Construction of Molecular Assemblies via Docking: Modeling of Tetramers with D2 Symmetry

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2003

Impact Factor: 2.63 · DOI: 10.1002/prot.10480 · Source: PubMed

PROTEINS: Structure, Function, and Genetics 53:817–829 (2003)

Construction of Molecular Assemblies via Docking: Modeling of Tetramers with D_2 Symmetry

Alexander Berchanski[1] and Miriam Eisenstein[2]*
[1]*Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel*
[2]*Department of Chemical Services, Weizmann Institute of Science, Rehovot 76100, Israel*

ABSTRACT **Comparative modeling methods are commonly used to construct models of homologous proteins or oligomers. However, comparative modeling may be inapplicable when the number of subunits in a modeled oligomer is different than in the modeling template. Thus, a dimer cannot be a template for a tetramer because a new monomer-monomer interface must be predicted. We present in this study a new prediction approach, which combines protein-protein docking with either of two tetramer-forming algorithms designed to predict the structures of tetramers with D_2 symmetry. Both algorithms impose symmetry constraints. However, one of them requires identification of two of the C_2 dimers within the tetramer in the docking step, whereas the other, less demanding algorithm, requires identification of only one such dimer. Starting from the structure of one subunit, the procedures successfully reconstructed 16 known D_2 tetramers, which crystallize with either a monomer, a dimer or a tetramer in the asymmetric unit. In some cases we obtained clusters of native-like tetramers that differ in the relative rotation of the two identical dimers within the tetramer. The predicted structural pliability for concanavalin-A, phosphofructokinase, and fructose-1,6-bisphosphatase agrees semiquantitatively with the observed differences between the several experimental structures of these tetramers. Hence, our procedure identifies a structural soft-mode that allows regulation via relative rigid-body movements of the dimers within these tetramers. The algorithm also predicted three nearly correct tetramers from model structures of single subunits, which were constructed by comparative modeling from subunits of homologous tetrameric, dimeric, or hexameric systems. Proteins 2003;53: 817–829.** © 2003 Wiley-Liss, Inc.

Key words: molecular docking; tetramers; comparative modeling; soft modes; D_2 symmetry; allosteric regulation

INTRODUCTION

Knowledge of the three-dimensional (3D) structures of proteins has proved indispensable in modern research. In particular, the structures of molecular complexes and oligomers provide detailed information regarding the interactions between molecules and allow analysis and comparison of the forces that stabilize nonpermanent versus permanent complexes.[1–3] Experimental methods, such as X-ray crystallography and NMR spectroscopy, are successfully used to determine the 3D structures of oligomeric complexes. Yet, in view of the fast accumulation of new sequence data it is important to develop theoretical tools for predicting the structures of complexes and assemblies starting from the structures of the individual molecules. Oligomers present a special problem in modeling. Although a single subunit can be modeled based on the structure of a homologous protein (comparative modeling), the oligomerization mode of the subunits may be different in different tissues and species. This problem can be overcome by combining comparative modeling with docking.

Complex formation and oligomerization require that the component molecules specifically recognize one another. Such recognition is attained through geometric and chemical complementarity of the interacting surfaces.[4–6] Both soluble and membrane-bound proteins commonly form symmetric oligomeric assemblies *in vivo,* with two or more identical subunits. Oligomers provide significant advantages over monomeric proteins, such as stability and regulation, as recently reviewed by Goodsell and Olson.[7] Assemblies of proteins can only have rotational symmetry because other symmetry operations convert chiral L-residues to D-residues. Several kinds of rotational symmetry occur in oligomeric proteins: cyclic symmetry, dihedral symmetry, and cubic symmetry. In the case of cyclic symmetry, C_n, the n subunits are related by rotations of $360°/n$ about a single axis. C_2 symmetry is particularly common in proteins (homodimers). Combination of an n-fold rotation axis with perpendicular twofold rotation axes produces dihedral symmetry D_n. Oligomers with an even number of subunits are often n-mers of symmetric dimers. Tetramers with D_2 symmetry are very common among soluble proteins.[7]

Grant sponsor: Daat Consortium for Developing Generic Technologies for Design and Development of Drugs and Diagnostics.

*Correspondence to: Miriam Eisenstein, Department of Chemical Services, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: miriam.eisenstein@weizmann.ac.il.

Received 10 February 2003; Accepted 1 April 2003

Prediction of the structures of oligomers can be achieved by combining binary docking with different aggregate forming algorithms. The underlying assumption of this approach is that the molecules in an oligomeric aggregate form good contacts with each other, but they do not necessarily interact in the oligomer as they do in an isolated binary complex. Thus, the molecules in an aggregate do not always utilize the best surface match, and the most stable structure of a supra-molecular complex is characterized by the best combined-fit, which is the sum of all the binary matches within a given system. Previously, our group verified this postulate for helical assemblies, represented by the protein coat of the tobacco mosaic virus. Thus, geometric docking was combined with a helix-forming algorithm appropriate for constructing supramolecular helices, in which all the subunits are identical.[8] The procedure successfully reconstructed the helical protein coat of the tobacco mosaic virus, starting from a single subunit. Moreover, it also predicted the structure of a disc of 17 subunits, which closely resembles one layer in the double-layered 34-mer that initiates the virus assembly, and of possible intermediates in the virus assembly process. The latter can explain how the initially formed discs transform into a helix without dissociating into subunits.[9] We are aware of only a few studies in which prediction of oligomeric structures was attempted. Docking under C_2 symmetry constraint was used to study the R-T transition pathway in hemoglobin.[10,11] More recently Harris et al.[12] constructed a helical assembly of the influenza matrix protein M1, based on the M1-M1 interactions in the crystal and Gardiner et al.[13] reassembled in part the heptamer of the proteosome activator reg-alpha.

We choose to further apply our aggregate forming approach to tetramers. Tetrameric proteins are common and they have important biological functions. Therefore, it is of interest to develop a modeling method for such proteins and to study their oligomerization characteristics. Examination of the arrangement of the subunits in 76 tetramers, whose structures were deposited in the Protein Data Bank[14] (PDB; version of August 1998), indicated that 72 of them were D_2 tetramers. Hence, a tetramer-forming algorithm, which predicts the structures of D_2 tetramers starting from a single subunit, is likely to be useful, especially when combined with comparative modeling. For example, the Fe-Mn superoxide dismutase from *Propionibacterium shermanii*[15] (PDB code: 1ar5) forms a tetramer, whereas the homologous protein, Mn superoxide dismutase from *Escherichia coli*[16] (PDB code 1vew) forms a dimer. If the structure of the tetrameric form of the Fe-Mn superoxide dismutase was not known, it could be predicted by applying a tetramer-forming algorithm to a model structure of one subunit, constructed by means of comparative modeling from the known structure of the homologous dimeric Mn superoxide dismutase (see below).

We describe in the present study two homotetramer-forming procedures, which combine docking with symmetry considerations to form D_2 tetramers. Both algorithms impose symmetry constraints. However, they have different starting requirements. One algorithm requires the identification of two different homodimers within the tetramer. These are combined, using vector algebra considerations to form the tetramer. The other algorithm requires the identification of only one homodimer. Two copies of this dimer are combined in a second docking step to form a tetramer. The second algorithm is seemingly less demanding, but it requires considerably longer computation times. In this study we examine the performance of these algorithms for reassembling tetramers from a single subunit and for predicting the structures of tetramers starting from the model structure of a single subunit.

The procedures were applied to 16 tetramers and successfully reconstructed all of them. These test systems were selected randomly form the PDB provided their sequence identity was $<30\%$. We, however, made sure that the list included D_2 tetramers with a near planar, near tetrahedral and intermediate arrangements of the centers of mass of the subunits and tetramers that crystallize with either a monomer, a dimer or a tetramer in the asymmetric unit. The systems are listed in Table I.

A detailed analysis of the predicted structures indicates that in some cases clusters of similar nearly correct solutions are formed, in which two dimers in the tetramer are rotated with respect to one another, while conforming to the D_2 symmetry. The predicted range of rotation for concanavalin-A, phosphofructokinase, and fructose-1,6-bisphosphatase is similar to the observed differences between the several experimental structures for these tetramer. Hence, our tetramer-forming algorithm designed to be a predictive tool, can also serve as an analysis tool for studying the interfaces in tetrameric oligomers.

