BIOINFORMATICS ORIGINAL PAPER Vol. 21 no. 8 2005, pages 1472–1478
doi:10.1093/bioinformatics/bti229

Structural bioinformatics

M-ZDOCK: a grid-based approach for C_n symmetric multimer docking

Brian Pierce[1], Weiwei Tong[2] and Zhiping Weng[1,2,*]

[1]Bioinformatics Program and [2]Department of Biomedical Engineering, Boston University, Boston, MA, USA

Received on October 13, 2004; revised on November 10, 2004; accepted on December 13, 2004
Advance Access publication December 21, 2004

ABSTRACT

Summary: Computational protein docking is a useful technique for gaining insights into protein interactions. We have developed an algorithm M-ZDOCK for predicting the structure of cyclically symmetric (C_n) multimers based on the structure of an unbound (or partially bound) monomer. Using a grid-based Fast Fourier Transform approach, a space of exclusively symmetric multimers is searched for the best structure. This leads to improvements both in accuracy and running time over the alternative, which is to run a binary docking program ZDOCK and filter the results for near-symmetry. The accuracy is improved because fewer false positives are considered in the search, thus hits are not as easily overlooked. By searching four instead of six degrees of freedom, the required amount of computation is reduced. This program has been tested on several known multimer complexes from the Protein DataBank, including four unbound multimers: three trimers and a pentamer. For all of these cases, M-ZDOCK was able to find at least one hit, whereas only two of the four testcases had hits when using ZDOCK and a symmetry filter. In addition, the running times are 30–40% faster for M-ZDOCK.

Availability: M-ZDOCK is freely available to academic users at http://zlab.bu.edu/m-zdock/

Contact: zhiping@bu.edu

Supplementary information: http://zlab.bu.edu/m-zdock

INTRODUCTION

Much of the activity of cells is guided by interactions between proteins. In order to better understand the workings of cells and for rational drug development, it is useful to understand these protein–protein interactions. One means of revealing information about protein interactions is the prediction of the structure of a protein complex based on the structures of two individually crystallized subunits. This problem is referred to as unbound docking, as opposed to the simpler (and largely solved) bound docking which is to predict the structure based on subunit coordinates taken directly from the bound structure. In order to simplify unbound docking, it is generally divided into two steps, the initial stage and the refinement stage. The initial stage is a full search of the six-dimensional (6D) space (three rotational degrees and three translational degrees) for the possible relative orientations of the two molecules. In order to make this search tractable, the proteins are assumed to be rigid during this stage, with allowance for some clash between the proteins (referred to as soft docking). The next stage, the refinement stage, performs slight improvements on a subset of the predictions from the initial docking stage. In addition to slight movements of the rigid bodies in 6D space, the refinement stage sometimes allows for movements of side chains and backbones (referred to as flexible in this case).

For initial stage docking, a variety of approaches have been developed; they are discussed in several reviews (Lengauer and Rarey, 1996; Halperin *et al*., 2002). A popular approach, using a fast Fourier transform (FFT) correlation-based method to test for surface complementarity, was first proposed by Katchalski-Katzir *et al*. (1992). The programs DOT (Mandell *et al*., 2001), GRAMM (Vakser, 1995), FTDOCK (Gabb *et al*., 1997) and ZDOCK (Chen *et al*., 2003a) all use (and expand upon) this concept successfully to predict protein complex structures. ZDOCK, developed by our lab, uses FFT correlations to find complexes based on desolvation and electrostatics, in addition to a surface complementarity metric called PSC (Chen and Weng, 2003).

A subclass of interactions between proteins is the case where two or more identical proteins interact to form a homomultimer. A common form of symmetry found in homomultimers is C_n symmetry or cyclic symmetry, which delineates a ring-shaped complex. For symmetric dimers, trimers, pentamers and heptamers, this symmetry is necessarily the case, while this symmetry is also found for other numbers of protein subunits. For instance, membrane channels and chaperonins often have oligomers with C_n symmetry.

To efficiently and accurately predict C_n multimer complexes, we have implemented a program called Multimer ZDOCK (M-ZDOCK). This program takes advantage of the properties of C_n symmetry to perform a simplified search for the correct complex.

There are many instances where this program can be applied. A number of proteins have been solved as monomers or in a complex with another protein but exist in a homomultimeric state under different conditions *in vivo* (e.g. heat shock, pH changes, viral fusion).

The recently solved crystal structure of adeno-associated virus Rep40 provides evidence that it oligomerizes for nucleotide binding, possibly as a C_n hexamer (James *et al*., 2004). Using M-ZDOCK with this monomeric crystal structure as input, the structure of the hexamer can be modeled. Another example is the protein Chaperonin-60 (Cpn60), which is expressed under heat shock and other forms of stress, is a homolog of *Escherichia coli* GroEL and is typically found in a double ring structure composed of 14 protomers. However, it has been found that *Mycobacterium tuberculosis* has lower order oligomers of this protein (Qamra *et al*., 2004). A ring-shaped model for

*To whom correspondence should be addressed.

this structure can be obtained by multimer docking. Also, Korkhov *et al*. (2004) have devised a model for the dimeric structure of GABA transporter-1 (GAT1). By computationally predicting possible structures of dimeric GAT1, multimer docking would help to support this model or provide new ones regarding the structure.

