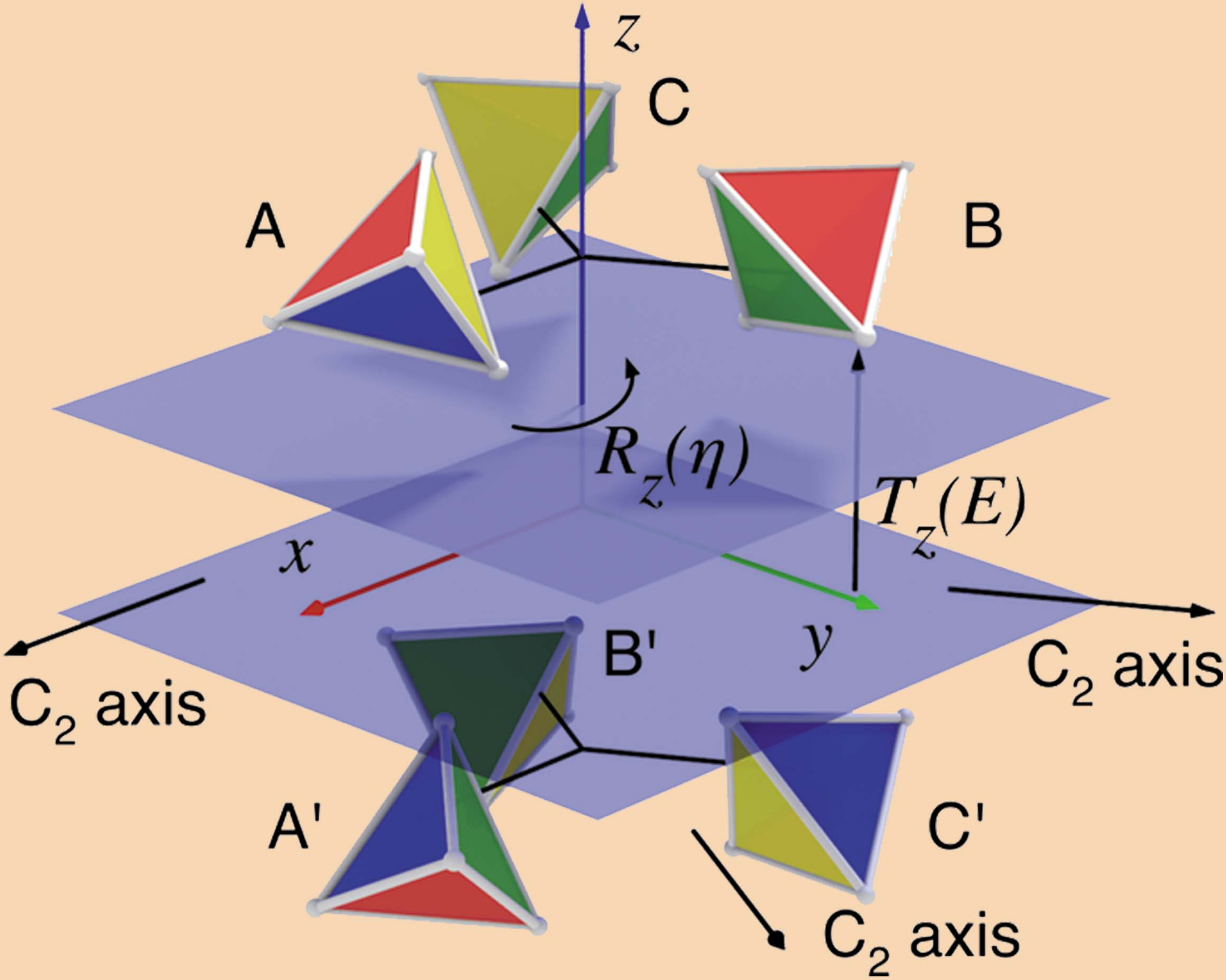
Journal of APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Volume 49 | Part 1 | February 2016



IUCr Journals | Wiley

JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Spherical polar Fourier assembly of protein complexes with arbitrary point group symmetry

David W. Ritchie[a]* and Sergei Grudinin[b,c,d]

[a]Inria Nancy – Grand Est, 615 Rue du Jardin Botanique, 54600, Villers-les-Nancy, France, [b]University of Grenoble Alpes, LJK, F-38000 Grenoble, France, [c]CNRS, LJK, F-38000 Grenoble, France, and [d]Inria Grenoble, France. *Correspondence e-mail: dave.ritchie@inria.fr

Received 19 September 2015
Accepted 30 November 2015

Edited by D. I. Svergun, European Molecular Biology Laboratory, Hamburg, Germany

Keywords: protein docking; polar Fourier correlation; point group symmetry; symmetry docking.

A novel fast Fourier transform-based *ab inito* docking algorithm called *SAM* is presented, for building perfectly symmetrical models of protein complexes with arbitrary point group symmetry. The basic approach uses a novel and very fast one-dimensional symmetry-constrained spherical polar Fourier search to assemble cyclic C_n systems from a given protein monomer. Structures with higher-order (D_n, T, O and I) point group symmetries may be built using a subsequent symmetry-constrained Fourier domain search to assemble trimeric sub-units. The results reported here show that the *SAM* algorithm can correctly assemble monomers of up to around 500 residues to produce a near-native complex structure with the given point group symmetry in 17 out of 18 test cases. The *SAM* program may be downloaded for academic use at http://sam.loria.fr/.

1. Introduction

Many protein complexes in the Protein Data Bank (PDB; Berman *et al.*, 2002; http://www.rcsb.org/) are symmetric homo-oligomers (Levy *et al.*, 2006). Indeed, it appears that large symmetrical protein structures have evolved in many organisms because they carry specific morphological and functional advantages compared to small individual protein molecules (Goodsell & Olsen, 2000; Levy *et al.*, 2008). There is therefore considerable interest in studying and modelling the structures of these large bio-molecular complexes. Although many symmetrical complexes have been solved by X-ray crystallography and cryo-electron microscopy, this can often be a difficult and time-consuming process, and it would be useful to be able to generate high-quality candidate complex structures for use as templates in molecular replacement (MR) techniques (Rossmann, 1990; Navaza, 2001), to provide angular parameters for locked MR search functions (Tong, 2001) or to dock high-resolution structural models into low-resolution cryo-EM density maps (Roseman, 2000), for example. From a protein design point of view, it would also be very useful to be able to predict computationally whether or not a given monomer might self-assemble into a symmetrical structure (Huang *et al.*, 2005).