Finally, we used our methods to form D_2 tetramers of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from *Pseudomonas* sp,[33] Fe-Mn superoxide dismutase from *P. shermanii,*[15] and nucleoside diphosphate kinase from *Myxococcus xanthus*[34] from model-structures of single subunits. The models were constructed by comparative modeling using one subunit form the tetrameric tropinone reductase-I from *Datura stramonium*,[35] the dimeric Mn superoxide dismutase from *E. coli*[16] and the hexameric nucleoside diphosphate kinase from *Dictyostelium discoideum*,[36] respectively. The structures of the modeling templates are known and so are the structures of the modeled tetramers, allowing validation of our predictions. Despite the limited accuracy of the modeled subunits our procedure predicted nearly correct structures of the tetramers in every case.

METHODS

Formation of Tetramers with D_2 Symmetry

The formation of homotetramers with D_2 symmetry is achieved in two steps. The first step is the identification of homodimers via molecular docking using an algorithm previously developed in our group.[37] In this step different docking procedures can be used, such that rely solely on geometric recognition, as described before for helical assemblies.[8] Alternatively, procedures relying on geometric-electrostatic[38] or geometric-hydrophobic recognition (Berchanski and Eisenstein, unpublished results) can be used. In the current study only geometric recognition is consid-

TABLE I. List of the Tetrameric Systems Used to Develop and Validate our Tetramer-Forming Procedures

System	PDB code	Dimer-dimer angles (°)[a]	Reference
4-hydroxybenzoyl CoA thioesterase	1bvq	47.4; 60.0	17
Dialkylglycine decaboxylase	1dge	3.5; 2.2	18
Enol acyl carrier protein reductase	1enp	31.6; 41.6	19
Fructose-1,6-bisphosphatase	1bk4	9.7; 13.4	20
Concanavalin-A	1gic	80.4; 79.6	21
Glycine N-methyl transferase	1xva	5.3; 4.2	22
Tetrahydromethanopterin formyltransferase	1ftr	51.7; 37.7	23
Phosphoglycerate mutase 1	1bq4	18.2; 17.9	24
D-glyceraldehyde-3-phosphate dehydrogenase	1a7k	88.5; 84.9	25
Pyrrolidone carboxyl aminopeptidase	1a2z	35.2; 27.4	26
Fructose 1,6-bisphosphate aldolase	1ado	49.1; 50.1	27
Phosphofructokinase	6pfk	73.0; 53.3	28
Platelet factor 4	1rhp	63.9; 46.2	29
NADP-dependent alcohol dehydrogenase	1ykf	82.2; 88.8	30
3α, 20β-hydroxysteroid dehydrogenase	1hdc	25.6; 26.7	31
L-asparagine amidohydrilase	4eca	77.0; 56.0	32

[a]The dimer-dimer angles are the angles between the vectors connecting the centers of mass of the AB and CD and the AC and BD subunits, respectively (for example see angle ε in figure 1).

ered; hence, the homodimers formed in the docking step are evaluated only by the geometric complementarity at the interface (referred to as "score" hereafter). In line with the assembly-formation postulate, we save in the docking step the best homodimers and also additional ones, whose complementarity scores are lower.

In the second step the homodimers obtained via docking are assembled, under given symmetry constraints, to form tetramers with D_2 symmetry. Each subunit in a D_2 tetramer is related to the three other subunits by rotations about three perpendicular twofold-symmetry axes originating at the center of mass of the tetramer. Hence, three different symmetric homodimers are found in a D_2 tetramer, the **ab**, **ac**, and **ad** dimers. Interestingly, in most structures the subunits are in contact in only two of the three homodimers.

We present two tetramer-forming algorithms: The **ab/cd** algorithm, which forms tetramers on the basis of only one dimer (**ab**), and the **ab/ac** algorithm, which needs information regarding two homodimers, **ab** and **ac**, and both dimers must be identified in the docking step.

Identification of symmetric homodimers via docking

The geometric docking algorithm developed by our group was implemented in a computer program named MolFit, as described previously.[37] In the MolFit scan one molecule (denoted **b**) is rotated and translated in steps with respect to another molecule (denoted **a**), spanning the whole rotation/translation space. MolFit produces a list of binary matches, each associated with a score, which represents the quality of the match, and with six position parameters: Three Euler angles defining the rotation matrix of **b** with respect to **a** (M), and three components of the translation vector connecting the centers of mass of the two molecules (**V**). In the current study we require only symmetric binary matches with twofold symmetry. Therefore, we use only those combinations of Euler angles in which the rotation about the eigenvector is near 180°.

Geometric rotation/translation docking scans were performed for each of the 48 independent subunits in the 16 systems listed in Table I. The MolFit grid interval was in the range 1.08–1.29 Å and the angular interval was 12°. The twofold symmetry restriction limited the scan to 1229 relative orientations and for each orientation we saved either 1 or 5 solutions (L1 and L5 lists, respectively). Each scan required 3.5–4 cpu hours on a SGI workstation with a R12000 processor.

The **ab/cd** *tetramer-forming algorithm*

The **ab/cd** tetramer-forming algorithm relies on the fact that in a perfect D_2 tetramer the dimers **ab** and **cd** are structurally identical. Only **ab** dimers in which the translation vector $\mathbf{V}_i$ is perpendicular to $\mathbf{E}_i$ (the eigenvector of the rotation matrix M_i, which relates subunit **b** to **a**) are accepted. Each **ab** dimer i is transformed from the MolFit coordinates system to a new system ($\mathbf{X}_t$, $\mathbf{Y}_t$, $\mathbf{Z}_t$), whose origin, **O**, is the midpoint of the vector $\mathbf{V}_i$, its $\mathbf{Z}_t$ axis points along the eigenvector $\mathbf{E}_i$ and the $\mathbf{Y}_t$ axis points along the **OA** vector (see Fig. 1).

Rotation of the **ab** homodimer by 180° about the $\mathbf{Y}_t$ axis produces the corresponding **cd** dimer, and in order to obtain a tetramer we use MolFit again to dock dimers **ab** and **cd** against one another. However, only a part of the rotation-translation space is considered. The **cd** dimer is rotated only about the $\mathbf{Z}_t$ axis in 10° steps (in the range 0–170°) and for each relative orientation we keep five high-scoring solutions in which the **cd** dimer is translated approximately along the $\mathbf{Z}_t$ axis. The procedure is repeated for other top ranking **ab** dimers in the L5 list. The dimer-dimer docking step requires ~3 cpu hours for 150 **ab** dimers (on a SGI workstation with a R12000 processor). The predicted tetramers are sorted by the combined complementarity score, which is the dimer-dimer complementarity score plus twice the score for the **ab** contact. It was mentioned above that in most homotetramers the subunits are in contact in only two of the three possible

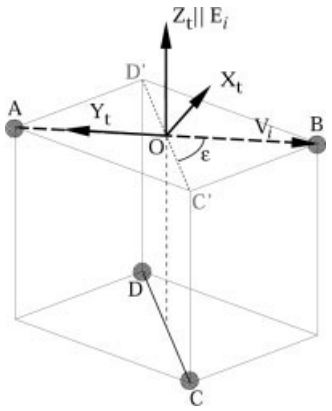

Fig. 1. The X_t, Y_t, Z_t coordinate system used in the **ab**/**cd** tetramer-forming algorithm. A, B, C, and D denote the centers of mass of the subunits in the tetramer. $\mathbf{V}_i$ is the inter subunit vector in the **ab** dimer and E_i is the eigenvector of the rotation matrix that relates molecule **b** to molecule **a**. Z_t is chosen to be parallel to E_i. Rotation of the **ab** dimer about the $\mathbf{Y}_t$ axis produces the **cd** dimer, which is then rotated and translated about the $\mathbf{Z}_t$ axis to produce the tetramer. **O** is the origin of the tetramer's coordinates system, at the midpoint of the vector $\mathbf{V}_i$. The C′–D′ dashed line is the projection of the inter subunit vector in the **cd** dimer onto the $\mathbf{X}_t\mathbf{Y}_t$ plane and ϵ is the angle between the **ab** and **cd** dimers.