Since the interface between two adjacent subunits is the same for all interfaces of the complex, only one of the n interfaces needs to be considered, reducing the problem to two monomers for any degree of C_n symmetry. In addition, since all C_n multimers can be aligned in a plane (as they are rotated around a single axis), one spatial degree of freedom can be ignored. Finally, since there is redundancy when rotating a C_n complex around its rotational axis (the resultant complex will be the same), this rotational degree of freedom is eliminated. Thus the problem becomes 4D instead of 6D; this reduces the amount of searching and the computational time.

Another type of symmetry seen in proteins is dihedral (D_2) symmetry, which is composed of two homodimers interacting symmetrically, or a dimer of dimers (four asymmetric units). From an interaction standpoint, this case differs from C_n symmetry in that there are two interfaces to predict rather than n identical interfaces as is seen in C_n symmetry. Recently Berchanski and Eisenstein (2003) filtered and combined the pairwise complexes between monomers generated with a FFT-based generic docking algorithm (Katchalski-Katzir *et al*., 1992) to predict the structures of D_2 multimers. They tested the subunits taken directly from the complex structure, as well as homology modeled monomers, and reported promising results. A similar approach was used earlier to construct the helically symmetric protein coat of the tobacco mosaic virus (Eisenstein *et al*., 1997). However, due to the discrete nature of the FFT algorithm, the vast majority of these binary predictions are not symmetric and even the ones that pass the filter are never truly symmetric.

Here we have developed a new docking algorithm M-ZDOCK, such that we explore only the part of search space that conforms to the C_n symmetry. We observe a significant improvement in accuracy, lower redundancy and fewer false positives, as shown in a direct comparison with docking and filtering. In addition, since only perfectly symmetric multimers are explored in the search space, less computational time is required.

In the process of developing M-ZDOCK, we have carefully curated a set of test cases that exist in both monomeric and multimeric forms in physiological conditions. Although small, this set represents an exhaustive search of such test cases in the Protein Data Bank (PDB) (Berman *et al*., 2000). It should prove useful for future docking studies on multimers.

METHODS

Scoring function

The scoring function used by this program is based on the scoring used in the latest version of ZDOCK (Chen *et al*., 2003a). ZDOCK is an initial stage docking algorithm designed to predict the structure of the complex of two proteins, referred to as the receptor and the ligand. It takes into account surface complementarity, electrostatics and desolvation to find the optimal fit between two proteins. Surface complementarity is calculated using pairwise shape complementarity (PSC), which consists of a favorable term determined by the number of atom pairs within a distance cutoff, and a penalty term determined by the number of clashes. Atomic Contact Energy (ACE) (Zhang *et al*., 1997) is used to score desolvation, and the electrostatic term is calculated by applying Coulomb's equation to the partial charges of the ligand in the electrostatic field of the receptor.

The search strategy of ZDOCK is to discretize both ligand and receptor onto a grid and use FFT to determine the best position of the ligand relative to the receptor. This discretization and FFT is performed for a complete set of angular orientations of the ligand (relative to a fixed receptor).

Results have demonstrated that this approach performs well against a docking benchmark and in the international blind test CAPRI (Chen *et al*., 2003b), Critical Assessment of Prediction of Interactions.

Euler angles

The Euler angle conventions used in this paper refer to these successive rotations from the initial configuration, described in Goldstein's Classical Mechanics (Goldstein, 1980):

(1) Rotation by ψ around the z-axis.

(2) Rotation by θ around the original x-axis.

See Figure 1 for a diagram of these rotations. Typically, Euler angles are sets of three angles; in this case, the third angle, ϕ, is not necessary as it would be redundant in the symmetric search (see the Search space section for an explanation).

Rotational/translational search

M-ZDOCK uses the convention that the rotational axis will be parallel to the z-axis, and searches in the x–y plane for the optimal position of this axis. To perform the search for the best conformation of a multimer based on the structure of a monomer, it has been necessary to make modifications to the search methodology that is used for ZDOCK 2.3. The new search algorithm is outlined below:

(1) Center the receptor (the input monomer) at the origin.

(2) Rotate the receptor by an angle ψ around the z-axis, and then θ around the x-axis.

(3) Copy the receptor, and rotate it by $360^\circ/n$ around the z-axis to create the ligand.

(4) Discretize both the ligand and receptor, with a grid spacing of 1.2 Å (the same as ZDOCK 2.3).

(5) Perform the 3D FFT and correlation, and search in the x–y plane for the best scoring multimer position for that rotational orientation.

(6) Repeat the steps 2–5 for various other sets of ψ and θ.

Search space

In order to fully explore the space of multimers, it is necessary to vary ψ from 0 to 360°, and θ from 0 to 90°. θ does not need to sample a full 360° because for a given ϕ there are redundancies at $180^\circ - \theta$, $180^\circ + \theta$, and $-\theta$ due to the symmetric nature of these angles around the z- and x-axes.

It is not necessary to sample ϕ angles (the third rotation around the z-axis), as these are symmetric around the z-axis and therefore would be redundant for the same values of ψ and θ. This corresponds to the loss of a rotational degree of freedom that is referred to in the Introduction section.

