Published in final edited form as:

Proteins. 2011 September ; 79(9): 2607–2623. doi:10.1002/prot.23082.

SymmRef: a Flexible Refinement Method for Symmetric Multimers

Efrat Mashiach-Farkash¹, Ruth Nussinov^{2,3}, and Haim J. Wolfson^{1,*}

¹ Blavatnik School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

² Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI - Frederick, Frederick, MD 21702, USA

³ Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Abstract

Symmetric protein complexes are abundant in the living cell. Predicting their atomic structure can shed light on the mechanism of many important biological processes. Symmetric docking methods aim to predict the structure of these complexes given the unbound structure of a single monomer, or its model. Symmetry constraints reduce the search-space of these methods and make the prediction easier compared to asymmetric protein-protein docking. However the challenge of modeling the conformational changes that the monomer might undergo is a major obstacle. In this paper we present SymmRef, a novel method for refinement and re-ranking of symmetric docking solutions. The method models backbone and side-chain movements and optimizes the rigid-body orientations of the monomers. The backbone movements are modeled by normal modes minimization and the conformations of the side-chains are modeled by selecting optimal rotamers. Since solved structures of symmetric multimers show asymmetric side-chain conformations, we do not use symmetry constraints in the side-chain optimization procedure. The refined models are re-ranked according to an energy score. We tested the method on a benchmark of unbound docking challenges. The results show that the method significantly improves the accuracy and the ranking of symmetric rigid docking solutions. SymmRef is available for download at <http://bioinfo3d.cs.tau.ac.il/SymmRef/download.html>.

Keywords

protein docking; symmetric complexes; docking refinement; side-chain optimization; backbone refinement; symmetric refinement

Introduction

The majority of protein-protein complexes are symmetric oligomers^{1,2}. Symmetric complexes have many functional, genetic and physicochemical advantages, which can explain their abundance in the living cell². Symmetric protein complexes often form structural arrangements with specific functions, such as channels, containers, and filaments.

* Address correspondence to H.J. Wolfson at School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. TEL: +972-3-640-8268, FAX: +972-3-640-6476, wolfson@tau.ac.il.

Monod et al.³ suggested that cyclic symmetric complexes are common because they occupy all of the available binding sites, which reduces aggregation. Moreover, once initially formed, the impact of mutations during evolution is doubled compared to heterooligomers, which might increase the evolutionary rate of improving the binding affinity. However, André et al.⁴ clarified that although the influence of each mutation is doubled, the number of unique amino acids in homo-oligomers is half of that in hetero-oligomers and therefore the mutation rate is halved. They suggested an alternative theory that explains the high frequency of symmetric complexes. When the mutation rate is halved but the impact of each mutation is doubled, the mean of the energetic contributions does not change, but the variance doubles. Increasing the variance drastically improves the probability of symmetric protein complexes with very-low energy to initially be formed by evolution.

In nature, one can find few types of symmetry. The most common is the *cyclic symmetry* (C_n), which contains a single axis of rotational symmetry². C_2 symmetry (homodimers) is the most frequent among the various C_n symmetries. *Dihedral symmetry* (D_n) is another common type. These complexes combine an axis of rotational symmetry and a perpendicular axis of two-fold symmetry. *Icosahedral symmetry* produces spherical assemblies and *helical symmetry* is produced by rotation and translation along a single symmetry axis.

Despite their high frequency, the large size of symmetric complexes makes them hard to be structurally characterized using experimental means. However, computational prediction is significantly facilitated when symmetry constraints are applied. In docking, symmetry provides powerful selectivity that considerably improves the success rate compared to general protein-protein docking. The number of degrees of freedom in the rigid-docking search-space of symmetric complexes is four, compared to six in heterodimers. André et al.⁴ have demonstrated the power of the symmetric selectivity by showing that only one out of 5,000 random orientations of two identical monomers is near-symmetric (with a symmetry measure that is typical for known symmetric dimers).

In practice, docking methods should attempt to predict the structure of oligomers given the unbound conformation of the monomers. Since proteins often undergo conformational changes upon binding which include both backbone and side-chain movements, this makes the docking problem much more difficult.

To date there are six major methods for docking prediction of symmetric multimers: MolFit^{5,6}, SymmDock^{7,8}, ClusPro⁹, M-ZDOCK¹⁰, ROSETTA¹¹ and HADDOCK¹². These methods can handle multimers with different types of symmetry and some of them model conformational changes of the monomers, as detailed in Table I.

Berchanski and Eisenstein were the first to tackle the challenge of symmetric docking by developing a method for predicting the structures of homo-multimers with D_2 symmetry⁶. First, structures of dimers are predicted by an FFT-based docking method¹³. Then, the solutions are filtered and only the symmetry-related homodimers are selected. Finally, the full D_2 complex is assembled by re-docking each dimer solution from the previous step to itself or by superimposing the monomers of two different dimer solutions. This method was later extended to handle C_n and D_n symmetries⁵.

SymmDock is a geometry-based rigid-docking method for protein complexes with cyclic symmetry^{7,8}. The method calculates matching pairs of surface points across the interface, computes the symmetry axes consistent with large numbers of such matching pairs, and then ranks the resulting candidate symmetric solutions by a shape complementarity score. SymmDock is extremely fast. It is able to successfully predict the structure of symmetric complexes with C_n symmetry in a few minutes on a standard desktop PC, for any n . The

efficiency of the method results from the fact that it a-priori limits the search space to symmetric transformations.

M-ZDOCK performs symmetric rigid-docking by an FFT approach that explores only the search space of C_n symmetric rigid-transformation (four degrees of freedom)¹⁰. The method searches for symmetric complexes that optimize not only a surface complementarity metric but also desolvation and electrostatics energy terms.

The ClusPro symmetric docking method attempts to predict both the symmetry type (cyclic or dihedral) and the structure of the complex⁹. It starts by generating a large set of rigid body docking solutions of a monomer with itself, using the DOT program¹⁴. It identifies the 2000 energetically favorable solutions and scores all the solutions of DOT according to the number of their energetically favorable neighbor solutions. Then, the algorithm generates all possible symmetric complexes that are induced by the rigid-docking solutions (cyclic and dihedral symmetry). Finally, it clusters and ranks them according to the scores of the docking solutions, from which they were built.

The ROSETTA symmetric docking method optimizes the rigid-body orientations of the monomers and their backbone and side-chain conformations, while restricting the search space to symmetrical models¹¹. The optimizations are performed by a Monte-Carlo-Minimization (MCM) protocol, starting from random symmetrical configurations.

The HADDOCK multi-body docking method¹² requires experimental data (and/or bioinformatics data) for docking. The method minimizes a function which includes: various energetic terms, a term of ambiguous interaction restraints (AIRs), determined by the available experimental data, and terms of symmetry restraints. The minimization accounts for torsion angle flexibility in flexible segments, which are typically defined automatically by intermolecular contacts.

The ROSETTA and the HADDOCK symmetric docking methods restrict all the monomers to be identical. However, most of the crystal structures of symmetric protein complexes show local differences in side-chain conformations and small variations in backbone conformation². An example of such local differences is shown in Figure 3A.

In this paper we present SymmRef, a novel docking refinement method for symmetric protein complexes. The method refines a set of rigid docking solution candidates which can be generated by any symmetric docking method (e.g. SymmDock⁷ or M-ZDOCK¹⁰). It models both side-chains and backbone movements and re-ranks the refined models by an energy scoring function. The optimization of side-chain conformations is performed by an Integer Linear-Programming approach. This approach is very efficient and promises to assign the optimal rotamers to the interface side-chains. Unlike other methods, SymmRef does not apply symmetry constraints at the side-chain optimization level. Backbone refinement by SymmRef is performed by normal modes minimization. While previous methods model backbone movements only in pre-selected backbone segments^{11,12}, SymmRef is able to model both local and global conformational changes, which move the entire backbone simultaneously. The results show that the method improves both the accuracy and the ranking of rigid-docking solution candidates, and outperforms existing state-of-the-art symmetric docking methods.

Methods

The SymmRef method refines a set of rigid docking solutions of protein complexes with cyclic symmetry and improves the accuracy and the ranking of the near native models. The

method optimizes the rigid-body orientation of the monomers, assigns optimal rotamers to interface side-chains and minimizes the backbone conformation.

The input to the algorithm is a set of hundreds or thousands docking solutions produced by any docking method of choice (e.g. SymmDock⁷ or M-ZDOCK¹⁰). For each docking solution, the refinement includes the following steps:

1. Repeat N times
 - 1.1 Fast rigid-body Monte-Carlo minimization
 - 1.2 Restricted side-chain optimization
 - 1.3 Backbone refinement
 - 1.4 Extensive rigid-body Monte-Carlo minimization
 - 1.5 Full side-chain optimization
 - 1.6 Scoring
2. Return the solution with the best score

In this study we repeated stage 1 of the algorithm 10 times (N=10). Each iteration results in a different refined solution due to the random Monte-Carlo process. The method is expected to produce better results when more iterations are performed. However, there is a tradeoff between performance and running-time that the user can control. The running time of single refinement iteration depends on the size of the interface and the extent of steric clashes in the starting docking model. The average running time of a single iteration on our data set was around 23 seconds on a 2.33GHz Intel(R) Xeon(R) CPU with 16G of memory. The refinement can be performed on a set of hundreds or thousands of rigid docking solutions and it can be easily parallelized and be run on a cluster of computers, where each process refines a subset of the rigid-docking solution candidates.