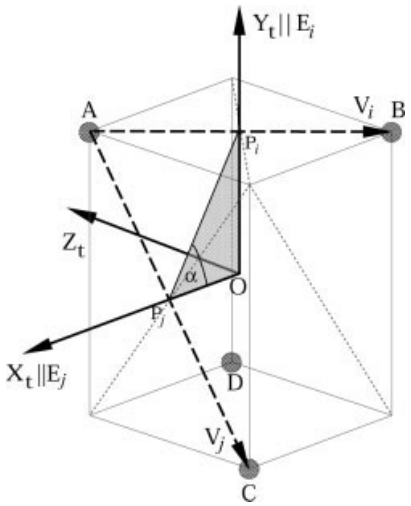

Fig. 2. The X_t, Y_t, Z_t coordinate system used in the **ab**/**ac** tetramer-forming algorithm. A, B, C, and D denote the centers of mass of the subunits in the tetramer. $\mathbf{V}_i$ and $\mathbf{V}_j$ are the inter subunit vectors in the **ab** and **ac** dimers, which are combined to form a tetramer. **O** is the origin of the tetramer's coordinates system and the meeting point of the eigenvectors $\mathbf{E}_i$ and $\mathbf{E}_j$. It is calculated as $\mathbf{O} = \mathbf{P}_i - |\mathbf{P}_{ij}| \cdot \sin\alpha \cdot \mathbf{E}_i$, where $\mathbf{P}_i$ and $\mathbf{P}_j$ are the midpoints of $\mathbf{V}_i$ and $\mathbf{V}_j$, respectively, $|\mathbf{P}_{ij}|$ is the distance between $\mathbf{P}_i$ and $\mathbf{P}_j$, and α is the angle between the vectors $\mathbf{P}_{ij} = \mathbf{P}_i - \mathbf{P}_j$ and $\mathbf{E}_j$.

homodimers. Hence, our combined complementarity score is the sum of scores for the dominant four interactions of the six possible subunit-subunit contacts in the tetramer.

The ab/ac *tetramer-forming algorithm*

The **ab/ac** tetramer-forming algorithm requires only one docking step, and we assume that the two C_2 homodimers **ab** and **ac** are identified in this scan. To test if a given pair of MolFit solutions, i and j, can form a D_2 tetramer we first check if the eigenvectors $\mathbf{E}_i$ and $\mathbf{E}_j$ of the corresponding rotation matrices M_i and M_j are perpendicular to one another. Next, we check if the translation vectors $\mathbf{V}_i$ and $\mathbf{V}_j$ are perpendicular to the eigenvectors $\mathbf{E}_i$ and $\mathbf{E}_j$, respectively (see Fig. 2). Pairs of homodimers, which satisfy these conditions, are used to form D_2 tetramers. The computations are done in the tetramer's coordinates system, whose origin, **O**, is the meeting point of eigenvectors $\mathbf{E}_i$ and $\mathbf{E}_j$. The position of **O** is calculated as shown in Figure 2.

The principal axes of the tetramer can point either in the positive direction of eigenvectors $\mathbf{E}_i$ and $\mathbf{E}_j$ or in the negative direction. Therefore, these eigenvectors can be combined in four alternative ways: (i) Both originate from the point **O**; (ii) Both point toward the point **O**; (iii) $\mathbf{E}_i$ points toward **O** and $\mathbf{E}_j$ points away from point **O**, and (iv) $\mathbf{E}_i$ points away from point **O** and $\mathbf{E}_j$ points toward it. All four options are considered in our algorithm, which also checks which ones are acceptable. For each option we calculate the coordinates of point **D** the putative center of mass of subunit **d** (see below) and compare the distances **AD** and **BC**. We note that if **D** is calculated using option (i) and **AD** is approximately equal to **BC**, then the **AD** distance calculated with option (ii) is very different from **BC** and option (ii) is eliminated. Correspondingly, if option (ii) is correct, then option (i) is eliminated. The situation is similar for the second pair of options, (iii) and (iv). We usually obtain two acceptable positions for the tetramer's center of mass, due to the approximate positioning of subunits **b** and **c** by MolFit. We find that the best approximation for **O** is the average of the two acceptable positions.

Next, the tetramer's coordinates system ($\mathbf{X}_t$, $\mathbf{Y}_t$, $\mathbf{Z}_t$) is built, in which $\mathbf{Z}_t$ is the vector product of the normalized eigenvectors $\mathbf{E}_j$ and $\mathbf{E}_i$, and $\mathbf{X}_t$ and $\mathbf{Y}_t$ are approximately parallel to $\mathbf{E}_j$ and $\mathbf{E}_i$, respectively (see Fig. 2). In this coordinates system the position of subunit **d** is obtained by rotation of subunit **a** by 180° about the $\mathbf{Z}_t$ axis. Tetramers are formed for all the pairs of homodimers, which conform to the D_2 symmetry restrictions. All the predicted tetramers are then sorted by the combined tetramer complementarity score, which is twice the sum of the scores for the **ab** and **ac** dimers. The **ab/ac** algorithm is fast, requiring 10–15 cpu minutes (on a SGI workstation with a R12000 processor) to process all the homodimers produced in the docking step.

Error-range parameters

MolFit uses a stepwise rotation/translation scanning method and therefore, the dimers formed through MolFit are approximate. Hence, the directions of the translation vectors $\mathbf{V}_i$ and $\mathbf{V}_j$, as well as the directions of the eigenvectors $\mathbf{E}_i$ and $\mathbf{E}_j$, derived from the corresponding rotation matrices, are not exact. To ensure that the tetramer-forming algorithms can use approximate dimers to form tetramers on the one hand, and to reject inappropriate dimers on the other hand, we allow limited error ranges. For example, the rotation about the eigenvector, which must be 180° in an exact C_2 dimer, is allowed to be 180 ± 6°. In addition, the angle between the translation vectors

$\mathbf{V}_i$ and $\mathbf{V}_j$, which is 60° in a perfect tetrahedron, 90° in a planar D_2 tetramer and adopts an intermediate value in most cases, is allowed a slightly wider range, between 54° and 92°. Also, the angles between the eigenvectors and the angles between the translation vectors and the corresponding eigenvectors, which must be 90° (see Figs. 1 and 2), are allowed a range of 90 ± 3°. Finally, the distance **A**–**D**, which must be equal to the distance **B**–**C**, is permitted to deviate by of up to 3 Å.

Determination of the Ranks of the Nearly Correct Predictions

The predicted tetramers are compared to the experimental structures by calculating the root mean square difference (RMSD) between the positions of corresponding C_α atoms. The rank of the nearly correct tetramer is its position in the sorted list of tetramers, and it is determined by searching for the highest scoring tetramer with a RMSD < 3.5 Å. In predictions based on modeled subunits, which deviate considerably from the experimental structures, we allow a larger deviation, up to 5 Å. Predictions within this RMSD are likely to provide significant structural information that can be used for planning and designing experiments.

Statistical Analysis

It is useful to provide each solution with a number that represents its statistical significance. We fit an extreme value distribution function[39] to the observed distribution of scores in each MolFit scan to obtain estimates for the mean score (μ) and for the standard deviation (σ). The statistical significance of the ith solution in the MolFit scan is $Z_i = (S_i - \mu)/\sigma$, where S_i is the score of that solution. Z_i is measured in units of σ.

The **ab/ac** tetramers are formed from pairs of dimers from the same distribution. The mean score for such a tetramer is therefore, equal to 2μ and the standard deviation is 2σ. Hence, the Z value for tetramer k is $Z_k = (S_i + S_j - 2\mu)/2\sigma$, where S_i and S_j are the scores of the homodimers i and j that were combined to form tetramer k. The Z values for tetramers formed via the **ab/cd** algorithm are less readily estimated. This algorithm involves two different distributions of solutions, the distribution of homodimers and the distribution of tetramers. The latter scan is confined to rotations about only one axis, and it is too limited to allow fitting to a theoretical distribution function.

Molecular Modeling

We modeled the structures of the single subunits of three oligomeric proteins (listed in the Introduction). The sequences of the modeled and template proteins were aligned using FASTA.[40] Initial models were built using the Homology module of InsightII (Accelrys Inc., San Diego, CA). First, residues in the more conserved segments were replaced with those of the modeled protein. Next, loops in which inserts or deletions occur were automatically generated. The initial models were then checked for close atom-atom contacts. Such clashes were eliminated by choosing different rotamers for the side chains. Finally, the model underwent energy minimization, in which the $C\alpha$ atoms were constrained to their initial positions, except in the generated loops. We followed the procedure proposed by InsightII in the production of the models. These are not necessarily the most accurate models; hence, better rotamer libraries could perhaps be used. We made no such effort because in this study we test the performance of the tetramer-forming algorithms when applied to standard models with limited accuracy.