M-ZDOCK uses 1500 angle sets, as this was found to be a good balance between computational time and predictive performance. In addition, given that ZDOCK 2.3 uses 54 000 angle sets for 3D angular freedom (6 degree sampling density), the number of angles that M-ZDOCK covers is mathematically reasonable as it is approximately $54\,000^{2/3}$.

Reconstructing the multimer

Based on the optimal relative position of two adjacent monomers in the x–y plane (output from the FFT), it is possible to reconstruct the full multimer. The only constraint is that the monomers need to be rotated by $360^\circ/n$ with respect to one another around the z-axis. Referring to Figure 2, the vector representing the displacement between the two adjacent monomers is **L** and the vector from the monomer to the symmetry axis (in the x–y plane) is **d**. β is the angle around the C_n symmetry axis between two multimer centers of mass, $360^\circ/n$. The angle between the vectors **L** and

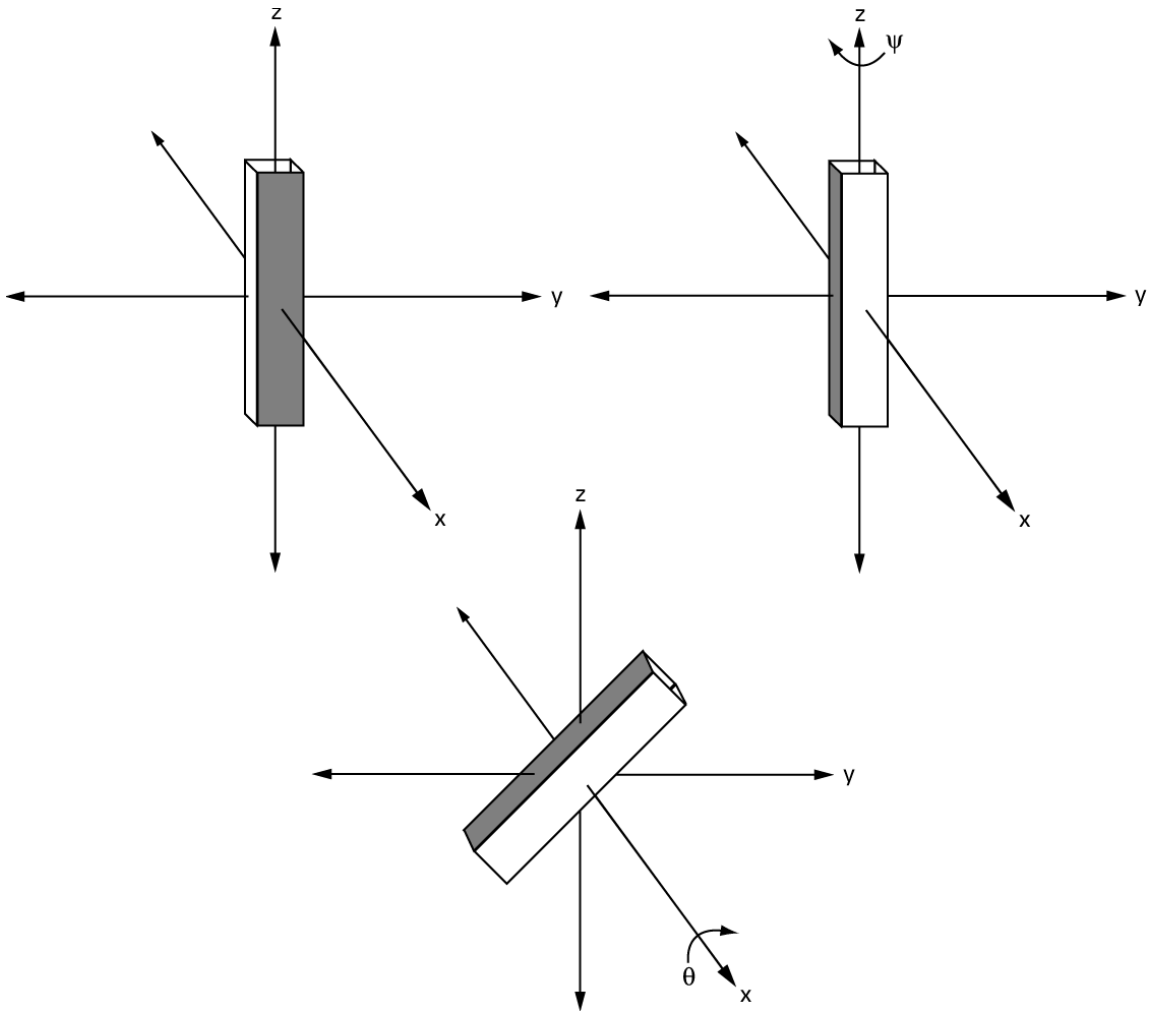

Fig. 1. Diagram of successive rotation through Euler angles ψ and θ the angles used to describe the rotational configuration of the ligand and receptor. In this case, $\psi = 90°$ and $\theta = 45°$.

d is α, given by $(180° - \beta)/2$. The magnitude of **d** can be computed as $L/(2^* \cos(\alpha))$.