Side-Chain Optimization

The side-chain optimization stage assigns the optimal conformation to each interface residue, from a set consisting of common rotamers (backbone dependent) and the unbound conformation. In a *restricted* side-chain optimization (step 1.2 in the SymmRef algorithm) only residues that are in a steric clash are allowed to be moved, and in the *full* optimization (step 1.5 in the SymmRef algorithm) the movements of all interface residues are modeled. SymmRef optimizes the side-chain conformations in the interface of a single monomer and its neighboring monomers. We do not restrict the side-chain conformations to be identical in each monomer, since solved crystal structures of homo-oligomers show that the conformation of side chains often vary between different monomers in a symmetric complex. For example, local differences in side-chain conformations of symmetric complexes can be seen in the crystal structure of Insulin¹⁵, HIV-1 protease² and the Flavivirus envelope glycoprotein (Figure 3A). An extensive analysis of side-chain conformations in symmetric complexes is detailed in the Results section and in Supplementary materials.

The optimization is performed by the integer linear programming approach described by Andrusier et al¹⁶ with a modification in the scoring function that is being minimized.

Minimize:

$$\sum_{(i,r)} E_{ISCO}(i_r) \cdot y_{i_r} + \sum_{(i,r,j,s): i < j, j \in N(i)} E_{ISCO}(i_r, j_s) \cdot x_{i_r j_s}$$

Subject to:

$$\begin{aligned} \sum_r y_{i_r} &= 1, i \in V \\ \sum_r x_{i_r j_s} &= y_{j_s}, i \in V, j \in N(i) \\ y_{i_r}, x_{i_r j_s} &\in \{0, 1\} \end{aligned}$$

Where y_{i_r} and $x_{i_r j_s}$ are decision variables. Setting y_{i_r} to 1 corresponds to choosing rotamer r for residue i , and similarly setting $x_{i_r j_s}$ to 1 corresponds to choosing to include the energy between rotamers r of residue i and rotamer s of residue j , in the minimized energy function.

$E_{ISCO}(i_r j_s)$ and $E_{ISCO}(i_r)$ are the pairwise and self energy terms used for interface side-chain optimization. $E_{ISCO}(i_r j_s)$ is the energy between two rotamers (r and s) of two movable residues (i and j), and $E_{ISCO}(i_r)$ is the energy between rotamer r of residue i and its static environment (backbone and other fixed atoms). These energy terms include repulsive van der Waals energy (E_{rep_vdW}), repulsive electrostatic energy (E_{rep_Elec}) and internal energy score of the chosen rotamers (E_{rot}), as described below:

$$E_{ISCO} = E_{rep_vdW} + E_{rep_Elec} + 6 \cdot E_{rot}$$

For the detailed formulas of these energy terms the reader is referred to Andrusier et al.¹⁶ While in our previous docking refinement methods^{16,17} the side chain optimization minimized a scoring function which was composed only of a repulsive van der Waals energy term and rotamer internal energy score, here we added electrostatics. We found the electrostatics to be particularly important in symmetric complexes since the symmetry often causes a residue to interact with the same residue in the neighboring monomer. This means that positive residues often interact with positive residues and negative residues interact with negative residues. Accounting for electrostatics in the side-chain optimization can resolve these electrostatic repulsive forces by choosing optimal rotamers.

Rigid-Body Monte-Carlo Minimization

Given a symmetric transformation of a docking solution candidate, the rigid-body minimization procedure aims to optimize the van der Waals binding energy of the complex by minimizing the energy in the four-dimensional space of symmetric rigid movements, one translational (the distance from the symmetry axis) and three rotational degrees of freedom. In this stage we perform Monte-Carlo (MC) sampling with local minimizations while restricting all transformations to be symmetric.

In a symmetric complex the interface between any monomer and its neighboring monomers is very similar (not identical since we assume that side-chain conformations may vary). Therefore, we optimize the orientation of a single monomer by minimizing the binding energy of its interaction with both its neighbors. In the case of a symmetric homodimer (C_2 symmetry), we minimize the binding energy of the single interface.

Given a model of a C_k symmetric complex, let M_I be the coordinates of one of the monomers, and let T be the symmetric transformation that transforms the monomer to the orientation of its neighboring monomer in the model. We can calculate the coordinates of the neighboring monomers, M_2 and M_k , by the following formulas:

$$M_2 = T \cdot M_1$$

$$M_k = T^{-1} \cdot M_1$$

Let P be a subtle rigid-body perturbation, sampled randomly from the four rigid-body degrees of freedom discussed above, and let M'_i be the coordinates of monomer i after the perturbation P , as described by the following formula:

$$M'_i = T^{i-1} \cdot P \cdot M_1$$

By applying the P^{-1} transformation on all the monomers M'_i , we will get a new symmetric complex which is slightly different from the original one, and in which the orientation of M_1 remains the same. The new symmetric transformation is $T(P) = P^{-1} \cdot T \cdot P$. It is still a C_k symmetric transformation, for any P , since $T(P)^k = (P^{-1} \cdot T \cdot P)^k = P^{-1} \cdot T^k \cdot P = P^{-1} \cdot P = I$. The rigid body minimization procedure aims to find a local perturbation P^* which minimizes the binding van der Waals energy of M_1 and its two neighbors: $E_{RBM}(T, P) = E_{RBM}(M_1, T(P) \cdot M_1) + E_{RBM}(M_1, T(P)^{-1} \cdot M_1)$. The van der Waals energy between two atoms is defined as the modified Lennard-Jones 6–12 potential with linear short-range repulsive score, as detailed in Mashiach et al¹⁷. The energy calculation accounts for different side-chain conformations of the neighboring monomers, as could be predicted by a previous run of the side-chain optimization procedure.

The rigid-body optimization procedure starts with a given C_k symmetric transformation T , and performs the following repeated steps:

1. Generate a random perturbation P .
2. Local minimization of $E_{RBM}(T, P)$ (in 4 degrees of freedom), by the BFGS quasi-Newton algorithm^{18,19}.
3. The obtained position is accepted or rejected by the Metropolis criterion: If the new position results in a lower energy score, the move is unconditionally accepted. Otherwise, it is accepted with some probability. If the position is accepted: $T = P^{-1} \cdot T \cdot P$
4. Go to 1.

The rigid-body MC Minimization procedure is performed twice in the SymmRef algorithm. At the beginning of the refinement (step 1.1 in the algorithm) a short MC minimization, with 10 iterations, is performed. The goal of this step is to remove major steric clashes that may occur in the initial docking model that is being refined. Later in the algorithm, after refining the side-chains and backbone conformations, a second minimization of the rigid-body orientation is performed (step 1.4). In the second MC minimization 50 iterations are performed. The goal here is to optimize the packing of monomers in the complex. 50 iterations were found to be adequate for finding a satisfactory local minimum in the energy landscape.

Backbone Refinement

Backbone refinement is performed by a normal mode based approach which was developed for general protein-protein docking refinement in the FiberDock method²⁰. Normal modes are a set of predicted movements that a protein is likely to undergo. The conformational change that a protein undergoes upon binding (from the unbound conformation to the bound conformation) can be described as a linear combination of normal modes. In the backbone refinement procedure of SymmRef we attempt to predict the coefficients of this linear

combination. We adjusted the backbone refinement algorithm of FiberDock to restrict the backbone conformations of all the monomers in the complex to be identical, in order to preserve the symmetry of the complex. This was done by applying the same linear combination of normal modes on all the monomers simultaneously during the refinement. The algorithm is briefly described below. For more details the reader is directed to the FiberDock paper²⁰.

In a pre-processing stage, the normal modes of the unbound structure of the monomer are calculated by using the anisotropic network model (ANM)²¹. During the backbone refinement procedure, the following steps are performed:

1. Rigid-body minimization by the BFGS quasi-Newton algorithm^{18,19}.
2. Repeat the following steps until the energy score converges:
 - 2.1 Calculate the van der Waals forces between two neighboring monomers.
 - 2.2 Identify the 10 normal modes with the best correlation to the calculated vdW forces and minimize the backbone conformation of the monomers along them and along the rigid-body degrees of freedom,
 - 2.3 Fast rigid-body Monte-Carlo minimization.

Step 2 is repeated until the energy score converges to a local minimum (5 iterations without energy improvement). In order to shorten the running time of the algorithm we restrict the number of iterations to be below 20, a sufficient number for most of the cases for producing satisfactory results. Additionally, we stop the backbone refinement procedure when all the steric clashes in the complex are solved, i.e. if the repulsive vdW energy is below a threshold.