RESULTS AND DISCUSSION

The tetramer-forming procedures described above were tested on 16 tetrameric systems, selected as described in the Introduction. The list includes 4 perfect tetramers formed through application of crystallographic symmetry operations to the single monomer in the asymmetric unit. These cases are comparable to docking of disassembled complexes, or "bound docking." We also tested 12 tetramers that crystallize as dimers or complete tetramers in the asymmetric unit. The individual subunits in such tetramers are not related to each other by crystallographic symmetry and exhibit small structural differences. Therefore, these predictions, based on the structure of a single subunit (we used each subunit as the starting structure and docked it against itself) may be considered as docking "unbound" structures. The correspondence is limited because the structural differences between subunits are generally smaller than the differences between the bound and unbound structures that we examined in a previous study[38] (0.1–1.9 vs. 0.6–2.5 Å). Finally, we applied the algorithm to three model structures of single subunits.

Reassembling Tetramers with the ab/ac Tetramer-forming Algorithm

The **ab/ac** tetramer-forming algorithm was applied to the L1 and L5 lists of homodimers, producing 110–140 or 700–800 tetramers, respectively, from each list. The results in terms of score, rank, and the statistical significance of the nearly correct tetramers and the dimers within them are shown in Table II. It is important to note that the **ab** dimers in Table II are not necessarily the A-B dimers in the experimental structures. They are homodimers identified by MolFit, which together with other homodimers, denoted **ac**, form nearly correct tetramers. Therefore, the **ab** dimer always has a higher score than the **ac** dimer.

In the 4 systems, in which crystallographic symmetry forms perfect tetramers, the nearly correct tetramers are ranked first. The **ab** dimers are ranked first in both, the L1 and L5 lists and the **ac** dimers are ranked among the top 28 solutions. We then applied the **ab/ac** tetramer-forming procedure to the independent subunits in 2 systems, which crystallize as dimers in the asymmetric unit and 10 systems, which crystallize as complete tetramers in the asymmetric unit (44 independent subunits). In 11 of these 12 systems at least one subunit gives very good results, with the nearly correct tetramer ranked first. In the12th system, 1gic, a nearly correct solution is ranked second

TABLE II. Reassembling D_2 Tetramers with the ab/ac Tetramer-forming Algorithm

	ab dimer			**ac** dimer				D_2 tetramer			
System[a]	Score	Rank[b]	Z[c]	Score	Rank[b]	Z[c]	Z_1	Score	Rank	Z[c]	Z_1
Monomer in the asymmetric unit											
1bvq	871	1	12.1	563	28	4.2	12.1	2868	1	8.2	8.2
1dge	3378	1	70.0	918	3	11.0	70.0	8592	1	40.3	40.3
1enp	833	1	11.9	560	16	4.8	11.9	2786	1	8.3	8.3
1bk4	1340	1	20.4	986	2	12.4	20.4	4652	1	16.4	16.4
Dimer in the asymmetric unit											
1gic/a	524	9	4.3	456	90	2.5	7.0	1960	9	3.4	4.3
1gic/b	585	4	5.4	523	20	3.7	7.3	2216	2	4.6	4.9
1xva/a	1098	1	19.5	480	83	2.7	19.5	3156	1	11.1	11.1
L5	1098	1	25.9	480	91	5.2	25.9	3156	1	15.4	15.4
1xva/b	1127	1	20.2	467	88	2.5	20.2	3188	1	11.4	11.4
L5	1127	1	28.9	467	94	4.9	28.9	3188	1	17.0	17.0
Tetramer in the asymmetric unit:											
1 ftr/a	716	6	7.6	485	173	1.9	10.4	2402	7	4.8	6.2
L5	802	2	14.4	406	1443	1.5	15.3	2416	10	7.9	9.7
1ftr/b	862	2	5.5	498	107	−0.1	6.8	2720	2	2.8	3.4
L5	943	1	19.0	473	223	3.6	19.0	2832	2	11.3	11.9
1ftr/c	776	4	9.0	475	236	1.6	14.5	2502	6	5.3	7.7
L5	776	4	13.0	475	291	3.4	19.9	2502	11	8.2	11.3
1ftr/d	872	1	12.0	550	31	3.7	12.0	2844	1	7.9	7.9
L5	872	1	16.2	550	32	6.0	16.2	2844	1	11.2	11.2
1bq4/a	675	1	6.8	545	28	3.5	6.8	2440	1	5.2	5.2
L5	675	1	10.2	545	31	6.0	10.2	2440	1	8.1	8.1
1bq4/b	559	17	4.0	513	60	2.8	5.3	2144	1	3.4	3.4
1bq4/c	652	5	5.9	602	13	4.7	9.7	2508	1	5.3	5.3
1bq4/d	504	134	2.2	458	347	1.1	7.1	1924	33	1.7	3.7
L5	504	143	4.3	458	444	2.9	10.4	1924	85	3.6	6.1
1a7k/a	1287	1	22.1	558	35	3.7	22.1	3690	1	12.9	12.9
L5	1287	1	28.6	558	36	5.9	28.6	3690	1	17.3	17.3
1a7k/b	1185	1	19.4	599	21	4.4	19.4	3568	1	11.9	11.9
1a7k/c	1251	1	22.2	683	5	7.3	22.2	3268	1	14.8	14.8
1a7k/d	1276	1	21.6	624	8	5.2	21.6	3800	1	13.4	13.4
1a2z/a	667	1	8.5	389	615	0.4	8.5	2112	3	4.5	5.5
L5	667	1	11.3	526	11	6.5	11.3	2386	1	8.9	8.9
1a2z/b	626	1	6.6	572	6	5.1	6.6	2396	1	5.9	5.9
1a2z/c	567	12	5.2	540	14	4.5	6.1	2214	1	4.9	4.9
L5	567	12	7.6	540	15	6.7	8.6	2214	1	7.1	7.1
1a2z/d	563	13	4.5	488	83	2.6	5.9	2102	2	3.6	5.4
1ado/a	743	1	8.6	681	3	7.0	8.6	2848	1	7.8	7.8
1ado/b	858	1	11.8	856	2	11.7	11.8	3428	1	11.8	11.8
1ado/c	806	1	8.7	679	3	5.8	8.7	2970	1	7.3	7.3
1ado/d	839	1	9.9	682	2	6.1	9.9	3042	1	8.1	8.1
6pfk/a	1096	1	16.2	615	12	4.4	16.2	3422	1	10.6	10.6
6pfk/b	1123	1	15.4	796	3	8.2	15.4	3838	1	11.8	11.8
6pfk/c	1067	1	14.1	691	5	6.0	14.1	3516	1	10.1	10.1
6pfk/d	989	1	13.1	683	6	6.1	13.1	3344	1	9.6	9.6
1rhp/a	593	1	7.6	462	20	4.0	7.6	2110	1	5.8	5.8
1rhp/b	401	103	2.4	381	197	1.8	8.9	1564	22	2.1	5.5
L5	401	112	4.3	381	237	3.5	11.9	1564	35	3.9	8.0
1rhp/c	469	2	7.9	458	4	7.4	8.7	1854	1	7.7	7.7

TABLE II. (Continued)

System[a]	**ab** dimer Score	**ab** dimer Rank[b]	**ab** dimer Z[c]	**ac** dimer Score	**ac** dimer Rank[b]	**ac** dimer Z[c]	Z_1	D_2 tetramer Score	D_2 tetramer Rank	D_2 tetramer Z[c]	Z_1
Tetramer in the asymmetric unit:											
1rhp/d	612	1	9.4	390	100	2.4	9.4	2004	1	5.9	5.9
L5	612	1	12.9	390	109	4.4	12.9	2004	1	8.6	8.6
1ykf/a	1174	1	15.4	714	4	5.5	15.4	3776	1	10.5	10.5
1ykf/b	1085	1	14.6	652	16	4.6	14.6	3474	1	9.6	9.6
1ykf/c	1203	1	14.8	882	2	8.3	14.8	4170	1	11.6	11.6
1ykf/d	1466	1	23.2	722	4	6.1	23.2	4376	1	14.7	14.7
1hdc/a	1518	1	26.2	521	99	2.3	26.2	4078	1	14.3	14.3
L5	1518	1	36.0	521	110	4.6	36.0	4078	1	20.3	20.3
1hdc/b	583	13	4.5	567	25	4.1	29.6	2300	7	4.3	5.6
L5	1568	1	38.4	395	1748	1.1	38.4	3926	1	19.8	19.8
1hdc/c	587	21	4.2	558	45	3.5	25.2	2290	15	3.9	6.8
L5	587	23	7.0	558	48	6.1	34.5	2290	28	6.6	17.3
1hdc/d	Not found			Not found				Not found			
4eca/a	1813	1	28.0	879	2	8.6	28.0	5384	1	18.3	18.3
4eca/b	2098	1	36.2	844	3	8.6	36.2	5884	1	22.4	22.4
4eca/c	2039	1	30.1	914	2	8.7	30.1	5906	1	19.6	19.6
4eca/d	2182	1	34.4	874	2	8.2	34.4	6112	1	21.3	21.3