Once the rotational axis is found, the monomer needs to be rotated around this axis n times by $\beta°$ to form the multimer. Thus, given the vector between two adjacent monomers in the C_n multimer (and the symmetry number), it is possible to reconstruct the entire multimer. To illustrate this concept, a java applet has been written and is publicly available at http://zlab.bu.edu/m-zdock.

Symmetry filter

In order to compare the results of M-ZDOCK with results from an existing method of docking, we implemented a symmetry filter that will choose only near-symmetric complexes. It is designed to process the results from a docking tool such as ZDOCK which produces many predictions (54 000 in the case of ZDOCK with dense sampling).

The filter determines the angle and axis between the monomers of the prediction, as well as the center of mass translation between the monomers. For perfect symmetry, the angle between the center of mass translation and the axis is 90°, and the angle of rotation around the axis is $360°/n$, but a certain range must be allowed as the predictions are not perfectly symmetric. In the case of Berchanski and Eisenstein (2003) the angular range for the rotation around the axis was ±6°, and between the axis and translation the angular range was ±3°. To allow for a comparison with the M-ZDOCK results so there would be approximately 1500 predictions per testcase these ranges were increased to ±18° and ±9°, respectively.

Multimer testcases

We tested M-ZDOCK with two categories of testcases, bound/quasi-bound and unbound.

Bound and quasi-bound testcases

To ensure that the search space is covered entirely and that the algorithm is valid for various types of C_n symmetry, both bound and quasi-bound docking testcases were used. The bound testcases were generated by extracting the monomer from the multimeric structure so that the docking algorithm can attempt to reassemble the multimer. These testcases should be relatively simple to dock as there is no conformational change to account for. If the correct structure is not found with these cases, it is possibly due to some problem with the searching algorithm. Though found in the PDB as both monomers and multimers, quasi-bound testcases are most likely biological multimers. Therefore the conformational change involved is of little or no significance, making these cases similar to (but slightly more difficult than)

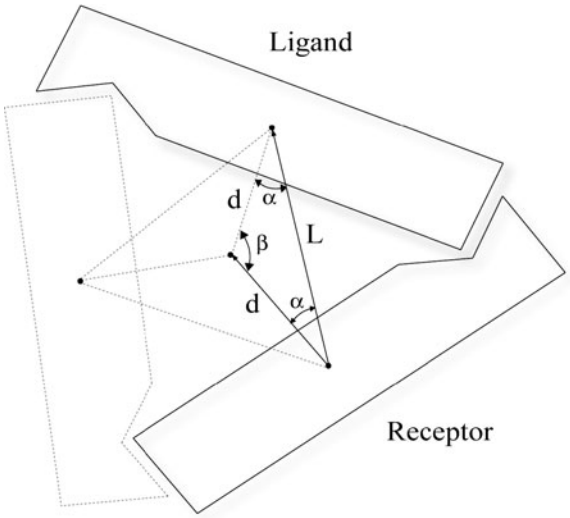

Fig. 2. The relative positions of the subunits of a C_3 multimer. The vector **L** is the relative position between the receptor and the ligand (which is the receptor rotated by β degrees; in this case $\beta = 120°$). The magnitude of vector **d** to the axis of symmetry and the angle α between vectors **L** and **d** can be determined algebraically. Thus, once the interface between the ligand and receptor is evaluated by M-ZDOCK, the rest of the multimer (in this case the subunit represented by the dashed lines) can be generated automatically.

the bound testcases. The monomer structure that is found in the PDB is used as input to the docking algorithm, while the multimer structure in the PDB is used to evaluate the docking results.

Unbound testcases

The second type of testcases is unbound structures. These testcases are significantly more difficult to predict, both due to the conformational change of the proteins inherent in unbound docking and because of the low affinity of the complexes, as these cases must exist in both monomer and multimer forms to be found experimentally. Four proteins were found in the PDB (Berman *et al*., 2000) for which different symmetric forms exist, according to Protein Quaternary Structure server classification. Here is a brief summary of these proteins:

RNase A. Bovine pancreatic RNase A was crystallized in monomeric (Tilton *et al*., 1992) and trimeric (Liu *et al*., 2002) forms. The trimer in this case is one of two trimeric forms thought to exist in mildly acidic conditions. Notable about this structure is a domain-swapped C-terminal beta strand.

Phospholipase A_2. The *Naja naja naja* (Indian cobra) phospholipase A_2 (PLA_2) was obtained from the venom and crystallized in trimeric form (Segelke *et al*., 1998) using random crystallization screening. The monomeric version (Scott *et al*., 1990) is the *Naja naja atra* (Chinese cobra) PLA_2, which was crystallized with a lower concentration of PLA_2 and higher concentration of Ca^{2+} (the Ca^{2+} is seen in the structure of the monomer but not the trimer). In Segelke *et al*. (1998) it is discussed that the trimeric form may be a means of shielding the active site and thus 'protecting the snake from its own venom'.

Flavivirus envelope protein. This is the fusion envelope protein of the tick-borne encephalitis virus (TBEV E protein). The input structure is taken from the homodimer structure (Rey *et al*., 1995). The trimeric form, which occurs at low pH during membrane fusion, was recently solved (Bressanelli *et al*., 2004).