In step 2.3 we perform 10 iterations of rigid-body MC minimization in order to optimize the packing of the monomers in the complex after the backbone movement.

Ranking

The ranking of the refined docking solutions is performed by an energy scoring function. We used the default energy score from our previous refinement methods, FireDock^{16,20} and FiberDock²⁰. The weights of this energy score were optimized by using a machine learning algorithm and heuristic optimization on a test set of asymmetric protein-protein complexes from a docking benchmark²², as described in the FireDock paper¹⁶. The performance of this energy score was shown to be successful on a docking benchmark and in the CAPRI experiments²³. Kastiris et al have recently shown that this energy score has the highest correlation to experimental binding affinities compared to the energy scores of other docking methods²⁴. However, this energy score is still not sufficiently accurate for predicting binding affinities.

We slightly adapted the weights of the energy terms to symmetric complexes. We could not use a machine learning algorithm to optimize the weights for symmetric multimers due to the small number of cases of symmetric complexes with solved unbound structures. Consequently, this was done manually based on specific properties of symmetric complexes as described below. Symmetric complexes often include charged residues which interact with the same residue in neighboring monomers. A slightly inaccurate modeling of the interface side-chain conformations may cause a high calculated repulsive electrostatic energy. Therefore, in order to improve the robustness of the energy score we penalize less for electrostatic repulsion. We have reduced the short range repulsive electrostatics weight from 0.21 to 0.15 and the long-range repulsive electrostatics weight from 0.69 to 0.1. These

values were manually adjusted by a process of trial and error. In addition, we removed the heuristic *insideness* energy term which is not relevant for symmetric complexes. The other weights were unchanged. The full energy score is detailed in the following formula:

$$E_{\text{symm}} = 1.5E_{\text{attrVdW}} + 0.8E_{\text{repVdW}} + 1.6E_{\text{ACE}} + 0.21E_{\text{s}_{\text{attr El}}} + 0.15E_{\text{s}_{\text{rep El}}} + 0.1E_{\text{t}_{\text{rep El}}} + 1.2E_{\text{HB}} + E_{\text{pypi}} + 0.7E_{\text{catpi}} + 2.5E_{\text{aliph}} + 0.05E_{\text{deform}}$$

The adjustment of the energy scoring function was tested on an independent dataset of arbitrarily selected bound docking cases in order to rule out over-fitting (see Results section).

Dataset

In order to test the performance of our refinement method we created a dataset of 16 unbound docking cases. There are very few cases in the Protein Data Bank in which the three dimensional structure of a monomer was solved both in a symmetric complex and alone in the unbound conformation. We identified 7 such cases and included them in our dataset. In addition we added 9 other cases in which the unbound structure of the monomer was not experimentally solved but a structure of a homologue protein was available. In these cases we modeled the structure of the monomer according to the homologue and used this model as the unbound structure in our unbound docking experiments. The homology modeling was performed by the MODELLER²⁵ method using sequence alignment that was generated by STACCATO²⁶. We trimmed edges and loops that had no template according to the sequence alignment. The details of the dataset are in Table II.

Some of the symmetric complexes in our dataset have dihedral symmetry. The current version of SymmRef can handle only cyclic symmetry. Therefore, in these cases we aimed to predict the structure of the dimers that compose the full dihedral complexes.

Our dataset includes two targets of the CAPRI experiment, target 10 (1URZ)²⁷ and 42 (2WQH)²⁸. The structure of the TBEV envelope protein of the tick-borne encephalitis virus (1URZ) includes a flexible C-terminal domain (from residue 299). In the CAPRI challenge, the predictors were advised to ignore this domain; therefore we removed it from the structure in our dataset as well.

The trimeric form of bovine pancreatic RNase (1JS0) reveals a domain-swap of a beta strand. This conformational change increases drastically the binding energy of the complex. However, we do not expect SymmRef, or any other existing docking method, to predict this conformational change without prior knowledge of the domain swap. Therefore we ignored the residues of this beta strand in the RMSD calculations.

Docking Evaluation

Following the CAPRI challenge evaluation protocol²⁷ we used several evaluation criteria for assessing the performance of docking methods, as detailed below. Let S be the native structure of a symmetric complex and let M be a predicted model of the same complex which we would like to evaluate. Let S_1 and S_2 be two adjacent monomers in the native structure and let M_1 and M_2 be the corresponding monomers in the evaluated model. Let $I(S_1, S_2)$ be the amino acids of the interface between the adjacent monomers S_1 and S_2 . An amino acid is considered to be in the interface if it has at least one atom within 10Å of the neighboring monomer.

RMSD – The root mean square deviation between the backbone atoms of S_2 and M_2 after superimposing the monomers S_1 and M_1 .

IRMSD – The root mean square deviation between the backbone atoms of I(S₁, S₂) and I(M₁, M₂) after superimposition of the two interfaces.

F_{nat} – The fraction of native contacts. A pair of residues from two adjacent monomers is considered to be in contact if any of their atoms are within 5 Å.

We classified the accuracy of the predicted models into four categories according to the CAPRI criteria²⁷: **(1)** Incorrect. **(2)** Acceptable accuracy (marked by one star – *). **(3)** Medium accuracy (**). **(4)** High accuracy (***).

Results

Analysis of backbone and side-chain conformations of symmetric complexes

Previous docking methods^{11,12} assume that the conformations of monomers in a symmetric complex are identical. However, most of the crystal structures of symmetric protein complexes show local differences in side-chain conformations and occasional minor differences in backbone conformations. We analyzed the backbone and side chain conformations of 65 crystal structures of symmetric complexes (Table S1).

This analysis shows that the backbone conformation in the interface of symmetric complexes is almost identical in different monomers of a symmetric complex. In 82% of the cases the C α interface RMSD was below 0.5 Å, and in 95% of the cases it was below 1 Å (Figure 1). Due to these findings we decided to use symmetry constraints on the backbone conformations of the monomers during the docking refinement process.

The analysis of the side-chain conformations showed that in 44 (68%) of the cases there was a variation in the side-chain conformation between two arbitrary monomers of the symmetric complex. Two side-chain conformations were considered to be different if the atomic RMSD exceeds 1 Å, after superimposing the two residues based on their backbone atoms.

There are few factors that can break the symmetry of side-chain conformations in a symmetric complex. The first is self-interaction of an amino-acid with itself in an adjacent monomer. We define a self-interaction to be a case in which two corresponding residues in adjacent monomers have atoms within a distance of 4 Å. For example, in the case of the Flavivirus envelope glycoprotein (Figure 3A), Arg217 has a self-interaction in the complex. One of the Arg217 residues (in chain A) adopts a conformation that enables it to form a favorable hydrogen-bond with a water molecule placed inside the complex. The other Arginines cannot adopt the same conformation without forming a steric clash with the first Arginine. In order to examine the impact of the self-interaction property on the symmetry of the side-chain conformations we compared the percentage of complexes with identical side-chain conformations in a group of symmetric complexes with and without self-interactions. Our dataset of 65 crystal structures of symmetric complexes included 28 cases without self-interactions. In this group 12 (43%) cases had identical side-chain conformations between two arbitrary monomers. In contrast, in the group of 37 structures with self-interaction, only 9 (24%) cases had identical side-chain conformations (Figure 2). Other factors that may break the symmetry of the side-chain conformations include side-chain flexibility on the surface of the protein due to interactions with the surrounding water molecules, backbone flexibility in flexible loops, asymmetric interactions with small ligands and interaction of a symmetric complex with another protein.

Due to these results we decided not to restrict the side-chain conformations of all the monomers to be identical during the side-chain optimization procedure. Figure 3B shows the

results of the side-chain optimization procedure on the Flavivirus envelope glycoprotein. We started the optimization by duplicating one of the monomers (chain A) and superimposing it on the other monomers in the complex. This initial complex had a high binding energy score of 782.27, mostly due to high repulsive vdW energy of 1045.99 and high repulsive electrostatic energy of 223.23. The figure focuses on a steric clash between Arginine 217 from two monomers in the complex. The side-chain optimization method rearranged the side-chain conformations and resulted in a model with a much lower energy score of -72.48 (repulsive vdW energy of 0.07 and repulsive electrostatic energy of 43.23). In this experiment we did not perform rigid-body minimization or backbone refinement.

The importance of symmetry constraints in docking refinement of symmetric multimers

Protein complexes with C_n symmetry have strict constraints on the rigid-body transformation between adjacent monomers. The transformation must be around a symmetry axis and the rotation angle must be exactly $360^\circ/n$. This constraint restricts the search space of the refinement algorithm and therefore increases the chance of finding a highly accurate solution. In order to demonstrate the importance of using these constraints we used SymmDock to predict the hexameric capsomer of HIV-1 (PDB ID: 3MGE) given a monomer in the bound conformation. Then, we refined the first acceptable solution of SymmDock (which was ranked in the 4th place) by SymmRef and by FiberDock. FiberDock²⁰ is a general docking refinement method that uses the same minimization techniques as SymmRef but with no symmetry constraints. The original SymmDock solution was highly accurate, with RMSD of 1.70Å and IRMSD of 0.69Å. SymmRef improved the accuracy of the docking solution to RMSD of 0.21 Å and IRMSD of 0.09Å. FiberDock refinement resulted in a less accurate solution with RMSD of 2.89Å and IRMSD of 1.03Å.