[a]The subunits are listed by their PDB code and chain identifier. All the results obtained with the L1 lists are shown. Results for the L5 lists are shown only when either the scores or the ranks of dimers that form the nearly correct tetramer differ from the corresponding values obtained with the L1 lists.
[b]In a former publication[38] we listed ranges of ranks when the score of the nearly correct dimer was identical to the score of other solutions. In this table the ranks of the dimers should sometimes be a range of ranks but we present only the central value of the range to save space.
[c]The statistical significance (Z) for each dimer and tetramer was calculated as described in the Methods section. Z_1 is the statistical significance for the dimer or tetramer with the highest score.

and its *Z* value is only 0.3 σ lower than that of the top-ranking false solution (see Table II). Notably, a nearly correct tetramer is formed for 43 of the 44 subunits, with 1hdc/d being the only exception. A somewhat less accurate solution for 1hdc/d is ranked 6 (RMSD of 3.75 Å). As mentioned above, the **ab** homodimers have by definition a higher score than the corresponding **ac** dimers. The ratio of the **ac** and **ab** scores has a very large range, from 0.25 to 1.0, suggesting that in some tetramers the interface areas for the two dimers differ much more than in others. Our procedure is only mildly sensitive to this difference. Thus, when both interfaces are small the nearly correct tetramer is often ranked lower than 1.

Next we compare the results obtained for the L1 and L5 lists of homodimers. For each of the 4 perfect tetramers, the same tetramers are predicted when either the L1 or the L5 lists are used, from the same **ab** and **ac** dimers, which also have the same ranks. Hence, saving of more than one solution per orientation in not essential in these systems, but it does not add new false positive predictions either. For the systems with a dimer or a tetramer in the asymmetric unit a nearly correct solution is ranked first in 32 of the 44 cases when the L1 lists are used. The number increases to 34 when the L5 lists are used. Hence, the L5 lists have a marginal advantage over the L1 lists.

We note that for each of the 10 systems in which a complete tetramer is found in the asymmetric unit at least one of the 4 predictions (that start from the 4 independent subunits) ranks the nearly correct tetramer first. Therefore, for 3 systems for which the results are less satisfactory for some subunits (1ftr, 1bq4, and 1hdc) we merged the four L5 lists of MolFit homodimers and used the united list as input to the **ab/ac** tetramer-forming algorithm. The results of these predictions are shown in Table III. In each case an **ab** dimer from one L5 list pairs with an **ac** dimer from another list to form the nearly correct tetramer, thereby increasing its combined score and elevating its rank. This procedure, in which different conformers of the same molecule are docked and the lists of solutions are united, may prove useful in "unbound docking," where conformation differences between bound and unbound molecules significantly affect the ability of docking algorithms to identify the correct solution.

Reassembling Tetramers with the ab/cd Tetramer-forming Algorithm

The **ab/cd** tetramer-forming procedure uses a second docking step. Therefore, it was applied to a limited number of cases: Two systems with a monomer in the asymmetric unit, and 10 subunits from systems with a dimer or a tetramer in the asymmetric unit, for which the nearly correct tetramer predicted with the **ab/ac** tetramer-forming algorithm was not ranked first.

TABLE III. Application of the ab/ac Tetramer-forming Algorithm to United L5 Lists of Homodimers

	ab dimer		**ac** dimer		D_2 tetramer	
System	Score	Original list	Score	Original list	Score	Rank
1ftr[a]	1108	1ftr/d	518	1ftr/b	3252	5
1bq4	684	1bq4/a	554	1bq4/b	2476	1
1hdc	1518	1hdc/a	576	1hdc/c	4188	1

[a]The nearly correct tetramer for 1ftr is ranked fifth, but the four solutions that precede it are almost identical, therefore in practice the nearly correct solution ranks 2nd.

TABLE IV. Comparison of the Nearly Correct Tetramers Formed with the ab/cd and the ab/ac Tetramer-forming Algorithms

	The **ab/cd** algorithm				The **ab/ac** algorithm			
	ab dimer		tetramer		**ab** dimer		tetramer	
System	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Monomer in the asymmetric unit								
1bvq	871	1	3168	1	871	1	2868	1
1enp	833	1	2644	1	833	1	2786	1
Dimer in the asymmetric unit								
1gic/a	524	9	1645	1	524	9	1960	9
1gic/b	585	4	1906	1	585	4	2216	2
Tetramer in the asymmetric unit								
1ftr/a	802	2	2525	3	802	2	2416	10
1ftr/b	862	2	2598	1	943	1	2832	2
1ftr/c	995	1	2884	1	776	4	2502	11
1bq4/d	504	145	1877	7	504	143	1924	85
1a2z/d	579	7	1858	2	563	13	2102	2
1rhp/b	465	9	2237	14	401	112	1564	35
1hdc/c	1439	1	3594	1	587	23	2290	28
1hdc/d	1242	1	3259	1	Not found			

The predictions obtained with the **ab/cd** tetramer-forming algorithm are compared in Table IV with the corresponding results obtained with the **ab/ac** tetramer-forming algorithm. In the two systems with a single subunit in the asymmetric unit a nearly correct tetramer is ranked first when either tetramer-forming algorithm is used, and both algorithms use the same **ab** homodimers for formation of the nearly correct tetramer. The **ab/cd** tetramer-forming algorithm performs better than the **ab/ac** algorithm for the 10 subunits in the less symmetric tetramers. Thus, the ranks of the nearly correct tetramers are higher for 9 of the 10 cases and unchanged in one case. Of particular interest is the subunit 1hdc/d, for which the **ab/ac** algorithm does not form a nearly correct tetramer (within the RMSD limit specified in Methods) and the **ab/cd** algorithm ranks it first. In addition, the **ab/cd** tetramer-forming algorithm was applied to 8 subunits for which the nearly correct solution obtained with the **ab/ac** algorithm was ranked first. This series of computations showed that the predictions obtained with the **ab/cd** algorithm are as good as those obtained with the **ab/ac** algorithm, using the same **ab** dimers (data not shown). The success of the **ab/cd** algorithm can be attributed to it being less demanding; it forms tetramers on the basis of only one homodimer, whereas the **ab/ac** algorithm requires identification of both the **ab** and the **ac** dimers in the docking step. However, as mentioned above, the **ab/cd** algorithm, which uses two docking steps, can only be applied to a limited set of top ranking dimers. In the reassembling tests it was applied to 150 top-ranking homodimers.

Evaluation of the Tetramer-reassembling Results and Comparison to Geometric Docking of Soluble Proteins

The scores obtained in the present study for the **ab** homodimers are considerably larger than the scores obtained for heterodimers, when similar grid intervals are used in the MolFit scan.[38] Thus, the scores and Z values for the **ab** dimers in the 4 perfect tetramers are 871–3378 and 11.9–70σ, respectively (using the L1 lists). The corresponding values in "bound docking" of soluble proteins are 387–839 and 4.4–17.7σ, respectively.[38] Similarly, the scores and Z values for the **ab** dimers in the 12 systems with a dimer or tetramer in the asymmetric unit (401–2182 and 2.2–36.2σ) are larger than the scores and Z

values obtained in unbound docking of soluble proteins (397–688 and 1.2–8.2σ). The comparison suggests that the **ab** dimer interfaces are larger than the interfaces in the heterodimers of soluble proteins studied by us previously and statistically more unique. The larger interface areas in oligomeric proteins were previously noted by Janin et al.[41] The scores for the **ac** dimers are lower than for the **ab** dimers (563–986 in the 4 perfect tetramers and 381–914 for the other tetramers) and closer to the values obtained for soluble proteins. The high *Z* values in Table II, for the **ab** and **ac** dimers, reflect the shape of the distribution of scores in the docking scans, which often have a very long tail at the high-scores end. Notably, the distribution of scores for the symmetric dimers is very similar to the distribution observed for all the orientations in a complete scan with an angular step of 12°. Hence, the symmetric homodimers are an evenly distributed subset of the full scan.