Bovine trypsin inhibitor. This testcase is the bovine pancreatic trypsin inhibitor (BPTI), which occurs as a monomer (Wlodawer *et al*., 1984) at basic pH and a decamer (Hamiaux *et al*., 2000) at acidic pH levels. As the decamer is comprised of two C_5 symmetric pentamers, the target for this case is one half of the decamer.

Table 1. The unbound multimer testcases

Testcase	PDB IDs[a]	Symmetry	RMSD[b]
RNase A	9RAT/1JS0	Trimer	0.33
Phospholipase A_2 (PLA2)	1POA/1A3F	Trimer	0.79
Flavivirus envelope protein (TBEV E)	1SVB/1OML	Trimer	2.08
Bovine trypsin inhibitor (BPTI)	3PTI/1B0C	Pentamer	0.41

[a]The first PDB code is for the structure used as input for docking, while the second one is the bound multimer.
[b]Interface Cα RMSD change between unbound/bound structures.

Table 1 summarizes these testcases. To provide a measure of the difficulty of docking each complex, interface Cα atoms from unbound monomers were fitted to two adjacent subunits of the complex. As M-ZDOCK is a rigid-body docking algorithm, the root mean square deviation (RMSD) in this case can be seen as the lower limit for the RMSD of the predictions.

RMSD calculations and hits

To evaluate bound and unbound predictions, the RMSDs of interface alpha Carbon (Cα) atoms were used. The interface Cα atoms were determined from the crystal structure of the multimer. If any atom of a residue is within 10 Å of any atom of another chain, the Cα atom from that residue is determined to be an interface Cα. In addition, to avoid false negatives due to large domain movements, regions of residues with large movement from unbound to bound (>4 Å) were removed before determining interface Cα atoms. See Supplementary information for the removed residues.

Once the Cα residues are known, two adjacent subunits of the predicted structure are fitted to two adjacent subunits of the complex using the interface Cαs, and the RMSD between the interface Cαs is computed. Hits are defined as predictions that have an interface Cα RMSD $\leq$ 2.5 Å.

RESULTS AND DISCUSSION

Structure prediction: quasi-bound and bound

Eight quasi-bound and bound testcases were used to ensure the coverage and basic functionality of M-ZDOCK for a variety of symmetries. The results in Table 2 clearly demonstrate that M-ZDOCK is capable of predicting structures with C_n symmetry. For all of the structures the number one ranked prediction was a hit, and in addition there were a number of hits in the top 20 for every testcase.

Structure prediction: unbound

The structure prediction capabilities of M-ZDOCK are shown to be superior to filtering normal docking predictions, across the unbound multimer benchmark (Table 3). For M-ZDOCK, all of the first hits are in the top third of the predictions, whereas for filtering there were two cases where no hit was found, and the two other cases were in the bottom third of the predictions.

M-ZDOCK successfully predicted a hit for RNase A (Fig. 3a), while the near-symmetric predictions of ZDOCK failed to produce a hit. This is despite the fact that 375 more predictions were produced by ZDOCK plus filtering. This complex was difficult to predict due to the strand swapping that takes place upon multimerization, which explains the relatively high rank of 476 for the first M-ZDOCK hit.

Table 2. M-ZDOCK results for quasi-bound and bound testcases

Testcase[a]	Symmetry	Hits[b]	Rank[c]	RMSD[d]	References
Quasi-bound testcases					
1NSP/1B99[e]	Trimer	11	1	1.06	(Morera *et al*., 1995; Gonin *et al*., 1999)
1KKU/1GZU[e]	Trimer	7	1	0.88	(Garavaglia *et al*., 2002; Werner *et al*., 2002)
1AUS/1AA1[e]	Tetramer	8	1	0.93	(Taylor and Andersson, 1997)
1EXB/1QRQ	Tetramer	7	1	1.15	(Gulbis *et al*., 1999, 2000)
Bound testcases					
1AF6	Trimer	6	1	0.73	(Wang *et al*., 1997)
1A8R[e]	Pentamer	10	1	0.78	(Auerbach *et al*., 2000)
1QNU	Pentamer	16	1	0.75	(Kitov *et al*., 2000)
1G31	Heptamer	17	1	1.81	(Hunt *et al*., 1997)

[a]PDB IDs of the testcases, with the PDB ID of the input structure for M-ZDOCK listed first for the quasi-bound testcases.
[b]Number of hits in the top 20 (out of 1500) predictions, as ranked by M-ZDOCK.
[c]Rank of the first hit.
[d]RMSD (in Å) of the first hit.
[e]The bound structures in these cases are in fact dimers of the C_n multimer; just the C_n contacts are predicted so the other interface is ignored.

Table 3. M-ZDOCK results for unbound testcases

Testcase	M-ZDOCK				ZDOCK + filtering			
	N_p^a	Hits[b]	Rank[c]	RMSD[d]	N_p^e	Hits[b]	Rank[c]	RMSD[d]
Rnase A	1500	1	476	2.44	1875	0	—	—
PLA_2	1500	6	33	1.11	1595	1	1417	2.05
TBEV E	1500	2	62	2.31	1476	0	—	—
BPTI	1500	20	384	2.25	1164	1	1064	1.83

[a]Number of predictions produced by M-ZDOCK (the number is always 1500).
[b]Number of hits among the predictions.
[c]Rank of the first hit.
[d]RMSD (in Å) of the first hit.
[e]Number of predictions remaining after running ZDOCK and filtering the 54 000 predictions for symmetry.