Both refinement methods minimize the rigid-body transformation between two adjacent monomers. The refined transformation can be used for building the full hexameric complex by applying it five times on the structure of the first monomer. Since FiberDock does not restrict the final transformation to be symmetric, using it to build the complete complex often results in an open complex, where the last monomer does not interact with the first monomer, or in a severe clash between the last monomer and the first one. In the case of the hexameric capsomer of HIV-1 the refinement by FiberDock resulted in an open complex (see Figure S1 in the supplementary material).

Next, we used FiberDock to refine the top 1000 docking solutions of SymmDock, given the bound structure of one of the monomers (chosen arbitrarily) for each case in our dataset. In cases where SymmDock generated less than 1000 solutions, we refined all of them. The results show that FiberDock improved SymmDock results in only half of the cases (see Table S2 in the supplementary material). SymmRef however, improved both the ranking and the accuracy of SymmDock results in 12 out of the 16 cases, and overall produce significantly better results than FiberDock. SymmRef results for bound docking are described in details in the next section and in Table III.

Bound Docking Experiments

In order to test the performance of SymmRef we first used it to refine the top 1000 docking solutions of SymmDock⁷, given the bound structure of one of the monomers (chosen arbitrarily). In cases where SymmDock generated less than 1000 solutions, we refined all of them. In this experiment we used SymmRef without the backbone refinement procedure (step 1.3 in the algorithm described in the Methods section), and then compared the performance of the algorithm with and without the restriction of identical side-chain conformations assigned to each monomer.

The results are presented in Table III. In 12 out of the 16 cases, SymmRef significantly improved both the ranking and the accuracy of the SymmDock results, and ranked a medium or a high quality model in the 1st or 2nd place.

In many of the cases, the first acceptable solution of SymmRef originated from a SymmDock solution which was poorly ranked, even though near native results existed in the top 10 results of SymmDock. For example, in the case of 3MGE the first acceptable solution of SymmDock was ranked in the 4th place with RMSD of 1.70Å and IRMSD of 0.69Å. Refinement and re-ranking of this solution resulted in a high accuracy solution with RMSD of 0.21Å and IRMSD of 0.09Å, which was still ranked in the 4th place by SymmRef. The top solution by SymmRef had a similar accuracy but a slightly better energy and it was created by refining a less accurate SymmDock solution, with RMSD of 8.84Å which was ranked 691. This and other results demonstrate that during the refinement by SymmRef, solutions with RMSD in the range of 0Å – 10Å converge to a highly accurate solution with low energy score. Only in rare cases in which the initial model before the refinement is relatively inaccurate, the accuracy of the refined model might get slightly worse (e.g. 1B0C, 1BK6, 1D01).

In the cases of 1B0C and 1DD1, the top near native results were ranked 188 and 597 (respectively) by SymmRef, which is a poor ranking for bound docking experiments. In these cases the size of the interface of the native structure are very small, with Δ ASA (Accessible Surface Area) of 238.4Å² for 1B0C and 601Å² for 1D01 as calculated by the PROTORP server³⁰. The standard average interface surface³¹ is approximately 1000Å² and complexes with Δ ASA < 1400Å² are considered to be very difficult for docking³². Additionally, in solution, the 1B0C complex actually forms a decamer (dimer of pentamers)³³. The interaction between two pentamers stabilizes the binding of the monomers within a pentamer. We believe that this could be the reason for the relatively poor ranking and high binding energy of the first acceptable model that SymmRef predicted for this pentamer. SymmDock ranked in the 7th place an acceptable solution for the 1B0C case, according to the CAPRI criteria. However, detailed examination of this solution showed that the solution is quite far from the native structure, with RMSD of 15.13Å. This solution was considered to be acceptable due to the IRMSD which is slightly below 4Å. However, since the interface in this case is very small, the IRMSD by itself is not a sufficient measure for the quality of the model. SymmRef ranked a more accurate solution with RMSD of 11.61Å in the 10th place but this solution is not considered to be acceptable according to CAPRI criteria and thus it does not appear in Table III.

Evaluation of the use of side-chains symmetry constraints—A comparison of the performance of SymmRef with and without symmetry constraints on the side-chain conformations shows that in specific cases these constraints prevent the method from creating high accuracy results with low energy score. This is shown most significantly in the case of 1URZ in which a comparison between monomers in the crystal structure of the complex reveals interface side-chains with different conformations. In this case, SymmRef without symmetric side-chains restriction ranked 4 high accuracy and two medium accuracy results in the top 10 solutions, while the refinement with this symmetric restriction didn't generate a high accuracy result at any rank (among all the 1000 results), and only one medium accuracy model was generated, but it was ranked in place 771. In addition to the 1URZ case, an improvement in the results of SymmRef without restricting the side-chain conformations can also be seen in cases 3D57, 1A3F, 3MGE and 1AP9. However, the restriction of the side-chain conformation slightly improved the results in cases 2WQH and 1BK6.

Evaluation of the energy scoring function adjustments—In this study we slightly adjusted the weights of the electrostatics terms of the energy scoring function that was used in our previously developed FireDock and FiberDock methods, as described in Methods section. These values were manually adjusted by a process of trial and error on a subset of our dataset which includes the first 11 cases in Table II (marked with stars). For these 11 cases the adjustment of the weights improved the ranking in 2 cases (1D2N and 1A3F). In the rest of the cases the ranking have hardly changed (see Table S3 in the supplementary material). We also noticed that there was a greater improvement in the ranking when the constraint of symmetric side-chain conformation was enforced. As shown before, this constraint causes SymmRef to generate less accurate models. For these models the adjusted scoring function was able to drastically improve the ranking, compared to the original scoring function. The adjusted scoring function reduces the influence of the repulsive electrostatic forces. In symmetric complexes two residues with the same charge often interact with each other. Therefore, inaccurate modeling of their conformation may cause a high calculated repulsive electrostatics energy. Reducing the weights of the repulsive electrostatics terms improves the robustness of the scoring function, which is especially important for unbound docking cases.

In order to rule out over fitting to the data we examined the affect of this adjustment on the other 5 cases in our dataset and on additional 10 independent cases of symmetric complexes. For all these cases we ran SymmDock, using the bound conformation of one of the monomers, and then we refined and re-ranked the solutions by the original and the adjusted energy scoring function. The results are detailed in Table S4 in the supplementary material. In 3 out of the 21 test cases (1A3F, 1EUA and 1RRE) the ranking has improved when the adjusted scoring function was used, and in the other cases the ranking remained almost identical.

Unbound Docking Experiments

In an unbound experiment, we docked the unbound monomers by SymmDock and refined the top 1000 results by SymmRef. The results (Table IV) show the drastic improvement in the accuracy and ranking of the docking solutions. SymmDock ranked a near native solution in the top 10 solutions in 4 out of the 16 cases. The refinement and rescoring by SymmRef increased this number to 7, and in 4 of these cases the near native solution was ranked in the 1st place. In the case of 2WQH, for example, the first acceptable solution of SymmDock was ranked 275. After running SymmRef, a high accuracy model (according to CAPRI criteria) was ranked in the 1st place, three other high accuracy solutions were also ranked in the top 10 solutions. Figure 4 shows the structures of some of SymmDock solutions, before and after SymmRef refinement.

Comparison of SymmRef to the refinement and rescoring by RosettaDock

The RosettaDock method can be used for global and local docking, and also for refining and re-ranking of given symmetric docking solutions. The method uses Monte-Carlo minimization to optimize the relative orientation and the conformation of the monomers. We used the symmetric docking protocol of RosettaDock for refining and re-ranking SymmDock unbound docking solutions and compared the results with SymmRef. The currently available download version of Rosetta does not allow backbone refinement during symmetric docking. Therefore, only side-chain flexibility was modeled by the RosettaDock refinement in this experiment. For each one of the top 1000 solutions of SymmDock, we performed 10 local refinements by the symmetric RosettaDock method and selected the lowest energy model. This is similar to the SymmRef method which also performs the refinement 10 times and chooses the solution with the best score (see the Methods section).

Finally we ranked the 1000 refined solutions by the RosettaDock energy score and compared the results to the results of SymmRef.