The independent subunits in systems that crystallize as dimers or tetramers in the asymmetric unit exhibit small structural differences with RMSDs, calculated for all the common atoms, ranging from 0.1 Å for subunits **a** and **b** in system 1ykf, to 1.9 Å for subunits **a** and **d** in 1rhp (the C-termini in 1ado, which are not involved in inter-subunit contacts, were excluded from these calculations). The structural differences between the subunits in the tetramers, although smaller than the differences between bound and unbound structures in the soluble proteins that we studied[38] (0.6–2.5 Å) are not negligible. Nevertheless, the ranks of the predicted nearly correct tetramers are very high compared with the ranks obtained in docking of soluble proteins, especially when unbound docking results are compared. The success of the predictions in the current study is probably related to the use of symmetry constraints in the docking search and in the tetramer forming algorithms. Interestingly, one cannot relate the success or accuracy of the predictions in this study to the structural differences between the subunits in a straightforward manner. For example, in the system 1bq4 a nearly correct solution is ranked first for subunits a, b, and c, whereas for subunit d its rank is considerably lower. Calculation of the RMSDs for the interface residues indicates that the difference between subunit d and any other subunit (0.7–0.9 Å) is not larger than the differences among subunits a, b, and c (0.7–1.0 Å).

Clustering of Solutions

Several top ranking predictions for concanavalin-A form a cluster of nearly correct tetramers, in which one dimer is slightly rotated with respect to the other dimer about a tetramer symmetry axis. We tested these results in more detail by calculating the complementarity scores for small relative rotations of the dimers in the perfect concanavalin-A tetramer 2cna.[42] The angle between the inter-subunit vectors in the **ab** and **cd** dimers in the highest scoring tetramer for 2cna is 76° (ϵ in Fig. 1). Rotation of the dimers by 2° away from this position reduces the score by 1%, rotation of 3° reduces it by 10% (1.5σ) and rotations by 4° lowers the score by 17%. Yet, it is 4.2σ higher than the mean score and 0.2σ higher than the highest scoring false solution. Hence, our docking predicts that small dimer-dimer rotations of 4° are likely to occur in concanavalin-A.

We further pursued this result by analyzing the 23 experimental structures of concanavalin-A deposited in the PDB. We superposed all 23 structures and calculated the angle ϵ. This analysis revealed a wide range for ϵ, from 71° to 82°, with one exception of 104°. Hence, our predicted range of angles, 76 ± 4°, embraces practically all the experimental values.

The dimer-dimer rotation in lectins, of which concanavalin-A is a member, is discussed by Goodsell and Olson.[7] These authors describe similar phenomena in other tetramers, two of which are of interest to us because they are allosterically regulated through rigid-body rotation of the dimers about a perpendicular twofold axis. The two tetramers are phosphofructokinase, in which a rotation of 7° relates the relaxed (R) and the tense (T) states,[43] and fructose-1,6-bisphosphatase, in which the rotation is 19°.[44] We find that the complementarity score for phosphofructokinase (6pfk) drops to 80% of the maximum when the **cd** dimer is rotated by 4° with respect to the **ab** dimer. The corresponding dependence for fructose-1,6-bisphosphatase (1bk4) is considerably less steep, allowing a rotation of 8°. The predicted ranges for phosphofructokinase and fructose-1,6-bisphosphatase, match semi-quantitatively the observed structural differences.

It appears that the clusters of native like solutions for concanavalin-A, phosphofructokinase, and fructose-1,6-bisphosphatase represent a structural soft-mode, i.e., a relatively easy distortion in the average structure of the tetramer, which corresponds to real relative movements of the more stable dimers in the tetramer. Such movements are important for the activity and regulation of the protein. Our procedures predict small clusters of nearly correct tetramers for other tetramers, which have two or more PDB structures and which do not display different dimer-dimer angles (1xva, 1bq4, and 1a7k). However, the different structures of these tetramers occur in the same crystallographic space groups (unlike 2cna, 6pfk, and 1bk4) and internal dimer-dimer motions, if they exist, are not expressed.

Prediction of the Structures of D_2 Tetramers from Model Structures of Single Subunits

The success of the tetramer-forming algorithm in reassembling D_2 tetramers led us to the next step—combining comparative modeling with tetramer formation. We searched the PDB for pairs of homologous proteins, which have different oligomerization modes, so that one of them cannot be used as a comparative modeling template for the other. It can, however, be used as a template for modeling one subunit of the other protein, which is then used as a starting structure in our algorithm. As the structures of both, the template and the modeled oligomers are known our predictions can be validated.

Three examples were tested: We first modeled one subunit of the enzyme *cis*-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from *Pseudomonas* sp[33] (PDB code 1bdb).

TABLE V. Prediction of the Structures of D_2 Tetramers from Modeled Subunits

		Modeled subunit[a]			Modeled D_2 tetramer			
					ab/ac algorithm			**ab/cd** algorithm
Modeled Protein	Modeling template	Sequence identity (%) Overall/**ab/ac**	No. of inserts/ deletions Overall/**ab/ac**	RMSD (Å) (all atoms) Overall/**ab/ac**	Rank/No. of tetramers	**ab** dimer Rank/Z	**ac** dimer Rank/Z	Rank/No. of tetramers
1bdb	1ae1 (tetramer)	31/33/20	8/1/1	3.5/1.9/2.5	23/737	94/4.6	615/2.5	3/89
1ar5	1vew (dimer)	41/58/28	4/0/3	2.4/2.0/3.4	132/792	144/4.2	1624/1.2	Not found
united list					78/12099			21/600
1nhk	1b99 (hexamer)	46/40/30	1/1/0[b]	2.0/2.6/1.9	104/3842	20/7.1	2968/0.5	193/473

[a]The columns below list sequence identity, the number of inserts and deletions in the modeled sequence and the RMSD of the modeled structure compared to the experimental subunit A (R for 1 nhk) for the whole subunit, the **ab** interface and the **ac** interface, respectively. The **ab** and **ac** interfaces in this table correspond to the A-C and A-D interfaces in the experimental structure of 1bdb; the A-B and A-D interfaces in 1ar5 and the R-L and R-A interfaces in 1nhk.
[b]The truncated C-terminus (see text) is counted as a deletion at the **ab** interface.

This enzyme is a biological tetramer with low sequence similarity to tropinone reductase-I from *D. stramonium*[35] (PDB code 1ae1), which is a biological dimer. Inspection of the structure of 1ae1 indicates that a D_2 tetramer is formed via crystallographic symmetry. Hence, in this case, a tetramer of 1bdb can also be formed by comparative modeling with 1ae1 as a template, assuming that the crystallographic dimer-dimer interface in the latter is biologically relevant to 1bdb. Next, we modeled the enzyme Fe-Mn superoxide dismutase from *P. shermanii*[15] (PDB code 1ar5), which acts and crystallizes as a tetramer. It is homologous to the Mn superoxide dismutase in *E. coli*[16] (PDB code 1vew), which acts and crystallizes as a dimer. In this case a new interface has to be predicted. Finally, we modeled the D_2 tetrameric enzyme nucleoside diphosphate kinase from *M. xanthus*[34] (PDB code 1nhk) using the homologous subunit from the hexameric D_3 nucleoside diphosphate kinase from *D. discoideum*[36] (PDB code 1b99) as a template. Interestingly, all known eukaryotic nucleoside diphosphate kinases are hexamers, whereas the homologous prokaryotic kinases are either tetramers or hexamers. Similar C_2 dimers form both the D_2 and D_3 oligomers,[45] but in this study we form the D_2 tetramer from one modeled subunit. In the predictions based on modeled subunits we used the same error range parameters as described in Methods, except for 1nhk for which a slightly larger deviations from perpendicularity ($\pm 6°$) was allowed. The **ab/cd** algorithm was applied to 750 top ranking homodimers (instead of 150), because the rank of a nearly correct homodimer in docking of modeled subunits is likely to be low. Table V summarizes the comparative modeling and docking results for the three tests.

The sequence similarity between 1bdb and 1ae1 is low, only 31% identity, but it extends evenly along the whole sequence, excluding the 15 C-terminal residues in 1bdb. The latter were not modeled because 1ae1 cannot provide a modeling template for this segment. Another segment (amino acids 199–207) was not modeled because the corresponding loop in the template protein is much shorter. Notably, this segment is not resolved in the X-ray structure of 1bdb. Figure 3(a) presents a superposition of the model onto subunit A in the experimental structure of 1bdb. The RMSD between the common Cα atoms in the two structures is 2.9 Å and for all the atoms it is 3.5 Å (see more details in Table V). Next, the modeled subunit was docked against itself and the L5 MolFit list was used as input to the **ab/ac** and the **ab/cd** tetramer forming algorithms. The **ab/ac** algorithm formed a nearly correct tetramer (RMSD of 4.5 Å) ranked 23. The **ab/cd** algorithm was even more successful, ranking a similar nearly correct tetramer 3rd.