In the past few years, several *ab initio* protein–protein docking programs, such as *MolFit* (Berchanski & Eisenstein, 2003), *ClusPro* (Comeau & Camacho, 2004), *M-Zdock* (Pierce *et al.*, 2005) and *SymmDock* (Schneidman-Duhovny *et al.*, 2005), have been adapted to apply various geometric filtering constraints to extract approximately symmetrical pair-wise docking orientations. Symmetry-constraint protocols may be applied to refine the coordinates of a given symmetric structure using *RosettaDock* (André *et al.*, 2007). The *Haddock* docking engine allows up to six distance restraints to be

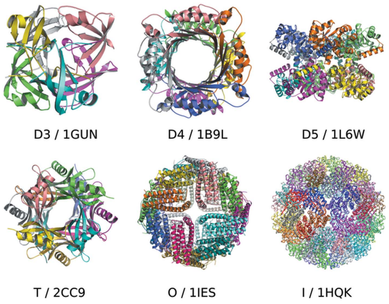


Table 1
The number of C_n and D_n complexes in the 3D-Complex database.

3D-Complex also reports 86 tetrahedral, 47 octahedral and six icosahedral complexes (the 3D-Complex database excludes all viral structures).

n	2	3	4	5	6	7	8	11
C_n	8740	992	223	107	76	29	5	10
D_n	2111	585	173	46	20	23	6	0

defined when refining oligomeric complexes with certain cyclic or dihedral symmetries (Karaca *et al.*, 2010). However, to our knowledge, there does not yet exist an *ab initio* docking algorithm which can automatically generate perfectly symmetrical protein complexes for arbitrary point group symmetry types.

We previously described the 'Hex' polar Fourier correlation method for rigid-body protein docking. Unlike conventional three-dimensional Cartesian grid-based fast Fourier transform (FFT) docking algorithms, Hex uses a spherical polar Fourier (SPF) representation which favours rotational correlation searches. Here, we present several developments of the SPF correlation approach, which allow protein complexes with point group symmetry to be assembled very rapidly by using a series of one-dimensional FFTs. We call this approach 'SAM', being short for 'symmetry assembler'.

In order to build symmetrical protein complexes, it is necessary to locate a certain number of protein monomers in orientations that satisfy the symmetry elements of a given point group. We are mainly concerned with cyclic (C_n) and dihedral (D_n) point groups, but it is desirable to have a general method which can build complexes with tetrahedral (T), octahedral (O) and icosahedral (I) symmetries as well. Table 1 summarizes the number of symmetric complexes reported by the 3D-Complex database (Levy *et al.*, 2006). This shows that C_2 homo-dimers consitute the majority of known homo-oligomers. However, many complexes have higher-order rotational symmetry (*i.e.* $C_{n>2}$), and a good number have multiple rotational symmetry axes, namely those with D_n, T, O and I point group symmetries. Thus, starting from a given protein monomer, the overall aim of the present work is to develop and implement exact and efficient SPF docking expressions to generate candidate protein complexes having one of the naturally occurring C_n, D_n, T, O and I symmetries.

2. Methods

2.1. Coordinate operators and SPF representations

In order to generate symmetry-related docking configurations, it is convenient to start with a single protein monomer at the origin, and to generate and score copies of the monomer in different candidate symmetry-related orientations. To generate orientations in a systematic way, we introduce the notion of a translation and a rotation coordinate operator, $\hat{T}(x,y,z)$ and $\hat{R}(\alpha,\beta,\gamma)$, respectively. These represent the actions of translating an object by an amount $\underline{x}=(x,y,z)$ and rotating the object about the origin using the Euler rotation angles (α,β,γ). Here, we use the Euler 'zyz' convention, in which a general rotation can be expressed as three consecutive rotations about the z and y axes:

$$\hat{R}(\alpha,\beta,\gamma)=\hat{R}_z(\alpha)\hat{R}_y(\beta)\hat{R}_z(\gamma), \qquad (1)$$

where $\hat{R}_z(\gamma)$ is applied first. Normally, the Euler rotation angles are restricted to the ranges $(0\le\alpha<2\pi,\ 0\le\beta<\pi,\ 0\le\gamma<2\pi)$, but we show below that one of these ranges should be reduced in the presence of symmetry.

In order to develop the equations necessary for a docking search, it is useful to introduce a 'docking operator', ↔, such that the notation

$$A(\underline{r})\leftrightarrow B(\underline{r}) \qquad (2)$$

is taken to mean a docking interaction between proteins A and B. In the present work, the functions $A(\underline{r})$ and $B(\underline{r})$ represent three-dimensional shape-density functions of the two proteins, while $\underline{r}$ represents a polar coordinate in three-dimensional space, $\underline{r}=(r,\theta,\varphi)\equiv(x,y,z)$. Following the original Hex docking algorithm, $A(\underline{r})$ and $B(\underline{r})$ consist of linear combinations of three-dimensional interior and surface skin density functions (Ritchie & Kemp, 2000). Thus, by introducing a scale factor, K, with units of kJ mol^{-1}, a three-dimensional overlap integral of the form

$$S=K\int A(\underline{r})^{*}B(\underline{r})\,\mathrm{d}\underline{r} \qquad (3)$$

may be treated as a shape-based docking score or pseudo interaction energy [the asterisk denotes the complex conjugation of $A(\underline{r})$]. While the functions $A(\underline{r})$ and $B(\underline{r})$ are initially entirely real, adopting the convention of conjugating one of these functions in the above overlap expression ensures that the docking score (taken as the real part of S) remains meaningful with complex functions. Indeed, by treating $A(\underline{r})$ and $B(\underline{r})$ as complex quantities, it is possible to accelerate the search over multiple candidate docking orientations using FFT techniques (Ritchie & Kemp, 2000; Ritchie *et al.*, 2008). In the subsequent analysis, we will drop the scale factor K and we will use only the symbols $A(\underline{r})$ and $B(\underline{r})$ instead of the actual linear combinations for the sake of clarity.

In the rigid-body docking problem where the relative orientations of A and B are unknown, we adopt the convention that the centres of mass of proteins A and B are initially located at the origin, and we let the expression

$$A(\underline{r})\leftrightarrow\hat{T}(x,y,z)\hat{R}(\alpha,\beta,\gamma)B(\underline{r}) \qquad (4)$$

represent a general interaction between protein A and a rotated and translated version of protein B. Consequently, the aim is to find the six parameters $(x,y,z,\alpha,\beta,\gamma)$ that give the most favourable interaction.