Both refinement methods significantly improved SymmDock results. However, the refinement by SymmRef produced better results on the tested benchmark. In 5 cases (1URZ, 1VIM, 2WQH, 1BK6, 1F1G) SymmRef ranked a near native solution in the top 10 results, while RosettaDock failed to do so. 1URZ, 1VIM and 1BK6 are cases with relatively large backbone conformational change, with IRMSD of 4.97Å, 3.53Å, and 2.18Å between their bound and unbound structure. This can explain the superior results of SymmRef which models backbone movements. On the other hand, in two other cases (3D57, 1A3F) RosettaDock ranked near native solutions in the top 10 results, while SymmRef failed to do so.

Another important difference between the performances of the methods is the running times. The average running time of a single refinement by SymmRef is 23 seconds on a 2.33GHz Intel(R) Xeon(R) CPU with 16G of memory. The average running time of a single refinement by the symmetric RosettaDock is 6 minutes.

Comparison to other symmetric docking methods

Our full symmetric docking protocol uses the SymmDock rigid symmetric docking procedure for generating a set of 1000 models and then uses SymmRef to refine and re-score these models. The performance of this docking protocol on the dataset of unbound structures was promising. In 7 out of the 16 cases SymmRef produced a near native model in the top 10 solutions, in 4 of these cases the first near-native solution was ranked in the 1st place. In one other case a near native model was ranked in the top 20 solutions. We compared this protocol to other state-of-the-art methods of symmetric docking: HADDOCK¹² and M-ZDOCK¹⁰. The results are presented in Table V.

The HADDOCK method requires information about the location of the interface to guide the docking. For that we used the CPORT web-server, which integrates the results of five interface prediction methods^{34–38}. The CPORT web-server was previously used¹² for predicting the interface of symmetric complexes prior to using HADDOCK and we used it in the same manner. However, one should note that HADDOCK is mostly used in cases where some biological information on the approximate location of the binding site is available. Since we wanted to compare between purely computational methods, we used the CPORT computational method for predicting the binding site. SymmDock can also use information on the interface location (not shown in the results). Given a set of residues that are suspected to be in the interface, SymmDock will only generate solutions in which at least one of these residues is located in the interface.

The HADDOCK webserver cannot predict hexamers, therefore we tested it on 14 out of the 16 cases in the dataset. Additionally, we analyzed only the best structure in the top 10 clusters (which are given as output by the webserver).

HADDOCK's performance is comparable to our protocol. However, it greatly depends on the interface prediction. In 3 of the cases more than 50% of the residues predicted by CPORT were indeed in the interface. In all these cases HADDOCK predicted a near-native docking model in its top 10 clusters. In 11 cases, less than 50% of the residues predicted by CPORT were correct, and in only 4 of these cases HADDOCK predicted a near native docking model in the top 10 clusters.

The M-ZDOCK method produced a near-native model in the top 10 solutions in 4 out of the 16 cases. In 3 cases (1URZ, 1VIM and 2WQH) in which M-ZDOCK did not produce a near-

native result in the top 50 solutions, our docking protocol ranked a near-native model in the top 10 solutions, in two of these cases in the 1st place. On the other hand, in the case of 1D2N, M-ZDOCK produced an acceptable solution which was ranked in the 2nd place, while SymmDock did not generate an acceptable solution in the top 1000 solutions, and therefore no near-native solutions were refined by SymmRef.

M-ZDOCK is a rigid-body symmetric docking method, which is based on the FFT approach. Similar to SymmDock it does not model any side-chain or backbone movements, but allows a certain amount of steric clashes (soft-docking). The scoring function of M-ZDOCK is based on surface complementarity, electrostatics and desolvation. However, since the structure of the monomers is not refined the values of the energy terms are less accurate. SymmRef models induced-fit conformational changes and calculates a more accurate energy score, and therefore achieved better results in most of the cases in our benchmark.

The SymmRef algorithm can refine rigid-docking solutions that are generated by any symmetric docking method. Therefore we tested its performance of refining and re-ranking the M-ZDOCK solutions. The results are presented in Table VI. The results show that SymmRef improves the accuracy and ranking of the solutions of M-ZDOCK. In 7 cases the number and quality of the near-native models in the top 10 solutions improved after SymmRef was used. In the case of 2WQH, for example, while M-ZDOCK did not rank any acceptable model in the top 100 solutions, SymmRef ranked 8 high accuracy models in the top 10 solutions.

Summary

The majority of protein complexes in the living cell are composed of several proteins that simultaneously interact with each other^{39,40}. Most of them are symmetric multimers². In this paper we described SymmRef, a novel method for symmetric docking refinement. The method refines and re-ranks symmetric docking solutions that can be generated by any global symmetric docking method. SymmRef models backbone and side-chain conformational changes and performs rigid-body minimization. The method is based on our previously developed protein-protein refinement method, FiberDock^{20,41}. It finds the optimal rotamers for the interface residues by an Integer Linear Programming approach, and refines the backbone conformation by normal modes minimization. SymmRef is available for download at <http://bioinfo3d.cs.tau.ac.il/SymmRef/download.html>.

Previous symmetric docking methods restrict the monomers to be identical. However, most of the solved crystal structures of symmetric homo-oligomers show local differences in side-chain conformations. We analyzed this phenomenon and discovered that self-interaction of an amino-acid in a symmetric complex tends to break the symmetry at the level of side-chain conformations. Therefore, we concluded that the side-chain optimization procedure should not be restricted to select symmetric rotamers. In cases without self-interactions there is a better chance for the side-chain conformations to be identical in each monomer. However, in these cases the interface between two monomers contains different residues in each side. Therefore restricting the side-chain optimization procedure in these cases, to select symmetric rotamers, will not reduce the search space.

We tested the performance of SymmRef on unbound docking cases by refining the symmetric rigid-docking solutions of SymmDock^{8,10} and M-ZDOCK¹⁰. The results show that in both cases SymmRef improves both the accuracy and the ranking of the symmetric rigid-docking solutions with RMSD in the range of 0Å – 10Å. Comparison of SymmRef with the refinement and re-scoring performance of RosettaDock showed that SymmRef produces better results on the tested benchmark. Moreover, the running time of SymmRef is

much faster than the running time of RosettaDock (by a factor of 15) and therefore SymmRef is able to perform a significantly larger number of refinement iterations with the same computational resources. Alternatively, the great efficiency of SymmRef can enable the user to refine and re-score significantly more rigid-docking solutions.

We also compared the performance of M-ZDOCK and HADDOCK with our full symmetric docking protocol of refining and re-ranking the top 1000 solutions of SymmDock by SymmRef. Our experiments on the unbound cases dataset show that our protocol outperforms M-ZDOCK and is comparable to HADDOCK. Our docking protocol produced a near native model in the top three ranked solution in 6 out of the 16 cases. M-ZDOCK succeeded to rank a near native model in the top three ranked solution in 4 cases and HADDOCK in 5 cases. When comparing the number of cases in which each docking method ranked a near native model in the top 10 solutions, our protocol succeeded in 7 cases, M-ZDOCK in 4 cases and HADDOCK in 7 cases.

Up until now, the docking community mainly concentrated on predicting the structure of binary protein-protein complexes. Studies have shown that the majority of protein complexes are composed of more than two proteins, and about half of all complexes contain more than five proteins⁴⁰. Therefore, in order to model real-life complexes we have to be able to predict the structure of multi-molecular complexes. Predicting the structure of multi-molecular symmetric complexes is only the first step toward this direction. Only a few methods were developed so far for predicting hetero-multi-molecular complexes^{12,42–44}. In the future we plan to adapt our docking refinement technique for hetero-multi-molecular complexes. The refinement method could be used for improving the performance of current docking methods and for improving the accuracy of rigid fitting of protein structures into electron-microscopy density maps.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

E.M is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities and carried out her research in partial fulfillment of the requirements for the Ph.D. degree at Tel Aviv University. The research of HJW has been supported in part by the Israel Science Foundation (grant no. 1403/09) and the Hermann Minkowski-Minerva Center for Geometry at Tel Aviv University. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

References

1. Plaxco KW, Gross M. Protein complexes: the evolution of symmetry. *Curr Biol.* 2009; 19(1):R25–26. [PubMed: 19138586]
2. Goodsell DS, Olson AJ. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct.* 2000; 29:105–153. [PubMed: 10940245]
3. Monod J, Wyman J, Changeux JP. On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol.* 1965; 12:88–118. [PubMed: 14343300]
4. Andre I, Strauss CE, Kaplan DB, Bradley P, Baker D. Emergence of symmetry in homooligomeric biological assemblies. *Proc Natl Acad Sci U S A.* 2008; 105(42):16148–16152. [PubMed: 18849473]