It appears that MolFit successfully identified both the A-C and the A-D interfaces, despite the low sequence similarity between 1bdb and 1ae1, and the resulting considerable structural errors in the modeled subunit of 1bdb. Thus, 19 of the 33 A-C interface residues and 19 of the 40 A-D interface residues are found in the corresponding predicted interfaces. Interestingly, the larger yet less conserved A-D interface in 1bdb corresponds to the interface in the biological dimer of 1ae1 (the dimer in the asymmetric unit). However, MolFit ranks a dimer similar to the experimental A-C dimer higher than the dimer that corresponds to the A-D dimer. The excellent prediction obtained with the **ab/cd** algorithm is based on the A-C like dimer.

In our second attempt a model of one subunit of the tetramer 1ar5 was constructed from the structure of one subunit of the dimer 1vew. The sequence identity between 1ar5 and 1vew is 41% (see Table V). The more conserved A-B interface of 1ar5 (58% identity) corresponds to the dimer interface in 1vew. The A-D interface, which has to be predicted, is considerably less conserved (28% identity) and it includes two large deletions in the 1ar5 sequence and one insert, making the modeling of this interface and of the whole tetramer difficult. The modeled subunit of 1ar5 is compared to subunit A in the experimental structure in Figure 3(b). The RMSD between the common Cα atoms in the two structures is 1.6 Å and for all the atoms it is 2.4 Å. However, the A-D interface deviates considerably more from the experimental structure, as reflected in the RMSD values in Table V.

The **ab/ac** tetramer-forming algorithm succeeded to construct a nearly correct tetramer for 1ar5 ranked 132, despite the inaccurate model structure of the single subunit. The RMSD for this tetramer (all common Cα atoms) is 4.5 Å and we find that 14 of the 19 A-B interface residues

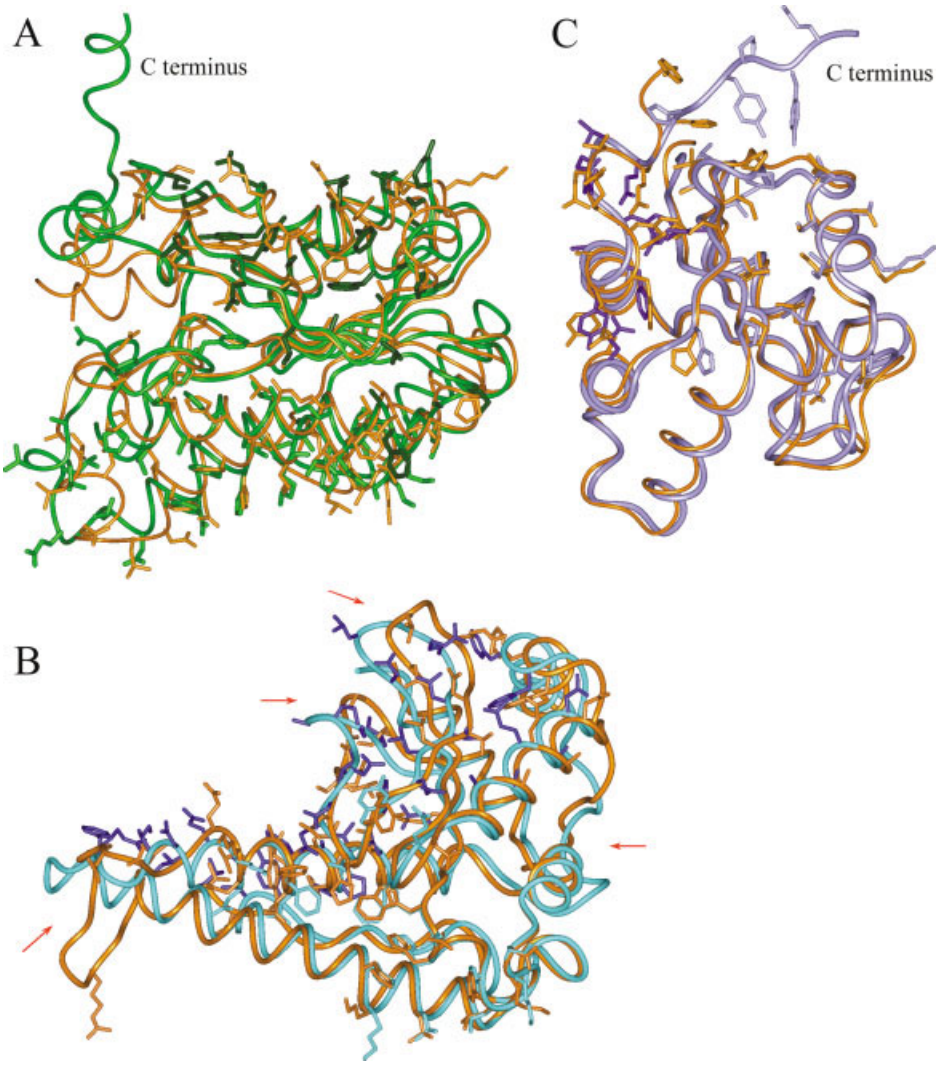

Fig. 3. Superposition of the modeled and experimental structures of subunits of 1bdb, 1ar5, and 1nhk. The models are colored in orange showing a ribbon diagram and the side chains of residues at the subunit-subunit interface. (A) The experimental structure of subunit A of 1bdb is shown in green with the side chains at the AC and AD interface in dark green and green, respectively. This model is based on the structure of one subunit in 1ae1. (B) The experimental structure of subunit A in 1ar5 is shown in cyan with the side chains of residues at the AB and AD interfaces in cyan and blue, respectively. The model is based on the structure of one subunit in 1vew. Loops with insert/deletion are indicated by the red arrows (C). The experimental structure of subunit R in 1nhk is shown in light violet with the side chains of residues at the RL and RA interface in light and dark violet, respectively. The model is base on the structure of one subunit in 1b99.

and 15 of the 46 A-D interface residues are predicted to be in the corresponding interface. In contrast to the former modeling test (1bdb), the **ab/cd** tetramer-forming algorithm failed to form a nearly correct tetramer for 1ar5. We tested in this case the idea suggested above of combining lists of homodimers formed by docking different conformers of the same subunit. To this end we docked each of four different conformers of the modeled subunit 1ar5 against itself. The different conformers were formed by omitting one of the four loops with insert/deletion (residues 49–57, 83–93, 133–135, and 150–154) irrespectively of it being at the interface or on the non-interacting surface [see Fig. 3(b)]. The united list of homodimers, sorted by the Z values of the solutions, was then used as input to the tetramer-forming algorithms. This procedure produced significantly better results: Hence, with the **ab/ac** algorithm a nearly correct tetramer is now ranked 78 (out of 12,099 tetramers formed from this list). This solution was formed by combining dimers with different conformations. The **ab/cd** algorithm ranked a nearly correct tetramer 21. This algorithm uses only one homodimer hence the latter result indicates that omission of one loop allowed identification of a reasonably accurate dimer-dimer interface.

Our third test was the construction of a D_2 tetramer of 1nhk based on one subunit from the hexameric 1b99. The overall sequence similarity is low, 46% identity, and it does not include the four C-terminal amino acids, which were therefore omitted from the model. One interface (the R-L

interface in 1nhk) is structurally similar in the two oligomers, although the sequence similarity is low (40% identity). The other interface (R-A interface in 1nhk) is different and it is also less conserved (30% identity). The predicted nearly correct tetramer obtained with the **ab/ac** tetramer-forming algorithm is ranked 104 (RMSD of 3.7 Å for all the common Cα atoms). We find that 13 of the 20 R-L interface residues are also at the interface in our prediction and so are 9 of the 10 residues in the R-A interface. Interestingly this was the only case in which we needed to slightly enlarge the error range, from 90 ± 3° to 90 ± 6°, in order to obtain an acceptable prediction. The **ab/cd** algorithm forms a nearly correct tetramer (RMSD of 2.5 Å) ranked 193.

The three modeling attempts show several interesting features: First, nearly correct tetramers are formed readily by the **ab/ac** tetramer-forming algorithm and their ranks are very high when compared with the results of unbound docking of soluble proteins.[38] This most likely results form the imposition of symmetry constraints in both the docking step and the tetramer formation step. Second, the **ab/ac** tetramer-forming algorithm forms a nearly correct solution in every case whereas the **ab/cd** algorithm fails for 1ar5. Initially we expected that the less demanding **ab/cd** algorithm, which requires the identification of only one dimer in the docking step, would form tetramers more readily than the **ab/ac** algorithm. It appears, however, that the ability of our docking algorithm MolFit to form nearly correct C_2 dimers from model structures, although impaired (low ranks of the dimers as shown in Table V), is not abolished. The **ab/ac** tetramer-forming algorithm can easily "fish" the appropriate low-ranking dimers and combine them to form tetramers.