Note that the symbol ↔ can be treated like an equality in the sense that applying an inverse translation to each side of equation (4),

$$\hat{T}(x,y,z)^{-1}A(\underline{r})\leftrightarrow\hat{R}(\alpha,\beta,\gamma)B(\underline{r}), \qquad (5)$$

represents exactly the same relative orientation of the two protein monomers as in the previous expression. In either case, the corresponding pair-wise docking score, S, would be calculated as a three-dimensional overlap integral of the form

$$S = \textstyle\int A(\underline{r})^* \left[\hat{T}(x, y, z)\hat{R}(\alpha, \beta, \gamma)B(\underline{r})\right]\mathrm{d}\underline{r}. \tag{6}$$

Here, we represent protein shapes as SPF expansions of complex spherical harmonic, $Y_{lm}(\theta, \varphi)$, and Gauss–Laguerre, $R_{nl}(r)$, basis functions:

$$A(\underline{r}) = \sum_{nlm} A_{nlm} R_{nl}(r) Y_{lm}(\theta, \varphi), \tag{7}$$

where A_{nlm} are complex expansion coefficients (see Appendix *B*). Nonetheless, when working in the SPF domain, it is often more efficient to calculate one side of a given 'docking equation' than the other. Thus, it is important to consider the most efficient order of operators for a given symmetry type.

2.2. Cyclic C_n complexes

With SPF basis functions, rotations and translations of SPF representations are most easily implemented with respect to the z axis. Hence, it is convenient to associate the z axis with the main (one-dimensional FFT) rotational and translational degrees of freedom (DOFs) and to associate the y axis with the principal rotational symmetry axis.

Because an individual protein monomer is asymmetric, we normally have to assume that it can take any orientation in space relative to a set of fixed coordinate axes. Thus, describing a particular orientation of a given monomer, A, with respect to a random starting orientation will absorb three rotational DOFs. Let us suppose that the operator associated with that description is $\hat{R}(\alpha, \beta, \gamma)$. If we then copy the rotated A into an equally rotated monomer B, we can describe the docking interaction between a pair of C_n symmetry mates by applying the following transformations:

$$\hat{R}_y(\omega_{j+1})\hat{T}_z(D)\hat{R}(\alpha, \beta, \gamma)B(\underline{r}) \leftrightarrow \hat{R}_y(\omega_j)\hat{T}_z(D)\hat{R}(\alpha, \beta, \gamma)A(\underline{r}), \tag{8}$$

where the angles $\omega_j = 2\pi j/n$ are rotations around the principal symmetry axis. This equation highlights the fact that there exist only four degrees of freedom $(D, \alpha, \beta, \gamma)$ between the monomers in a complex with C_n symmetry. It is shown in Appendix *A* that the range of the α rotation angle must be restricted to $0 \le \alpha < \pi$.

For a symmetric dimer or trimer, the above pair-wise A ↔ B interaction is the only interaction that needs to be calculated. For $C_{n>3}$, there may also exist additional higher-order [*i.e.* $1 \leftrightarrow 3, \ldots, 1 \leftrightarrow (n/2 + 1)$] interactions which should in principle be taken into account. However, these are likely to be small or negligible in most cases, and are ignored in the current work.

Nonetheless, a weakness of the above approach is that when n becomes large it becomes necessary to translate each monomer far from the origin in order to achieve the desired separation between consecutive pairs of monomers. Such large translations can seriously reduce the resolution of the shape-density representations because of the exponential fall-off in the SPF radial basis functions. Therefore, in order to have expressions which involve only small translations, it is desirable to perform the SPF docking search near the origin and to transform only the top solutions back to the symmetry frame. Fig. 1 describes the problem graphically.

Thus, with the aid of Fig. 1, it is preferable to begin instead with

$$\hat{T}_z(S)\hat{R}(\alpha, \beta, \gamma)A(\underline{r}) \leftrightarrow \hat{R}_y(\omega)\hat{T}_z(S)\hat{R}(\alpha, \beta, \gamma)B(\underline{r}), \tag{9}$$

where $\omega = 2\pi/n$ and $S = D/[2\sin(\omega/2)]$. To calculate this equation with A at the origin, we apply $\hat{T}_z(S)^{-1}$ to each side to give

$$\hat{R}(\alpha, \beta, \gamma)A(\underline{r}) \leftrightarrow \hat{T}_z(S)^{-1}\hat{R}_y(\omega)\hat{T}_z(S)\hat{R}(\alpha, \beta, \gamma)B(\underline{r}). \tag{10}$$

Then, to locate B on the positive z axis, we apply $R_y(-\psi)$ to each side, where $\psi = \pi/2 + \omega/2$ (see Fig. 1), to obtain

$$\begin{aligned}&\hat{R}_y(-\psi)\hat{R}(\alpha, \beta, \gamma)A(\underline{r})\\ &\quad\leftrightarrow \hat{R}_y(-\psi)\hat{T}_z(S)^{-1}\hat{R}_y(\omega)\hat{T}_z(S)\hat{R}(\alpha, \beta, \gamma)B(\underline{r}).\end{aligned} \tag{11}$$

It can then be shown that

$$\hat{R}_y(-\psi)\hat{T}_z(S)^{-1}\hat{R}_y(\omega)\hat{T}_z(S) = \hat{T}_z(D)\hat{R}_y(\omega)\hat{R}_y(-\psi), \tag{12}$$

where D is the distance between the two monomers. Furthermore, if we assume that we are starting from a random monomer orientation, we can 'bury' the y rotation by putting

$$\hat{R}_y(-\psi)\hat{R}(\alpha, \beta, \gamma) = \hat{R}(\alpha', \beta', \gamma') \tag{13}$$

to give

$$\hat{R}(\alpha', \beta', \gamma')A(\underline{r}) \leftrightarrow \hat{T}_z(D)\hat{R}_y(\omega)\hat{R}(\alpha', \beta', \gamma')B(\underline{r}). \tag{14}$$

As shown below, we can use a one-dimensional FFT search near the origin to determine the parameters $(D, \alpha', \beta', \gamma')$. We can then transform the solution back to the original coordinate frame by applying the operator $\hat{T}_z(S)\hat{R}_y(\psi)$ to each side. In other words, if the FFT search finds solutions $(D, \alpha', \beta', \gamma')_k$, the transformation matrix, $\underline{M}_k^A$, that should be applied to

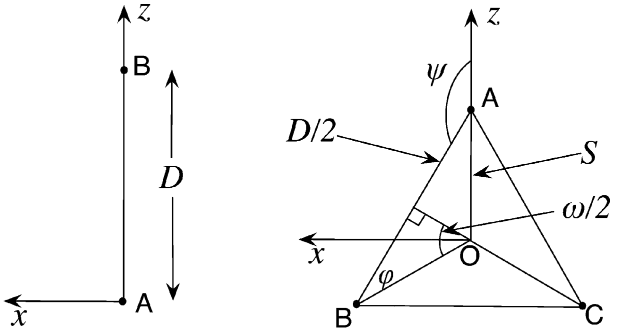


Figure 1
The coordinate systems used for pair-wise docking in C_n. The figure on the left shows the computational coordinate frame for a pair of monomers, A and B, with A at the origin in the xz plane and B at a distance D along the positive z axis. The figure on the right shows the symmetry frame of a C_3 trimer with the monomers arranged about the y axis (which points out of the plane towards the viewer). Here, $\omega = 2\pi/n$ is the C_n symmetry angle. From basic geometry, $S = D/(2\cos\varphi) = D/[2\sin(\omega/2)]$ is the distance from the principal symmetry axis to the centre of each monomer. We also have $\psi = \pi - \varphi = (\pi/2 + \omega/2)$, which defines the rotation that relates the two coordinate systems.