Although the swapped strands were not included in the RMSD calculation, they were clearly a part of the interface making the prediction non-trivial. The swapped strands are highlighted in Figure 3a.

The symmetric trimer PLA_2 was successfully prediced by M-ZDOCK. In this case M-ZDOCK predicted six hits, one of them with the particularly high rank of 33. While ZDOCK + filtering obtained a hit, the rank of the hit was 1417 and the RMSD of this hit was higher.

Perhaps the most striking results are for the TBEV E protein, where two hits were found by M-ZDOCK, while no hits were found with ZDOCK. This protein is somewhat difficult to dock due to the large C-terminal conformational change upon trimerization that helps to stabilize the interaction. The difficulty is also reflected in the lower bound for the RMSD of 2.08 Å (Table 1), which leaves little room for error for rigid-body docking to obtain a hit (under 2.5 Å). However M-ZDOCK is able to predict this structure, giving the first-ranked hit an impressive rank of 62 (see Fig. 3c for the structure).

The BPTI pentamer (Fig. 3d) had a large number of predictions produced by M-ZDOCK. As with the other testcases, M-ZDOCK performed better with regard to hits and the rank of the first hit. In this case the RMSD of the first hit was slightly better for the ZDOCK prediction. But of the 20 M-ZDOCK hits, 6 of them had a better rank and RMSD than the top ZDOCK prediction, so clearly M-ZDOCK is superior in this case as well.

Computational performance

The docking predictions reported in this study were performed on an IBM p690 workstation with 32 1.3 Ghz Power4 processors, using MPI for parallelization. Due to the increased efficiency of the M-ZDOCK search, a significant reduction in running time can be seen using this approach. The discretization of the receptor at every angle set (as described in the Methods section), which costs more than regular ZDOCK, is more than compensated by the faster search. On average, M-ZDOCK runs 30–40% faster than ZDOCK.

M-ZDOCK was also compiled and run on Linux in serial and parallel, and on Mac OS X in serial. Versions of M-ZDOCK for all of these platforms are available at: http://zlab.bu.edu/m-zdock.

Future work

A possible future modification to the M-ZDOCK algorithm would be to incorporate the degree of packing into the algorithm. Since the algorithm currently used considers only the interface between