CONCLUSIONS

We demonstrate in this study that comparative modeling in combination with a tetramer-forming algorithm, in which symmetry constraints are incorporated, can be used to predict the structure of a D_2 tetramer from the model structure of a single subunit. The predicted tetramers are similar to the experimental structure and they provide information regarding the intersubunit interactions. The algorithm that we present is also useful for analyzing D_2 tetramers. Thus, the observed clusters of nearly correct tetramers often reflect genuine structural distortions related to the protein activity and regulation.

ACKNOWLEDGMENTS

We thank Prof. Ephraim Katchalski-Katzir for his support and advice throughout this study.

REFERENCES

1. Argos P. An investigation of protein subunit and domain interfaces. Protein Eng 1988;2:101–113.
2. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
3. Stites WE. Protein-protein interactions: Interface structure, binding thermodynamics, and mutational analysis. Chem Rev 1997;97: 1233–1250.
4. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Protein-protein interfaces: architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences. Crit Rev Biochem Mol Biol 1996;31:127–152.
5. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng 1997;10:999–1012.
6. Conte LL, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
7. Goodsell DS, Olson AJ. Structural symmetry and protein function. Annu Rev Biophys Biomol Struct 2000;29:105–153.
8. Eisenstein M, Shariv I, Koren G, Friesem AA, Katchalski-Katzir E. 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 1997;266:135–43.
9. Butler PJ, Klug A. The assembly of a virus. Sci Am 1978;239:62–69.
10. Janin J, Wodak SJ. Reaction pathway for the quaternary structure change in hemoglobin. Biopolymers 1985;24:509–526.
11. Janin J, Wodak SJ. The quaternary structure of carbonmonoxy hemoglobin ypsilanti. Proteins 1993;15:1–4.
12. Harris A, Forouhar F, Qiu S, Sha B, Luo M. The crystal structure of the influenza matrix protein M1 at neutral pH: M1-M1 protein interfaces can rotate in the oligomeric structures of M1. Virology 2001;289:34–44.
13. Gardiner EJ, Willett P, Artymiuk PJ. Protein docking using a genetic algorithm. Proteins 2001;44:44–56.
14. Berman HM, Westbrook J, Feng Z, Gilliland, G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
15. Schmidt M, Meier B, Parak F. X-ray structure of the cambialistic superoxide dismutase from Propionibacterium shermanii active with Fe or Mn. J Biol Inorg Chem 1996;1:532.
16. Edwards RA, Baker HM, Whittaker MM, Whittaker JW, Jameson GB, Baker EN. Crystal structure of Escherichia coli manganese superoxide dismutase at 2.1-angstrom resolution. J Biol Inorg Chem 1998;3:161.
17. Benning MM, Wesenberg G, Liu R, Taylor KL, Dunaway-Mariano D, Holden HM. The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from Pseudomonas sp. strain CBS-3. J Biol Chem 1998;273:33572–33579.
18. Hohenester E, Keller JW, Jansonius JN. An alkali metal ion size-dependent switch in the active site structure of dialkylglycine decarboxylase. Biochemistry 1994;33:13561–13570.
19. Rafferty JB, Simon JW, Baldock C, Artymiuk PJ, Baker PJ, Stuitje AR, Slabas AR, Rice DW. Common themes in redox chemistry emerge from the X-ray structure of oilseed rape (Brassica napus) enoyl acyl carrier protein reductase. Structure 1995;3: 927–938.
20. Weeks CM, Roszak AW, Erman M, Kaiser R, Jornvall H, Ghosh D. Structure of rabbit liver fructose 1,6-bisphosphatase at 2.3 A resolution. Acta Crystallogr D Biol Crystallogr 1999;55:93–102.
21. Harrop SJ, Helliwell JR, Wan TCM, Kalb AJ, Tong L, Yariv J. Structure solution of a cubic crystal of concanavalin a complexed with methyl alpha-D-glucopyranoside. Acta Crystallogr D 1996;52: 143.
22. Fu Z, Hu Y, Konishi K, Takata Y, Ogawa H, Gomi T, Fujioka M, Takusagawa F. Crystal structure of glycine *N*-methyltransferase from rat liver. Biochemistry 1996;35:11985–11993.
23. Ermler U, Merckel M, Thauer R, Shima S. Formylmethanofuran: tetrahydromethanopterin formyltransferase from Methanopyrus kandleri—new insights into salt-dependence and thermostability. Structure 1997;5:635–646.
24. Rigden DJ, Walter RA, Phillips SE, Fothergill-Gilmore LA. Polyanionic inhibitors of phosphoglycerate mutase: combined structural and biochemical analysis. J Mol Biol 1999;289:691–699.
25. Kim H, Hol WG. Crystal structure of Leishmania mexicana glycosomal glyceraldehyde-3-phosphate dehydrogenase in a new crystal form confirms the putative physiological active site structure. J Mol Biol 1998;278:5–11.
26. Singleton M, Isupov M, Littlechild J. X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon Thermococcus litoralis. Structure Fold Des 1999;7:237–244.
27. Blom N, Sygusch J. Product binding and role of the C-terminal region in class I D-fructose 1,6-bisphosphate aldolase. Nat Struct Biol 1997;4:36–39.
28. Schirmer T, Evans PR. Structural basis of the allosteric behaviour of phosphofructokinase. Nature 1990;343:140–145.
29. Zhang X, Chen L, Bancroft DP, Lai CK, Maione TE. Crystal

structure of recombinant human platelet factor 4. Biochemistry 1994;33:8361–8366.
30. Korkhin Y, Kalb AJ, Peretz M, Bogin O, Burstein Y, Frolow F. NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of Clostridium beijerinckii and Thermoanaerobacter brockii. J Mol Biol 1998;278:967–981.
31. Ghosh D, Erman M, Wawrzak Z, Duax WL, Pangborn W. Mechanism of inhibition of 3 alpha, 20 beta-hydroxysteroid dehydrogenase by a licorice-derived steroidal inhibitor. Structure 1994;2:973–980.
32. Palm GJ, Lubkowski J, Derst C, Schleper S, Rohm KH, Wlodawer A. A covalently bound catalytic intermediate in Escherichia coli asparaginase: crystal structure of a Thr-89-Val mutant. FEBS Lett 1996;390:211–216.
33. Hulsmeyer M, Hecht HJ, Niefind K, Hofer B, Eltis LD, Timmis KN, Schomburg D. Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from a PCB degrader at 2.0 A resolution. Protein Sci 1998;7:1286–1293.
34. Munoz Dorado J, Arias JM. Protein kinases and phosphatases during the developmental cycle of myxobacteria. Microbiologia 1995;11:376–378.
35. Nakajima K, Yamashita A, Akama H, Nakatsu T, Kato H, Hashimoto T, Oda J, Yamada Y. Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold. Proc Natl Acad Sci USA 1998;95:4876–4881.
36. Gonin P, Xu Y, Milon L, Dabernat S, Morr M, Kumar R, Lacombe ML, Janin J, Lascu I. Catalytic mechanism of nucleoside diphosphate kinase investigated using nucleotide analogues, viscosity effects, and X-ray crystallography. Biochemistry 1999;38:7265–7272.
37. 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 USA 1992;89:2195–2199.
38. Heifetz A, Katchalski-Katzir E, Eisenstein M. Electrostatics in protein-protein docking. Protein Sci 2002;11:571–587.
39. Levitt M, Gerstein M. A unified statistical framework for sequence comparison and structure comparison. Proc Natl Acad Sci USA 1998;95:5913–5920.
40. Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 1990;183:63–98.
41. Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. J Mol Biol 1988;204:155–164.
42. Reeke GN Jr, Becker JW, Edelman GM. The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. J Biol Chem 1975;250:1525–1547.
43. Evans PR. Structural aspects of allostery. Curr Opin Struct Biol 1991;1:773–779.
44. Lipscomb WN. Structure and function of allosteric enzymes. CHEMTRACTS Biochem Mol Biol 1991;2:1–15.
45. Janin J, Dumas C, Morera S, Xu Y, Meyer P, Chiadmi M, Cherfils J. Three-dimensional structure of nucleoside diphosphate kinase. J Bioenerg Biomembr 2000;32:215–225.