locate the A monomer on the positive z axis for the kth docking solution is given by

$$\underline{M}_k^A = \underline{T}_z(S_k)\underline{R}_y(\pi/2+\omega/2)\underline{R}(\alpha'_k, \beta'_k, \gamma'_k). \tag{15}$$

Similarly, the docked B monomer may be located by applying the matrix

$$\underline{M}_k^B = \underline{T}_z(S_k)\underline{R}_y(\pi/2+\omega/2)\underline{T}_z(D_k)\underline{R}_y(\omega)\underline{R}(\alpha'_k, \beta'_k, \gamma'_k). \tag{16}$$

Because it can be seen that $\underline{M}_k^B = \underline{R}_y(\omega)\underline{M}_k^A$, it follows that all remaining symmetry mates may be generated from the coordinates of the A monomer.

Regarding the actual FFT calculation, by putting

$$A(\underline{r})' = B(\underline{r})' = \hat{R}(0, \beta', \gamma')A(\underline{r}), \tag{17}$$

and by exploiting the fact that $\hat{R}_z(\alpha')$ and $\hat{T}_z(D)$ commute, the docking equation in the computational frame becomes

$$\hat{T}_z(D)^{-1}A(\underline{r})' \leftrightarrow \hat{R}_z(\alpha')^{-1}\hat{R}_y(\omega)\hat{R}_z(\alpha')B(\underline{r})' \tag{18}$$

or more simply

$$A(\underline{r})'' \leftrightarrow \hat{R}_z(\alpha')^{-1}\hat{R}_y(\omega)\hat{R}_z(\alpha')B(\underline{r})'. \tag{19}$$

The Fourier series representation of the A monomer may be rotated and translated using

$$A'_{nlm} = \sum_{m'} D^{(l)}_{mm'}(0, \beta', \gamma')A_{nlm'} \tag{20}$$

and

$$A''_{nlm} = \sum_{kj} T^{|m|}_{nl,kj}(-D)A'_{kjm}, \tag{21}$$

where $D^{(l)}_{mm'}(\alpha, \beta, \gamma)$ are matrix elements of the Wigner rotation matrices for the spherical harmonics (Biedenharn & Louck, 1981) and each $T^{|m|}_{nl,kj}(D)$ is a translation matrix element for the SPF basis functions (Ritchie, 2005). Then, writing the rotations for monomer B in terms of the Wigner rotation matrix elements (Appendix B) gives

$$\hat{R}_z(\alpha')^{-1}\hat{R}_y(\omega)\hat{R}_z(\alpha')B(\underline{r})' = \sum_{nlm}\sum_{rpq} D^{(l)}_{mr}(-\alpha', 0, 0) \times D^{(l)}_{rp}(0, \omega, 0)D^{(l)}_{pq}(\alpha', 0, 0)B'_{nlq}R_{nl}(r)Y_{lm}(\theta, \varphi), \tag{22}$$

and hence

$$\hat{R}_z(\alpha')^{-1}\hat{R}_y(\omega)\hat{R}_z(\alpha')B(\underline{r})' = \sum_{nlmp} \exp[-i(p-m)\alpha'] \times d^{(l)}_{mp}(\omega)B'_{nlp}R_{nl}(r)Y_{lm}(\theta, \varphi). \tag{23}$$

Taking the complex conjugate of $A(\underline{r})''$ and integrating over the product with B then gives an $O(N^4)$ complexity docking score

$$S(\alpha'; \omega, D, \beta', \gamma') = \sum_{nlmp} \exp[-i(p-m)\alpha']\, d^{(l)}_{mp}(\omega)B'_{nlp}A''^{*}_{nlm}. \tag{24}$$

Summing over n and l using

$$C_{mp} = \sum_{nl} d^{(l)}_{mp}(\omega)B'_{nlp}A''^{*}_{nlm} \tag{25}$$

reduces this to

$$S(\alpha'; \omega, D, \beta', \gamma') = \sum_{mp} C_{mp}\exp[-i(p-m)\alpha']. \tag{26}$$

The α' rotation (which here is restricted by symmetry to the range $0 \le \alpha' < \pi$) may be scaled back onto the natural range of the FFT (see Appendix B) by putting $\alpha'' = 2\alpha'$ and writing

$$\exp(-is\alpha') = \sum_t \lambda^{(\pi)}_{st}\exp(-it\alpha'') \tag{27}$$

to obtain

$$S(\alpha'; \omega, D, \beta', \gamma') = \sum_{mpt} C_{mp}\lambda^{(\pi)}_{p-m,t}\exp(-it\alpha''). \tag{28}$$

Finally, summing over m and p as

$$Q_t = \sum_{mp} C_{mp}\lambda^{(\pi)}_{p-m,t} \tag{29}$$

gives a one-dimensional Fourier series in α'':

$$S(\alpha'; \omega, D, \beta', \gamma') = \sum_t Q_t\exp(-it\alpha''). \tag{30}$$

Because we now have a simple complex exponential on the right-hand side, this expression shows that for a given translation D and rotation (β', γ') the pair-wise docking score in an arbitrary C_n system may be calculated over a range of samples in α'' by using a one-dimensional FFT.

2.3. Dihedral complexes

In addition to having a principal C_n axis, the D_n point groups also have n twofold rotational axes in the plane perpendicular to the C_n axis. Thus, applying a flip about any one of these twofold axes will produce an indistinguishable arrangement. It is therefore natural to consider making a D_3 structure from two C_3 solutions, for example. However, this involves determining one additional translational and one rotational DOF with respect to the principal rotation axis. Thus, in the SPF basis, it is convenient to let the principal rotation axis in D_n coincide with the global z axis. This is shown in Fig. 2, where $T_z(E)$ and $R_z(\eta)$ denote the two additional DOFs.