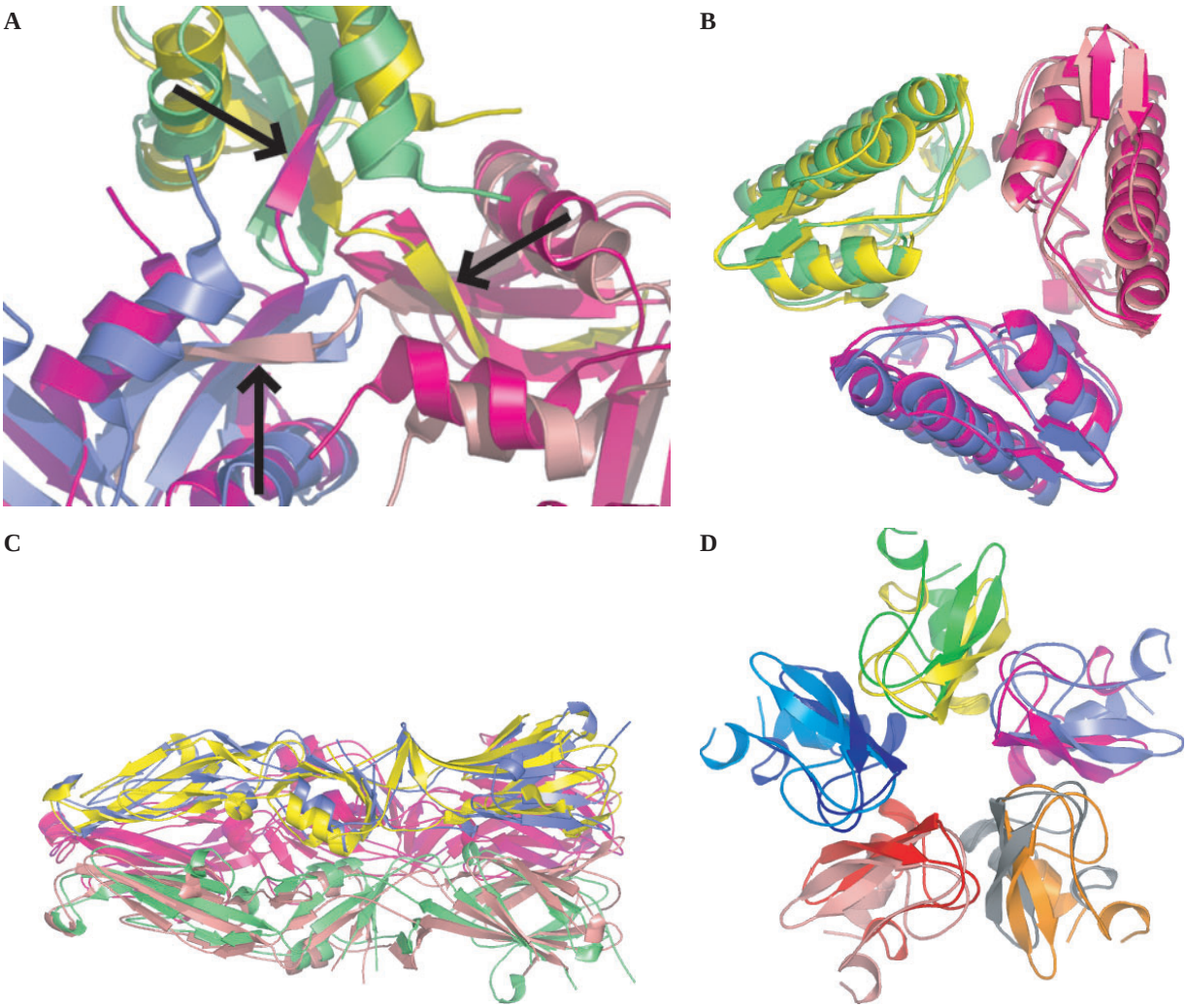

Fig. 3. The highest ranked hits of M-ZDOCK superposed onto the structures of the complexes, with each predicted and actual chain colored separately. Images generated with PyMOL (Delano, 2002). (**a**) The interface region of RNase A, with arrows indicating the β-strands that swap in the trimeric structure. Predicted: magenta, green, blue; Actual: purple, pink, yellow (**b**) Phospholipase A_2. Predicted: purple, pink, yellow; Actual: magenta, green, blue (**c**) TBEV E Protein, truncated at the highly mobile C-terminal domain. Predicted: purple, pink, yellow; Actual: magenta, green, blue (**d**) Bovine pancreatic trypsin inhibitor. Predicted: magenta, yellow, blue, red, grey; Actual: slate, green, aqua, peach, orange.

adjacent subunits, interactions between non-adjacent subunits is ignored. This would possibly be an issue in the case of a multimer that has a structure similar to the spokes of a wheel, i.e. tightly packed (versus a 'daisy chain' that is shown in Fig. 2). However, these non-neighbor interactions would clearly be less significant overall than the interactions taking place in the interface between adjacent subunits.

SUMMARY

There are several advantages to using M-ZDOCK versus filtering docking predictions from a normal binary docking program:

(1) Greater accuracy, with improvements in both hit count and the rank of the first hit.

(2) Increased efficiency, due to a reduced search space based on the knowledge of C_n symmetry.

(3) Perfectly symmetric multimers are automatically output; there is no need for approximation/fitting to generate the other subunits.

We have shown that it is possible to perform an intelligent search of the space of exclusively symmetric C_n multimers, and have incorporated this into the M-ZDOCK program. Based on its performance, with regards to both accuracy and speed, M-ZDOCK is an effective program for predicting complexes of C_n symmetry based on the structure of its subunit, showing clear superiority over traditional docking and subsequent filtering.

ACKNOWLEDGEMENT

This work was partially funded by NSF grants DBI-0078194, DBI-0133834 and DBI-0116574.

REFERENCES

Auerbach,G., Herrmann,A., Bracher,A., Bader,G., Gutlich,M., Fischer,M., Neukamm,M., Garrido-Franco,M., Richardson,J., Nar,H., Huber,R. and Bacher,A. (2000) Zinc plays a key role in human and bacterial GTP cyclohydrolase I. *Proc. Natl Acad. Sci. USA*, **97**, 13567–13572.

Berchanski,A. and Eisenstein,M. (2003) Construction of molecular assemblies via docking: modeling of tetramers with D2 symmetry. *Proteins*, **53**, 817–829.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Bressanelli,S., Stiasny,K., Allison,S.L., Stura,E.A., Duquerroy,S., Lescar,J., Heinz,F.X. and Rey,F.A. (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J.*, **23**, 728–738.

Chen,R. and Weng,Z. (2003) A novel shape complementarity scoring function for protein–protein docking. *Proteins*, **51**, 397–408.

Chen,R., Li,L. and Weng,Z. (2003a) ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, **52**, 80–87.

Chen,R., Tong,W., Mintseris,J., Li,L. and Weng,Z. (2003b) ZDOCK predictions for the CAPRI challenge. *Proteins*, **52**, 68–73.

Delano,W.L. (2002). The PyMOL Molecular Graphics System.

Eisenstein,M., Shariv,I., Koren,G., Friesem,A.A. and Katchalski-Katzir,E. (1997) Modeling supra-molecular helices: extension of the molecular surface recognition algorithm and application to the protein coat of the tobacco mosaic virus. *J. Mol. Biol.*, **266**, 135–143.

Gabb,H.A., Jackson,R.M. and Sternberg,M.J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.*, **272**, 106–120.

Garavaglia,S., D'Angelo,I., Emanuelli,M., Carnevali,F., Pierella,F., Magni,G. and Rizzi,M. (2002) Structure of human NMN adenylyltransferase. A key nuclear enzyme for NAD homeostasis. *J. Biol. Chem.*, **277**, 8524–8530.

