmi I, Shehu A, Haspel N, Refinement of protein complex structures using evolutionary traces, *Proc. Comput Struct Biol Workshop (CSBW) in Conjunction with IEEE BIBM*, Atlanta GA, USA, pp. 400−405, 2011.
15. Craig JJ, *Introduction to Robotics: Mechanics and Control*, 2nd edn., Addison-Wesley Longman Publishing Co. Inc., Boston, MA, USA, 1989.
16. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J Comput Chem* **24**(16):1999−2012, 2003.

17. Hashmi I, Akbal-Delibas B, Haspel N, Shehu A, Protein docking with information on evolutionary conserved interfaces, *Proc Comput Struct Biol Workshop (CSBW)* in conjunction with IEEE BIBM, Atlanta GA, USA, pp. 358−365, 2011.
18. Hashmi I, Akbal-Delibas B, Haspel N, Shehu A, Guiding protein docking with geometric and evolutionary information, *J Bioinform Comput Biol* **10**(3):1242008.

Bahar Akbal-Delibas is pursuing her Ph.D. in Computer Science at University of Massachusetts Boston. She received her M.S. in Computer Science from University of Massachusetts Boston and B.S. in Computer Engineering from Fatih University in Turkey. Her research interests are in better understanding protein−protein interactions through modeling of conformational changes in protein chains and protein-based assemblies.

Irina Hashmi is pursuing her Ph.D. in Computer Science at George Mason University. She received her B.S. in 2007 and M.S. in 2009 in Computer Science and Engineering from University of Dhaka, Bangladesh. Her research interests are in computational structural biology, evolutionary computation, and reversible computing. Her current work focuses on high-dimensional search space problems related to protein docking.

Amarda Shehu is an Assistant Professor in the Department of Computer Science at George Mason University. She holds affiliated appointments in the Department of Bioinformatics and Computational Biology and the Department of Bioengineering at George Mason University. She received her B.S. in Computer Science and Mathematics from Clarkson University in Potsdam, NY, and her Ph.D. in Computer Science from Rice University in Houston, TX, where she was an NIH fellow of the Nanobiology Training Program of the Gulf Coast Consortia. Shehu's research contributions are in computational structural biology, biophysics, and bioinformatics, with a focus on issues concerning the relationship between sequence, structure, dynamics, and function in biological molecules. Shehu is a recent recipient of an NSF CAREER award for her research on probabilistic search algorithms for protein conformational spaces.

Nurit Haspel is an Assistant Professor in the Department of Computer Science at the University of Massachusetts Boston. She received her B.Sc. in Chemistry and Computer Science, her M.Sc in Human Genetics, and Ph.D in Computer Science from Tel Aviv University in Israel. She was trained as a postdoctoral research associate at Rice University in Houston, TX. Haspel's research contributions are in structural bioinformatics with emphasis on studying the structure, function, and dynamics of proteins and protein complexes, as well as the design of novel nano-structures and the study of molecular self-assembly. Her research is funded in part by the National Science Foundation.