If we make C_n dimers using the above FFT procedure, we will obtain a list of parameters, $(D_k, \alpha_k, \beta_k, \gamma_k)$, for each pairwise docking solution, k. However, it should be borne in mind that the two C_n structures could be arranged in either a 'head-to-head' or a 'tail-to-tail' orientation (Fig. 2). Taking into account the above transformations leads us to define the following operator:

$$\hat{M}^X_{ki} = \hat{R}_z(\omega_j)\hat{R}_y(\tau_i)\big[\hat{R}_x(-\pi/2)\hat{T}_z(S_k)\hat{R}(\alpha_k, \beta_k, \gamma_k)\big]. \tag{31}$$

Here, $X = A$ represents the A monomer $(\omega_0 = 0)$ on the positive y axis, $X = B$ represents the B monomer $(\omega_{+1} = 2\pi/n)$ and $X = C$ represents the C monomer $(\omega_{-1} = -2\pi/n)$ in $D_{n>2}$ systems. Similarly, $\tau_0 = 0$ and $\tau_1 = \pi$ represent a possible flip about the y axis, which will soon be useful. Using this notation, we can locate three monomers of the kth solution using

$$A_{ki}(\underline{r}) = \hat{M}^A_{ki}A(\underline{r}),\ \ B_{ki}(\underline{r}) = \hat{M}^B_{ki}A(\underline{r}),\ \ C_{ki}(\underline{r}) = \hat{M}^C_{ki}A(\underline{r}). \tag{32}$$

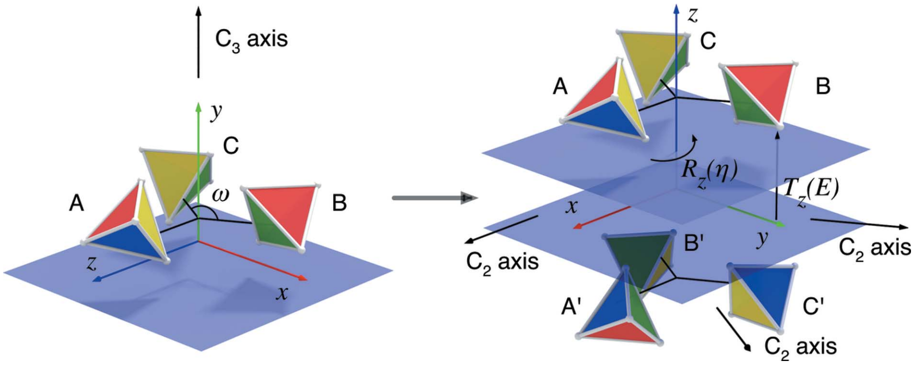


Figure 2
(Left) Illustration of the C_3 point group symmetry with the y axis as the principal rotational symmetry axis and $\omega = 2\pi/n$. Each asymmetric protein monomer is represented by a tetrahedron having four differently coloured faces (red, green, blue and yellow). (Right) A D_3 system may be generated from two planar C_3 solutions (but note the change of axes here with respect to the C_3 system on the left). When starting from a C_n solution, the D_n assembly problem has one translational and one rotational DOF, here denoted as $T_z(E)$ and $R_z(\eta)$, respectively. From symmetry, the rotational search range in $R_z(\eta)$ may be restricted to $0 \leq \eta < 2\pi/n$.

Furthermore, it is convenient to consider a new pseudo-molecule, denoted as $P_{1,ki}(\underline{r})$, constructed as the union of the A, B and C monomers of the first C_n ring system:

$$P_{1,ki}(\underline{r}) = A_{ki}(\underline{r})i + B_{ki}(\underline{r}) + C_{ki}(\underline{r}). \qquad (33)$$

A pair of such pseudo-molecules could then be docked by performing a restricted rigid-body docking search with respect to the z axis using two new operators $\hat{T}_z(E)$ and $\hat{R}_z(\eta)$. The presence of the perpendicular C_2 axes may be embedded in the search space by introducing an additional rotation, $\hat{R}_y(\pi)$. Consequently, a trimer of the second C_n system [denoted as $P_2(\underline{r})$] may be located using

$$P_{2,ki}(\underline{r}) = \hat{R}_z(\eta)\hat{T}_z(E)\hat{R}_y(\pi)P_{1,ki}(\underline{r}). \qquad (34)$$

Thus, taking into account the multiplicity of pair-wise interactions, for $n \geq 3$, the docking interaction between the two cyclic systems may be calculated as a series of point evaluations:

$$S_{ki} = n \textstyle\int A_{1,ki}(\underline{r})^* P_{2,ki}(\underline{r})\,\mathrm{d}\underline{r}. \qquad (35)$$

In principle, calculations involving equation (35) could be accelerated by a one-dimensional FFT search in η, but we currently sample η explicitly because we expect that a one-dimensional FFT will not give a large speedup when $n > 2$. Thus, by iterating over a range of samples (E, η, τ) for each C_n solution defined by $(D_k, \alpha_k, \beta_k, \gamma_k)$, we obtain a ranked list of parameters (E_m, η_m, τ_m) from which the D_n multimer may be built. For $n = 2$, we set the C coefficients to zero, and for $n > 3$, we shift the $A_{1,ki}(\underline{r}) \leftrightarrow P_{2,ki}(\underline{r})$ system to the origin for better numerical stability (details not shown).

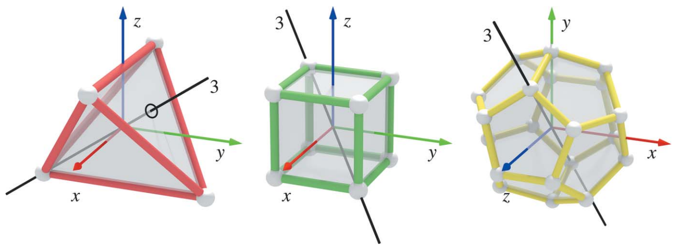


Figure 3
Computational orientations for tetrahedral (T), octahedral (O) and icosahedral (I) complexes. For each symmetry type, a solid black line shows one of the threefold axes. Candidate symmetrical complexes may be created by placing a C_3 trimer at each vertex (white sphere) of the desired symmetry type.

2.4. Tetrahedral complexes

Because the T, O and I groups all have multiple C_3 axes, a natural way to build complexes with these symmetries is to begin by making C_3 trimers, and then to assemble an appropriate number of trimeric copies to build the final complex. In particular, a tetrahedron has one threefold symmetry axis about each of four lines joining the four face centres and vertices, and one twofold rotational symmetry axis through each of three lines joining pairs of opposite edges. The dihedral angle between two faces is $\Gamma = \cos^{-1}(1/3) = 70.53°$. Seen from the origin, the angle between any two face centres is $\pi - \Gamma = 109.47°$, which is the classical tetrahedral bond angle in methane, for example. The angle between the base and the fourth vertex is half of this angle, *i.e.* $\gamma = (\pi - \Gamma)/2 = \sin^{-1}(2^{1/2}/3^{1/2}) = 54.74°$.