Goldstein,H. (1980). *Classical Mechanics*. Addison-Wesley Publishing Co., Reading, Mass.

Gonin,P., Xu,Y., Milon,L., Dabernat,S., Morr,M., Kumar,R., Lacombe,M.L., Janin,J. and Lascu,I. (1999) Catalytic mechanism of nucleoside diphosphate kinase investigated using nucleotide analogues, viscosity effects, and X-ray crystallography. *Biochemistry*, **38**, 7265–7272.

Gulbis,J.M., Mann,S. and MacKinnon,R. (1999) Structure of a voltage-dependent K^+ channel beta subunit. *Cell*, **97**, 943–952.

Gulbis,J.M., Zhou,M., Mann,S. and MacKinnon,R. (2000) Structure of the cytoplasmic beta subunit-T_1 assembly of voltage-dependent K^+ channels. *Science*, **289**, 123–127.

Halperin,I., Ma,B., Wolfson,H. and Nussinov,R. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*, **47**, 409–443.

Hamiaux,C., Perez,J., Prange,T., Veesler,S., Ries-Kautt,M. and Vachette,P. (2000) The BPTI decamer observed in acidic pH crystal forms pre-exists as a stable species in solution. *J. Mol. Biol.*, **297**, 697–712.

Hunt,J.F., van der Vies,S.M., Henry,L. and Deisenhofer,J. (1997) Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell*, **90**, 361–371.

James,J.A., Aggarwal,A.K., Linden,R.M. and Escalante,C.R. (2004) Structure of adeno-associated virus type 2 Rep40–ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases. *Proc. Natl Acad. Sci. USA*, **101**, 12455–12460.

Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A.A., Aflalo,C. and Vakser,I.A. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.

Kitov,P.I., Sadowska,J.M., Mulvey,G., Armstrong,G.D., Ling,H., Pannu,N.S., Read,R.J. and Bundle,D.R. (2000) Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature*, **403**, 669–672.

Korkhov,V.M., Farhan,H., Freissmuth,M. and Sitte,H.H. (2004) Oligomerization of the GABA transporter-1 is driven by an interplay of polar and hydrophobic interactions in transmembrane helix II. *J. Biol. Chem.*, **279**, 55728–55736.

Lengauer,T. and Rarey,M. (1996) Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.*, **6**, 402–406.

Liu,Y., Gotte,G., Libonati,M. and Eisenberg,D. (2002) Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci.*, **11**, 371–380.

Mandell,J.G., Roberts,V.A., Pique,M.E., Kotlovyi,V., Mitchell,J.C., Nelson,E., Tsigelny,I. and Ten Eyck,L.F. (2001) Protein docking using continuum electrostatics and geometric fit. *Protein Eng.*, **14**, 105–113.

Morera,S., Chiadmi,M., LeBras,G., Lascu,I. and Janin,J. (1995) Mechanism of phosphate transfer by nucleoside diphosphate kinase: X-ray structures of the phosphohistidine intermediate of the enzymes from *Drosophila* and *Dictyostelium*. *Biochemistry*, **34**, 11062–11070.

Qamra,R., Srinivas,V. and Mande,S.C. (2004) *Mycobacterium tuberculosis* GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. *J. Mol. Biol.*, **342**, 605–617.

Rey,F.A., Heinz,F.X., Mandl,C., Kunz,C. and Harrison,S.C. (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. *Nature*, **375**, 291–298.

Scott,D.L., White,S.P., Otwinowski,Z., Yuan,W., Gelb,M.H. and Sigler,P.B. (1990) Interfacial catalysis: the mechanism of phospholipase A2. *Science*, **250**, 1541–1546.

Segelke,B.W., Nguyen,D., Chee,R., Xuong,N.H. and Dennis,E.A. (1998) Structures of two novel crystal forms of *Naja naja naja* phospholipase A2 lacking Ca^{2+} reveal trimeric packing. *J. Mol. Biol.*, **279**, 223–232.

Taylor,T.C. and Andersson,I. (1997) Structure of a product complex of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry*, **36**, 4041–4046.

Tilton,R.F.,Jr Dewan,J.C. and Petsko,G.A. (1992) Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. *Biochemistry*, **31**, 2469–2481.

Vakser,I.A. (1995) Protein docking for low-resolution structures. *Protein Eng.*, **8**, 371–377.

Wang,Y.F., Dutzler,R., Rizkallah,P.J., Rosenbusch,J.P. and Schirmer,T. (1997) Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin. *J. Mol. Biol.*, **272**, 56–63.

Werner,E., Ziegler,M., Lerner,F., Schweiger,M. and Heinemann,U. (2002) Crystal structure of human nicotinamide mononucleotide adenylyltransferase in complex with NMN. *FEBS Lett.*, **516**, 239–244.

Wlodawer,A., Walter,J., Huber,R. and Sjolin,L. (1984) Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J. Mol. Biol.*, **180**, 301–329.

Zhang,C., Vasmatzis,G., Cornette,J.L. and DeLisi,C. (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.*, **267**, 707–726.