5. Berchanski A, Segal D, Eisenstein M. Modeling oligomers with Cn or Dn symmetry: application to CAPRI target 10. *Proteins*. 2005; 60(2):202–206. [PubMed: 15981250]
6. Berchanski A, Eisenstein M. Construction of molecular assemblies via docking: modeling of tetramers with D2 symmetry. *Proteins*. 2003; 53(4):817–829. [PubMed: 14635124]
7. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. Geometry-based flexible and symmetric protein docking. *Proteins*. 2005; 60(2):224–231. [PubMed: 15981269]
8. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*. 2005; 33(Web Server issue):W363–367. [PubMed: 15980490]
9. Comeau SR, Camacho CJ. Predicting oligomeric assemblies: N-mers a primer. *J Struct Biol*. 2005; 150(3):233–244. [PubMed: 15890272]
10. Pierce B, Tong W, Weng Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. *Bioinformatics*. 2005; 21(8):1472–1478. [PubMed: 15613396]
11. Andre I, Bradley P, Wang C, Baker D. Prediction of the structure of symmetrical protein assemblies. *Proc Natl Acad Sci U S A*. 2007; 104(45):17656–17661. [PubMed: 17978193]
12. Karaca E, Melquiond AS, de Vries SJ, Kastiris PL, Bonvin AM. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Mol Cell Proteomics*. 2010; 9(8):1784–1794. [PubMed: 20305088]
13. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A*. 1992; 89(6):2195–2199. [PubMed: 1549581]
14. Mandell JG, Roberts VA, Pique ME, Kotlovsky V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF. Protein docking using continuum electrostatics and geometric fit. *Protein Eng*. 2001; 14(2):105–113. [PubMed: 11297668]
15. Adams MJ, Blundell TL, Dodson EJ, Dodson GG, Vijayan M, Baker EN, Harding MM, Hodgkin DC, Rimmer B, Sheat S. Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature*. 1969; 224(5218):491.
16. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. *Proteins*. 2007; 69(1):139–159. [PubMed: 17598144]
17. Mashiach E, Nussinov R, Wolfson HJ. FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins*. 2010; 78(6):1503–1519. [PubMed: 20077569]
18. Zhou S. Rapidly convergent procedure to solve the density profile equation in the classical density functional theory. *J Comput Chem*. 2006; 27(8):941–947. [PubMed: 16586529]
19. Fletcher R. A new approach to variable metric algorithms. *Comput J*. 1970; 13:317–322.
20. Mashiach E, Nussinov R, Wolfson HJ. FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins*. 2009
21. Hinsen K. Analysis of domain motions by approximate normal mode calculations. *Proteins*. 1998; 33(3):417–429. [PubMed: 9829700]
22. Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J, Weng Z. Protein-Protein Docking Benchmark 2.0: an update. *Proteins*. 2005; 60(2):214–216. [PubMed: 15981264]
23. Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An integrated suite of fast docking algorithms. *Proteins*. 2010
24. Kastiris PL, Bonvin AM. Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J Proteome Res*. 2010; 9(5):2216–2225. [PubMed: 20329755]
25. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000; 29:291–325. [PubMed: 10940251]
26. Shatsky M, Nussinov R, Wolfson HJ. Optimization of multiple-sequence alignment based on multiple-structure alignment. *Proteins*. 2006; 62(1):209–217. [PubMed: 16294339]
27. Mendez R, Lepplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. *Proteins*. 2005; 60(2):150–169. [PubMed: 15981261]

28. Janin J. The targets of CAPRI Rounds 13–19. *Proteins*. 2010; 78(15):3067–3072. [PubMed: 20589643]
29. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25(13):1605–1612. [PubMed: 15264254]
30. Reynolds C, Damerell D, Jones S. ProtorP: a protein-protein interaction analysis server. *Bioinformatics*. 2009; 25(3):413–414. [PubMed: 19001476]
31. Bahadur RP, Zacharias M. The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cell Mol Life Sci*. 2008; 65(7–8):1059–1072. [PubMed: 18080088]
32. Vajda S. Classification of protein complexes based on docking difficulty. *Proteins*. 2005; 60(2): 176–180. [PubMed: 15981248]
33. Hamiaux C, Perez J, Prange T, Veesler S, Ries-Kautt M, Vachette P. The BPTI decamer observed in acidic pH crystal forms pre-exists as a stable species in solution. *J Mol Biol*. 2000; 297(3):697–712. [PubMed: 10731422]
34. Kufareva I, Budagyan L, Raush E, Totrov M, Abagyan R. PIER: protein interface recognition for structural proteomics. *Proteins*. 2007; 67(2):400–417. [PubMed: 17299750]
35. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol*. 2004; 338(1):181–199. [PubMed: 15050833]
36. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins*. 2007; 66(3):630–645. [PubMed: 17152079]
37. Liang S, Zhang C, Liu S, Zhou Y. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res*. 2006; 34(13):3698–3707. [PubMed: 16893954]
38. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. *Proteins*. 2005; 61(1):21–35. [PubMed: 16080151]
39. Abbott A. The society of proteins. *Nature*. 2002; 417(6892):894–896. [PubMed: 12087375]
40. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*. 2002; 415(6868):141–147. [PubMed: 11805826]
41. Mashiach E, Nussinov R, Wolfson HJ. FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking. *Nucleic Acids Res*. 2010; 38 (Suppl):W457–461. [PubMed: 20460459]
42. Lasker K, Sali A, Wolfson HJ. Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. *Proteins*. 2010; 78(15):3205–3211. [PubMed: 20827723]
43. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ. Protein structure prediction via combinatorial assembly of sub-structural units. *Bioinformatics*. 2003; 19 (Suppl 1):i158–168. [PubMed: 12855452]
44. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ. Prediction of multimolecular assemblies by multiple docking. *J Mol Biol*. 2005; 349(2):435–447. [PubMed: 15890207]

Figure 1.

The distribution of interface C α RMSD (IRMSD) between two arbitrary monomers in symmetric protein complexes. The horizontal axis shows the IRMSD intervals in angstroms and the vertical axis shows the number of complexes with each range of IRMSD. In 95% of the cases the IRMSD is below 1Å.

Figure 2.

The percentage of cases of symmetric protein complexes with and without conformational changes in the interface side-chains. In cases with self interactions of an interface amino acid, there is a higher chance for different side-chain conformations.

Figure 3. Local differences in side-chain conformations of the monomers of the Flavivirus envelope glycoprotein (PDB-ID: 1URZ). **(A)** Superimposition of the three monomers. The interface residues are shown in sticks presentation. **(B)** The interface between two monomers. The native residue conformations are in red, a duplication of the rotamers of one of the monomers (chain A) is in blue and the refined conformations by the side-chain optimization of SymmRef is in green. In the square, one can see that enforcing side-chain conformations to be identical in all the monomers results in steric clashes and high repulsive electrostatic energy (the squared area show a steric clash between Arg217 in monomers A and B). By allowing local differences between the monomers, SymmRef was able to find an alternative rotamer (similar to the native one) that resolved the clash. This image was produced using the UCSF Chimera package.²⁹

SymmDock solution Refinement by SymmRef

Figure 4.

Comparison between SymmDock solutions before and after refinement by SymmRef. The figure shows superimposition of the docking models on the native complex structure, presented in gray. The superimposition minimizes the RMSD of one of the monomers of the model and the native complex. The structure of the second monomer is shown in red for the SymmDock solution and in blue for the refined solution by SymmRef. **(A)** The docking solutions of 1VIM. On the left a SymmDock solution with RMSD of 8.97Å which was ranked 801. On the right, the refinement of this solution by SymmRef, with RMSD of 4.17Å which was ranked in the 1st place. **(B)** The docking solution of 2WQH. SymmDock solution with RMSD of 7.28Å and rank 455. SymmRef refined model with RMSD of 2.32Å and rank 1. **(C)** The docking solution of 1TR0. SymmDock solution with RMSD of 4.05Å and rank 96. SymmRef refined model with RMSD of 1.67Å and rank 1. This image was produced using the UCSF Chimera package.²⁹

Table I

Existing docking methods for symmetric multimers.

Method	Symmetry type	Flexibility	Reference
M-ZDOCK	Cyclic	None	Ref. 10
SymmDock	Cyclic	None	Ref. 7,8
MolFit	Cyclic, dihedral	None	Ref. 5,6
ClusPro	Cyclic, dihedral	None	Ref. 9
ROSETTA	Cyclic, dihedral, helical, icosahedral	Full interface flexibility	Ref. 11
HADDOCK	Oligomers with arbitrary symmetry	Full interface flexibility	Ref. 12

Table II

The dataset used for evaluating SymmRef performance.