When building a tetrahedral complex by placing a C_3 trimer at each vertex, it is easy to see that there is one rotational DOF, η, about its threefold axis with $0 \leq \eta < 2\pi/3$, and one translational DOF along that axis, which can be considered as the 'radius', E, of the tetrahedron. Furthermore, like the D_n case, there is an unknown 'flip' of each trimer perpendicular to the threefold axis which needs to be determined.

Computationally, it is convenient to start with two edges of the tetrahedron parallel to the x and y axes (Fig. 3) and to let the global z axis correspond to the two additional DOFs, $R_z(\eta)$ and $\hat{T}_z(E)$, as before. The following symmetry-preserving operations may then be used to locate the A monomers (repeat with M^B and M^C to locate the B and C monomers) at the four tetrahedral vertices:

$$A_1(\underline{r}) = \hat{R}_y(\gamma)\hat{R}_z(\eta)\hat{T}_z(E)\hat{M}^A_{ki}A(\underline{r}), \quad A_2(\underline{r}) = \hat{R}_z(\pi)A_1(\underline{r}),$$
$$A_3(\underline{r}) = \hat{R}_x(\pi)\hat{R}_z(+\pi/2)A_1(\underline{r}), \quad A_4(\underline{r}) = \hat{R}_x(\pi)\hat{R}_z(-\pi/2)A_1(\underline{r}), \qquad (36)$$

where E is the 'radius' of the tetrahedron (the distance between its centre and one vertex, or its centre and the centre of mass of one trimer). If D is the docking distance between a pair of trimers, this corresponds to the length of an edge on the tetrahedron and the face diagonal of its enclosing cube. Hence it can be shown from basic geometry that $E = (3/8)^{1/2}D$. Then, by calculating a trimeric pseudo-molecule [equation (33)], an SPF correlation search may be used to assemble tetrahedral complexes rather efficiently because it allows the principal monomer–monomer interactions $[(A_1 + B_1 + C_1) \leftrightarrow (A_2 + B_2 + C_2)$ *etc.*] to be calculated together, as shown above for the case of D_n.

2.5. Octahedral complexes

The O group has eight threefold rotational symmetry axes (as well as three fourfold and six twofold axes). Therefore, as before, it is natural to begin by making a C_3 trimer. In this case, it is convenient to start with the diagonals of the cube parallel to the coordinate axes. Then, from basic geometry, the angle between the vertical axis and each of the top vertices is given by $\gamma = \tan^{-1}(2^{1/2}) = 54.74^\circ$. The following symmetry-preserving operations may be used to locate the A monomers (repeat with M^B and M^C to locate the B and C monomers) at the eight octahedral vertices (Fig. 3):

$$\begin{aligned}
&A_1(\underline{r}) = \hat{R}_y(\gamma)\hat{R}_z(\eta)\hat{T}_z(E)\hat{M}^A_{ki}A(\underline{r}), \quad A_2(\underline{r}) = \hat{R}_z(1\pi/2)A_1(\underline{r}),\\
&A_3(\underline{r}) = \hat{R}_z(2\pi/2)A_1(\underline{r}), \quad A_4(\underline{r}) = \hat{R}_z(3\pi/2)A_1(\underline{r}),\\
&A_5(\underline{r}) = \hat{R}_y(\pi)A_1(\underline{r}), \quad A_6(\underline{r}) = \hat{R}_y(\pi)A_2(\underline{r}),\\
&A_7(\underline{r}) = \hat{R}_y(\pi)A_3(\underline{r}), \quad A_8(\underline{r}) = \hat{R}_y(\pi)A_4(\underline{r}).
\end{aligned} \tag{37}$$

As before, once a list of candidate C_3 trimers has been calculated the subsequent octahedral assembly step may be calculated using trimeric pseudo-molecules [equation (33)].

2.6. Icosahedral complexes

The I group has 20 threefold rotational symmetry axes (as well as 24 C_5 and 15 C_2 axes), and so assembling 20 C_3 trimers will give a complex of 60 monomers. In this case, it is convenient to consider 20 vertices of a dodecahedron, which is the dual of the icosahedron, in which the centre of one of its pentagonal faces is located on the positive z axis. By initially locating the first C_3 trimer on the positive z axis, we then need to rotate it onto each vertex of a dodecahedron to define the 20 C_3 axes.

The required rotations may be deduced from the geometry of the dodecahedron. More specifically, it is well known that the dihedral angle between two pentagon faces is $\Gamma = \cos^{-1}(-5^{1/2}) = 116.56^\circ$. Therefore, as seen from the origin, the angle between the centres of a pair of touching pentagons is $\gamma = \pi - \Gamma = \cos^{-1}(1/5^{1/2})$. This can be used to calculate the 'width' of a pentagon face (distance from the centre to the middle of an edge) as

$$W = R\tan(\gamma/2), \tag{38}$$

where R is the distance from the origin to the pentagonal face centre. Then, in the plane of the pentagon, its 'radius' P is given by

$$P = W/\cos(\pi/5). \tag{39}$$

So now we can rotate the initial trimer off the y axis and onto the first pentagon vertex by applying a rotation of $\hat{R}_x(-\beta)$, where

$$\beta = \tan^{-1}(P/R) = \tan^{-1}[\tan(\gamma/2)/\cos(\pi/5)]. \tag{40}$$

Now it can also be shown that the distance E from the origin to each vertex is given by

$$E = P/\sin\beta, \tag{41}$$

where $P = D/[2\sin(\omega/2)]$. Hence, the radial and docking distances, E and D, are related according to

$$E = \frac{D}{2\sin(2\pi/10)\sin\beta}. \tag{42}$$

Using these distances, the A monomers of C_3 trimers may be located at the dodecahedron vertices (Fig. 3) using the following symmetry-preserving operations (which should be repeated with M^B and M^C to locate the B and C monomers):

$$\begin{aligned}
&A_1(\underline{r}) = \hat{R}_y(-\beta)\hat{R}_z(\eta)\hat{T}_z(E)\hat{M}^A_{ki}A(\underline{r}), \quad A_2(\underline{r}) = \hat{R}_z(2\pi/5)A_1(\underline{r}),\\
&A_3(\underline{r}) = \hat{R}_z(4\pi/5)A_1(\underline{r}), \quad A_4(\underline{r}) = \hat{R}_z(6\pi/5)A_1(\underline{r}),\\
&A_5(\underline{r}) = \hat{R}_z(8\pi/5)A_1(\underline{r}), \quad A_6(\underline{r}) = \hat{R}_y(\gamma)\hat{R}_z(\pi)A_1(\underline{r}),\\
&A_7(\underline{r}) = \hat{R}_y(\gamma)\hat{R}_z(\pi)A_2(\underline{r}), \quad A_8(\underline{r}) = \hat{R}_z(2\pi/5)A_6(\underline{r}),\\
&A_9(\underline{r}) = \hat{R}_z(2\pi/5)A_7(\underline{r}), \quad A_{10}(\underline{r}) = \hat{R}_z(4\pi/5)A_6(\underline{r}),\\
&A_{11}(\underline{r}) = \hat{R}_z(4\pi/5)A_7(\underline{r}), \quad A_{12}(\underline{r}) = \hat{R}_z(6\pi/5)A_6(\underline{r}),\\
&A_{13}(\underline{r}) = \hat{R}_z(6\pi/5)A_7(\underline{r}), \quad A_{14}(\underline{r}) = \hat{R}_z(8\pi/5)A_6(\underline{r}),\\
&A_{15}(\underline{r}) = \hat{R}_z(8\pi/5)A_7(\underline{r}), \quad A_{16}(\underline{r}) = \hat{R}_y(\pi)A_1(\underline{r}),\\
&A_{17}(\underline{r}) = \hat{R}_y(\pi)A_2(\underline{r}), \quad A_{18}(\underline{r}) = \hat{R}_y(\pi)A_3(\underline{r}),\\
&A_{19}(\underline{r}) = \hat{R}_y(\pi)A_4(\underline{r}), \quad A_{20}(\underline{r}) = \hat{R}_y(\pi)A_5(\underline{r}).
\end{aligned} \tag{43}$$

As before, once a list of candidate C_3 trimers has been calculated, the subsequent icosahedral assembly step may be calculated using trimeric pseudo-molecules [equation (33)].

3. Results and discussion

To test our approach, we selected a representative example structure of each complex symmetry type for which three-dimensional structures exist in the 3D-Complex database. These examples are listed in Table 2. For each complex, we manually extracted the first monomer from the PDB file to serve as the A monomer, and we applied the SPF assembly algorithm for the given symmetry type using SPF expansions to polynomial order $N = 30$.

More specifically, for the C_n correlation search (and for the initial trimeric search in the higher-symmetry types), the (β, γ) angular samples were generated from an icosahedral tessellation of the sphere with 812 sample vertices with an angular separation between the vertices of approximately 7.5°. The FFT search in α was calculated using 64 steps of approximately 2.8° in the first hemisphere, and up to 64 translational steps of 0.8 Å were applied, starting from an initial inter-monomer distance estimated from the monomer radius. Thus, a total of approximately 6×10^6 trial $A_1 \leftrightarrow B_1$ orientations were generated and scored in the FFT search. The B_1 monomers of the generated solutions were then clustered using a greedy

Table 2
Example symmetrical complexes assembled from a single monomer by the *SAM* algorithm with $N = 30$.

Here, #Res denotes the number of residues in one monomer of each structure, B_1 denotes the B monomer of the first C_n system, and B_2 denotes a B monomer of the second ring system in D_n complexes or of an adjoining C_3 trimer for T, O and I complexes. All RMSD values are in ångström units and all times are elapsed seconds for a Linux workstation with dual six-core (2.67 GHz) Intel X5650 processors. 'N/F' denotes not found. A hyphen denotes not applicable.

			M-Zdock			*SymmDock*			*SAM*					
PDB	#Res	Sym	Rank-C_n	RMSD-B_1	Time	Rank-C_n	RMSD-B_1	Time	Rank-C_n	RMSD-B_1	RMSD-B_2	Rank	RMSD	Time
1m4g	182	C_2	N/F	N/F	5963	26	21.47	6	1	1.82	–	1	1.82	45
1f7o	117	C_3	1	2.33	4641	1	2.32	14	1	2.82	–	1	2.82	48
1f8c	389	C_4	1	2.00	11171	1	2.37	62	1	2.04	–	1	2.04	40
1g8z	104	C_5	1	1.87	3187	1	2.02	15	1	1.62	–	1	1.62	43
1gl7	412	C_6	1	1.41	14228	1	1.41	40	1	0.68	–	1	0.68	50
1i81	75	C_7	1	1.95	2571	1	4.02	7	1	1.17	–	1	1.17	43
1v5w	240	C_8	1	2.49	7354	1	2.93	14	1	2.51	–	1	2.51	44
1qaw	68	C_{11}	1	2.61	2196	1	1.75	5	1	1.09	–	1	1.09	43
1xib	389	D_2	–	–	–	–	–	–	1	1.01	0.68	1	0.86	319
1gun	68	D_3	–	–	–	–	–	–	2	1.35	0.99	1	1.19	308
1b9l	120	D_4	–	–	–	–	–	–	1	1.34	1.57	1	1.46	393
1l6w	221	D_5	–	–	–	–	–	–	1	1.26	3.61	5	2.70	479
1znn	246	D_6	–	–	–	–	–	–	1	1.34	1.92	1	1.66	439
1yg6	194	D_7	–	–	–	–	–	–	1	1.94	3.30	1	2.70	381
1q3r	519	D_8	–	–	–	–	–	–	2	3.65	10.83	25	7.98	397
2cc9	65	T	–	–	–	–	–	–	1	1.97	2.63	1	2.32	199
1ies	175	O	–	–	–	–	–	–	1	1.24	0.94	1	1.10	201
1hqk	155	I	–	–	–	–	–	–	1	1.45	1.88	1	1.68	200

clustering algorithm with a 3 Å root-mean-square deviation (RMSD) cluster threshold in order to remove near-duplicate solutions, and the top-scoring member of each of the first 100 clusters was retained as a distinct solution. For the C_n complexes, any remaining monomer coordinates were generated by symmetry, and the top 100 solutions were saved as PDB files. When calculating the FFT correlations in parallel using these parameters, it takes approximately 30 s to generate 100 C_n complexes on a dual processor workstation with two 2.3 GHz E4510 Intel Xeon processors (eight cores in total).

For the D_n, T, O and I complexes, similar angular and translational search parameters were then used again in the subsequent trimeric assembly search using the top 100 trimeric solutions. For these complexes, the calculation time is governed by the cost of constructing the trimeric pseudo-molecules and the cost of performing the subsequent correlation search explicitly, without the benefit of an FFT.

To assess the quality of the generated complexes, the coordinates of the crystallographically determined complex structure were used as a reference structure with which to calculate RMSDs between the calculated and reference monomer coordinates. For all of the examples in Table 2, the 'Rank-C_n' and 'RMSD-B_1' columns show the rank and RMSD for the first B monomer of the C_n complexes (or the trimeric component in the higher-symmetry cases) found within 10 Å of the crystal structure. This column shows that, in all but one case (1gun), our one-dimensional FFT search is correctly identifying a near-native interface between the A and B monomers. Given that this calculation is rigidly assembling monomers which should fit perfectly, these very good results are not especially surprising. Nonetheless, these figures confirm that our FFT correlation expressions are implemented correctly. Fig. 4 shows cartoon representations of the first near-native solution found for each complex.