System	PDB ^(a)	Unbound ^(b)	RMSD ^(c)	IRMSD ^(d)	C _n /D _n ^(e)
Flavivirus envelope glycoprotein *	1URZ	1SVB	4.33	4.97	C ₃
Hypothetical protein *	1VIM	1VIV ^(HB)	3.10	3.53	D ₂
HIV gp41 core *	1F23	3CP1 ^(HB)	0.63	0.63	C ₃
TR-beta variant *	3D57	1NAX ^(HU)	0.88	0.78	C ₂
A boiling stable protein SP1 *	1TR0	1Q4R ^(HB)	1.00	1.05	D ₆
N-ethylmaleimide-sensitive fusion protein *	1D2N	1IY2 ^(HU)	5.29	5.35	C ₆
Domain-swapped RNase *	1JS0	9RAT	6.97	8.69	C ₃
Half of BPTI decamer *	1B0C	3PTI	0.51	0.54	C ₅
Phospholipase A2 *	1A3F	1POA	0.77	0.82	C ₃
Hexameric Capsomer of HIV-1 *	3MGE	1E6J	10.05	10.12	C ₆
Self-association of TPR domains *	2WQH	1NA0	0.86	0.16	C ₂
Bacteriorhodopsin	1AP9	3HAR	1.66	1.39	C ₃
Karyopherin α	1BK6	1IAL ^(HU)	2.81	2.18	C ₂
TNF Receptor Associated Factor 2	1D01	1LB4 ^(HU)	2.68	1.23	C ₃
Smad4 Active Fragment	1DD1	1MJS ^(HU)	5.14	2.59	C ₃
Yeast cuznsod exposed to nitric oxide	1F1G	1MFM ^(HU)	1.15	1.67	C ₂

^(a)The PDB ID of the symmetric complex.

^(b)The PDB ID of the unbound monomer or the homologue of the monomer (marked by (HU) if the homologue is in an unbound conformation and by (HB) if the homologue is in a bound conformation).

^(c)The full backbone RMSD of the bound and unbound structure of the monomer.

^(d)The backbone RMSD of the interface residues of the bound and unbound structure of the monomer.

^(e)The Symmetry type.

* The cases that were used to manually adjust the electrostatics weights on the energy scoring function.

SymmDock and SymmRef docking results, by using the bound structure of one of the monomers, chosen arbitrarily. The table compares the performance of SymmRef, with and without a symmetric restriction on the interface side-chain conformations.

Table III

SymmDock				SymmRef				SymmRef with symmetric side-chain optimization			
PDB	First Acceptable(a)	Top 10 quality(b)	***/**/*	First acceptable(a)	Original SymmDock solution(c)	Top 10 quality(b)	***/**/*	First acceptable(a)	Original SymmDock solution(c)	Top 10 quality(b)	***/**/*
IURZ (C3)	1 (3.46,1.77,0.73)	0/3/0	1 (2.03,1.09,0.73)	1 (2.03,1.09,0.73)	134 (9.05,4.68,0.23)	4/2/3		4 (7.12,3.36,0.38)	95 (12.39,5.88,0.30)	0/0/4	
IVIM (D2)	1 (4.97,1.88,0.56)	1/1/4	1 (0.79,0.45,0.90)	1 (0.79,0.45,0.90)	70 (4.67,2.27,0.29)	10/0/0		1 (0.94,0.48,0.93)	6 (5.41,2.29,0.56)	10/0/0	
IF23 (C3)	1 (4.08,1.98,0.81)	0/3/0	1 (1.96,1.13,0.80)	1 (1.96,1.13,0.80)	16 (8.49,4.03,0.32)	0/10/0		1 (1.89,1.12,0.82)	80 (7.17,3.81,0.30)	0/10/0	
3D57 (C2)	2 (7.49,2.87,0.59)	0/0/2	1 (3.01,0.92,0.88)	1 (3.01,0.92,0.88)	10 (5.44,2.10,0.76)	5/1/0		1 (3.57,1.02,0.79)	27 (12.50,3.91,0.38)	3/3/0	
ITR0 (D6)	1 (1.86,0.71,0.95)	3/4/2	1 (0.58,0.28,0.91)	1 (0.58,0.28,0.91)	1 (1.86,0.71,0.95)	10/0/0		1 (0.41,0.19,1.00)	2 (2.12,0.74,0.78)	10/0/0	
1D2N (C6)	1 (1.78,0.55,0.91)	1/3/0	1 (0.26,0.11,0.98)	1 (0.26,0.11,0.98)	1 (1.78,0.55,0.91)	10/0/0		1 (0.66,0.23,0.99)	3 (5.75,1.97,0.74)	10/0/0	
1JS0 (C3)	9 (4.83,1.72,0.72)	0/1/0	1 (0.68,0.33,1.00)	1 (0.68,0.33,1.00)	290 (4.26,1.96,0.59)	3/0/0		1 (0.68,0.33,1.00)	290 (4.26,1.96,0.59)	3/0/0	
1B0C (C5)	7(15.13,3.70,0.22)	0/0/1	188(14.42,3.43,0.22)	324 (12.38,3.34,0.30)	324 (12.38,3.34,0.30)	0/0/0		465 (9.82,2.84,0.13)	843 (11.21,3.06,0.04)	0/0/0	
1A3F (C3)	6 (9.22,4.01,0.36)	0/0/1	1 (0.47,0.19,0.92)	112 (4.33,1.60,0.71)	112 (4.33,1.60,0.71)	8/0/0		4 (0.61,0.31,0.92)	38 (3.70,1.62,0.68)	7/0/0	
3MGE (C6)	4 (1.70,0.69,0.88)	3/1/1	1 (0.89,0.27,0.89)	691 (8.84,2.95,0.25)	691 (8.84,2.95,0.25)	10/0/0		2 (0.99,0.29,0.97)	691 (8.84,2.95,0.25)	6/0/0	
2WQH (C2)	105 (7.6,1.9,0.91)	0/0/0	1(0.90,0.34,1.00)	105 (7.6,1.90,0.91)	105 (7.6,1.90,0.91)	2/0/0		1 (0.90,0.34,1.00)	105 (7.6,1.90,0.91)	2/0/1	
1AP9 (C3)	1 (9.00,4.31,0.27)	0/0/3	3 (8.74,4.21,0.20)	1 (9.00,4.31,0.27)	1 (9.00,4.31,0.27)	0/0/1		19 (1.08,0.42,0.84)	49 (5.45,2.31,0.51)	0/0/0	
1BK6 (C2)	13(3.58,1.71,0.76)	0/0/0	6 (13.65,3.87,0.41)	15 (10.85,3.15,0.61)	15 (10.85,3.15,0.61)	0/0/1		7 (1.08,0.39,1.00)	48 (8.32,3.41,0.37)	1/0/0	
1D01 (C3)	76(9.38,4.20,0.43)	0/0/0	597 (9.88,3.41,0.20)	111 (8.61,3.62,0.36)	111 (8.61,3.62,0.36)	0/0/0		797 (9.24,3.67,0.13)	361(9.35,3.26,0.36)	0/0/0	
1DD1 (C3)	57(5.54,2.11,0.52)	0/0/0	1 (1.78,1.73,0.75)	57 (5.54,2.11,0.52)	57 (5.54,2.11,0.52)	0/2/0		1 (1.78,1.71,0.78)	625 (5.84,3.00,0.14)	0/2/0	
1F1G (C2)	1 (8.56,2.29,0.70)	0/0/2	2 (0.51,0.21,1.00)	40 (7.39,2.57,0.50)	40 (7.39,2.57,0.50)	3/0/0		1 (0.67,0.25,0.98)	40 (7.39,2.57,0.50)	3/0/0	

(a)The first acceptable solution (according to CAPRI criteria). The details of the solution are presented in the following format: rank (RMSD, IRMSD, Fnat).

(b)The number of high accuracy (***) medium accuracy (**) and acceptable (*) solutions in the top 10 solutions.

(c)The original SymmDock solution of the first acceptable solution of SymmRef.

Table IV

SymmDock, SymmRef and RosettaDock docking results, by using the unbound structure of a monomer.

SymmDock				SymmRef			RosettaDock		
PDB	First Acceptable ^(a)	Top 10 quality ^(b) ***/**/*	First acceptable ^(a)	Original SymmDock solution ^(c)	Top 10 quality ^(b) ***/**/*	First acceptable ^(a)	Original SymmDock solution ^(c)	Top 10 quality ^(b) ***/**/*	
IURZ (C3)	430(8.61,6.09,0.2)	0/0/0	8 (7.52,6.87,0.24)	622 (6.21,5.74,0.29)	0/0/1	136 (6.34,6.07,0.21)	622 (6.21,5.74,0.29)	0/0/0	
IYIM (D2)	2(4.45, 3.31, 0.15)	0/0/2	1 (4.17,3.80,0.40)	801 (8.97,4.57,0.11)	0/3/1	186 (3.22,3.42,0.46)	480 (29.03,14.96,0.00)	0/0/0	
IF23 (C3)	1 (1.54,0.95,0.76)	1/6/0	1 (2.49,1.57,0.72)	37 (3.17,1.51,0.59)	0/10/0	1 (1.90,1.06,0.70)	3 (2.20,1.25,0.82)	4/6/0	
3D57 (C2)	44(10.36,3.99,0.53)	0/0/0	18 (2.30,1.24,0.74)	81 (7.92,3.34,0.41)	0/0/0	1 (2.98,1.07,0.79)	690 (31.46,14.58,0.15)	0/4/1	
ITR0 (D6)	1 (2.68,1.53,0.66)	0/9/1	1 (1.67,1.08,0.75)	96(4.05,1.71,0.68)	0/10/0	1 (2.35,1.32,0.74)	328 (5.22,2.32,0.75)	0/10/0	
ID2N (C6)	X	0/0/0	X	X	0/0/0	X	X	0/0/0	
LJS0 (C3)	387 (9.80,3.71,0.33)	0/0/0	276 (8.42,3.11,0.25)	940 (14.56,9.79,0.04)	0/0/0	93 (6.11,2.76,0.33)	489 (11.12,9.36,0.11)	0/0/0	
IB0C (C5)	134(14.9,3.56,0.13)	0/0/0	189 (14.52,3.46,0.17)	648 (14.21,3.47,0.35)	0/0/0	17 (15.46,3.64,0.13)	512 (15.97,3.80,0.09)	0/0/0	
IA3F (C3)	29(6.05,2.60,0.62)	0/0/0	57 (3.51,1.69,0.42)	249 (8.11,2.98,0.62)	0/0/0	5 (3.54,1.91,0.39)	616 (5.31,2.39,0.68)	0/2/2	
3MGE (C6)	X	0/0/0	X	X	0/0/0	X	X	0/0/0	
2WQH (C2)	275(12.34,1.90,0.69)	0/0/0	1 (2.32,0.37,0.66)	455 (7.28,2.41,0.44)	4/0/0	75 (17.15,3.76,0.31)	532 (18.26,7.09,0.22)	0/0/0	
1AP9 (C3)	164 (4.48,1.95,0.44)	0/0/0	82 (2.17,1.45,0.47)	164 (4.48,1.95,0.44)	0/0/0	25 (2.32,1.48,0.44)	164 (4.48,1.95,0.44)	0/0/0	
IBK6 (C2)	6 (10.40,3.28,0.71)	0/0/1	2 (6.82,3.15,0.25)	6 (10.40,3.28,0.71)	0/0/1	X	X	0/0/0	
ID01 (C3)	65 (12.36,3.98,0.16)	0/0/0	22 (6.44,3.20,0.16)	438 (6.44,3.20,0.16)	0/0/0	32 (20.57,3.76,0.20)	156 (21.92,4.27,0.13)	0/0/0	
IDD1 (C3)	213 (9.63,4.54,0.21)	0/0/0	439 (7.39,3.85,0.21)	242 (8.40,4.17,0.17)	0/0/0	183 (6.10,3.33,0.26)	521 (6.91,3.65,0.23)	0/0/0	
IF1G (C2)	23 (14.12,3.54,0.40)	0/0/0	3 (8.20,3.72,0.30)	36 (9.42,3.12,0.56)	0/0/1	X	X	0/0/0	

^(a)The first acceptable solution (according to CAPRI criteria). The details of the solution are presented in the following format: rank (RMSD, IRMSD, Fnat). Cases with no acceptable solutions are denoted by 'X'.

^(b)The number of high accuracy (***) medium accuracy (**) and acceptable (*) solutions in the top 10 solutions.

^(c)The original SymmDock solution of the first acceptable solution of SymmRef/RosettaDock.

Table V

Comparison of the performance of our symmetric docking protocol (SymmDock+SymmRef) to other state-of-the-art methods: M-ZDOCK and HADDOCK. The 'X' sign represent a failure in producing any near native solution. The '–' sign represent cases on which HADDOCK could not be run on (hexamers).

PDB	SymmDock+ SymmRef ^(a)	HADDOCK ^(a)	CPORT TP/FP ^(b)	M-ZDOCK ^(a)
1URZ (C3)	8 (7.52,6.87)	X	13/25 (34%)	224 (5.6,6.33)
1VIM (D2)	1 (4.17,3.80)	7 (8.62,4.94)	30/8 (79%)	57 (3.28,3.42)
1F23 (C3)	1 (2.49,1.57)	1 (3.14,2.13)	23/9 (72%)	1 (1.91,1.09)
3D57 (C2)	18 (2.30,1.24)	X	0/27 (0%)	31 (13.93,3.88)
1TR0 (D6)	1 (1.67,1.08)	3 (2.81,1.67)	24/13 (65%)	1 (3.91,1.93)
1D2N (C6)	X	–	–	2 (7.04,5.56)
1JS0 (C3)	276 (8.42,3.11)	X	7/47 (13%)	126 (10.70,3.42)
1B0C (C5)	189 (14.42,3.46)	1 (11.39,3.96)	9/22 (29%)	52 (10.53,3.08)
1A3F (C3)	57 (3.51,1.69)	3 (0.99,1.23)	13/31 (30%)	49 (3.41,1.48)
3MGE (C6)	X	–	–	X
2WQH (C2)	1 (2.32,0.37)	X	15/38 (28%)	139 (3.58,0.87)
1AP9 (C3)	82 (2.17,1.45)	8 (9.83,4.12)	17/96 (15%)	13 (7.65,3.12)
1BK6 (C2)	2 (6.82,3.15)	X	5/44 (10%)	2 (11.74,3.94)
1D01 (C3)	22 (6.44,3.20)	X	8/34 (19%)	107 (5.88,3.17)
1DD1 (C3)	439 (7.39,3.85)	X	3/42 (7%)	97 (6.19,3.29)
1F1G (C2)	3 (8.20,3.72)	1 (8.48,3.75)	14/31 (31%)	19 (9.03,3.38)

^(a) The first acceptable solution (according to CAPRI criteria). The details of the solution are presented in the following format: rank (RMSD, IRMSD).

^(b) The quality of the interface prediction of CPORT, which was given as an input to the HADDOCK method. This column shows the amount of True-Positive (TP) predictions and the amount of False-Positive (FP) predictions. The number in the brackets shows the percentage of correct interface residues in the predicted interface (precision).

Table VI
M-ZDOCK and SymmRef docking results when using the unbound structure of a monomer.

PDB	M-ZDOCK			SymmRef		
	First Acceptable ^(a)	Top 10 quality ^(b) ***/**/*	First acceptable ^(a)	Original M-ZDOCK solution ^(c)	Top 10 quality ^(b) ***/**/*	
IURZ (C3)	224 (5.60,6.33,0.33)	0/0/0	6 (7.95,7.02,0.27)	224 (5.60,6.33,0.33)	0/0/1	
IYIM (D2)	57 (3.28,3.42,0.47)	0/0/0	65 (5.14,3.94,0.24)	57 (3.28,3.42,0.47)	0/0/0	
IF23 (C3)	1 (1.91,1.09,0.77)	1/8/0	1 (1.85,1.10,0.77)	1 (1.91,1.09,0.77)	3/7/0	
3D57 (C2)	31 (13.93,3.88,0.29)	0/0/0	8 (3.11,1.40,0.68)	132 (2.41,1.35,0.79)	0/1/0	
ITR0 (D6)	1 (3.91,1.93,0.63)	0/5/2	1 (1.80,1.78,0.62)	924 (3.68,1.44,0.61)	0/9/0	
ID2N (C6)	2 (7.04,5.56,0.21)	0/0/1	39 (8.30,6.16,0.18)	398 (7.40,5.67,0.22)	0/0/0	
IJS0 (C3)	126 (10.70,3.42,0.50)	0/0/0	589 (10.55,3.37,0.58)	241 (12.15,9.55,0.12)	0/0/0	
IB0C (C5)	52 (10.53,3.08,0.52)	0/0/0	158 (6.79,3.63,0.13)	602 (6.37,3.20,0.22)	0/0/0	
IA3F (C3)	49 (3.41,1.48,0.59)	0/0/0	17 (3.30, 1.66, 0.56)	419 (4.22,1.90,0.50)	0/0/0	
3MGE (C6)	X	0/0/0	X	X	0/0/0	
2WQH (C2)	139 (3.58,0.87,0.84)	0/0/0	3 (3.08,0.54,0.84)	341 (1.75,0.51,0.88)	8/0/0	
IAP9 (C3)	13 (7.65,3.12,0.42)	0/0/0	68 (2.24,1.58,0.44)	687 (2.69,1.58,0.71)	0/0/0	
IBK6 (C2)	2 (11.74,3.94,0.29)	0/0/3	1 (8.65,3.97,0.22)	8 (10.15,3.68,0.22)	0/0/8	
ID01 (C3)	107 (5.88,3.17,0.23)	0/0/0	79 (6.95,3.20,0.16)	107 (5.88,3.17,0.23)	0/0/0	
IDD1 (C3)	97 (6.19,3.29,0.30)	0/0/0	212 (6.59,3.19,0.28)	97 (6.19,3.29,0.30)	0/0/0	
IFIG (C2)	19 (9.03,3.38,0.50)	0/0/0	6 (8.65,3.30,0.44)	88 (9.17,2.84,0.84)	0/0/1	

(a) The first acceptable solution (according to CAPRI criteria). The details of the solution are presented in the following format: rank (RMSD, IRMSD, Fnat).
 (b) The number of high accuracy (***), medium accuracy (**) and acceptable (*) solutions in the top 10 solutions.
 (c) The original M-ZDOCK solution of the first acceptable solution of SymmRef.