In order to compare the performance of *SAM* with some existing symmetry docking algorithms, we selected *M-Zdock* (Pierce *et al.*, 2005) as a good example of an FFT-based algorithm and *SymmDock* because it is based on a geometric hashing technique (Schneidman-Duhovny *et al.*, 2005). Table 2 shows that these algorithms can also successfully find rank 1 solutions with low RMSDs for all of our C_n examples (both *M-Zdock* and *SymmDock* were designed only for C_n complexes) except for the first C_2 structure (1m4g), for which *M-Zdock* does not find a solution in its top ten predictions and for which *SymmDock* finds a very poor solution only at rank 26. However, if we consider the seven examples (C_3–C_{11}) for which all three algorithms produce rank 1 solutions, Table 2 shows that *SymmDock* is approximately twice as fast as *SAM*, while *SAM* is approximately 130 times faster than *M-Zdock*, with average execution times of 23 s for *SymmDock*, 44 s for *SAM* and 5734 s for *M-Zdock*. Furthermore, the RMSD-B_1 columns of this table show that *SAM* often gives considerably better quality solutions, with average RMSD values of 1.70 Å for *SAM*, 2.09 Å for *M-Zdock* and 2.40 Å for *SymmDock*. These results show that *SAM* performs quite favourably when compared with these previous approaches.

In order to assess the trimeric pseudo-molecule assembly step for the D_n, T, O and I complexes, the 'RMSD-B_2' column of Table 2 reports the best RMSD found by *SAM* for the calculated coordinates of the B_2 monomer. This column shows that our strategy of scoring the interactions between trimeric pseudo-molecules works very well for all of the examples except for the D_8 complex (PDB code 1q3r). Finally, the 'Rank' and 'RMSD' columns give the rank and overall RMSD of the first B_1 and B_2 solutions found within 10 Å of the crystal structure. These columns show that in 16 out of the 18 examples the first solution calculated by *SAM* corresponds very closely to the crystal structure. For the D_5 example (PDB

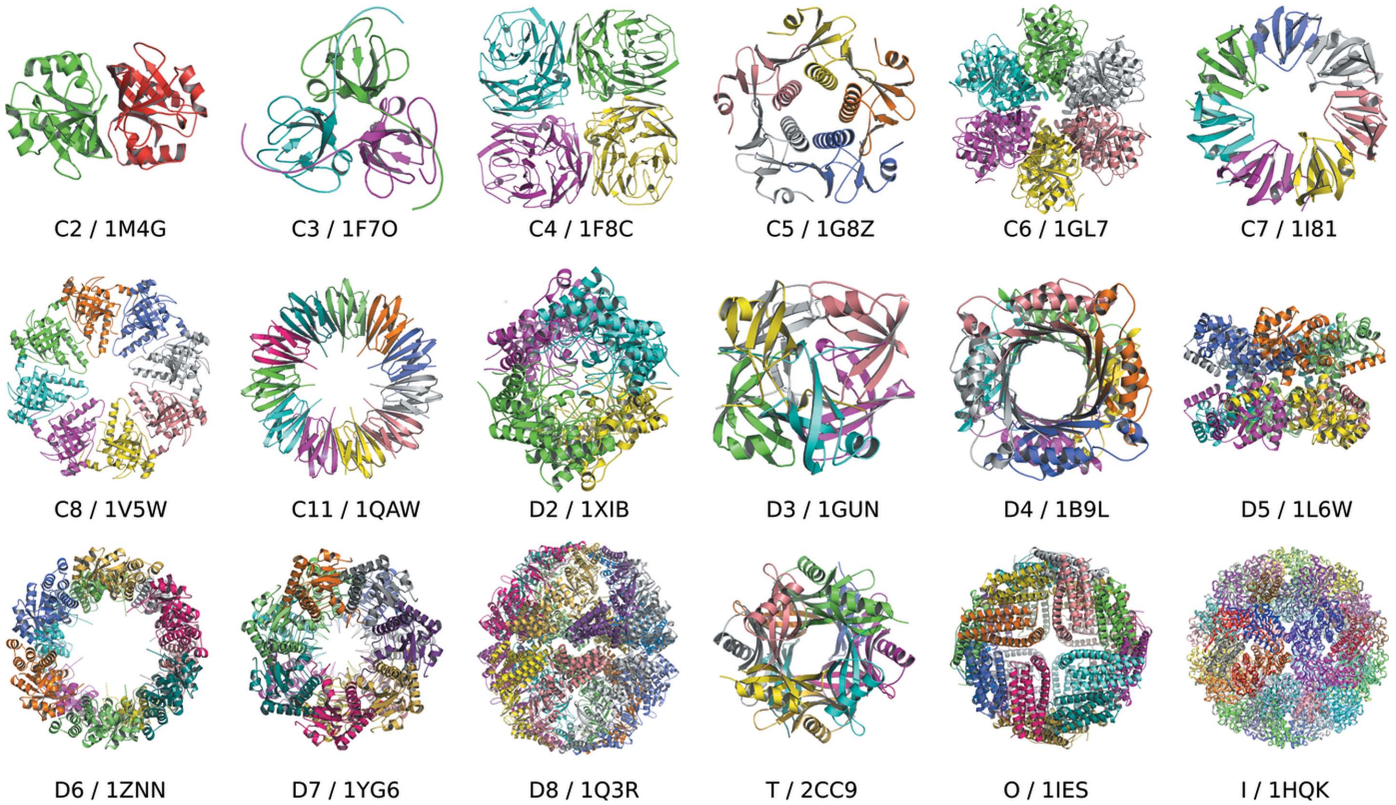


Figure 4
The example symmetrical complexes assembled by *SAM*, starting from a single monomer from the crystal structure. Computational details are provided in Table 2.

code 1l6w), the first near-native structure is found at rank 5. Although a good trimer is found at rank 2 for the D_8 example (PDB code 1q3r), the subsequent trimer assembly step finds a rather poor near-native orientation only at rank 25.

Despite these rather promising results, we know that one limitation of the SPF approach is that most of the zeros in the basis functions appear within about 50 Å from the origin. This means that very large protein domains, typically greater than about 500 residues, cannot be represented accurately by a single SPF polynomial expansion. We believe that this explains the poor performance of the 1q3r example (519 residues per monomer). Taking into account the possibility that one monomer might consist of several chains, we have calculated that 87% (9024/10 176) of C_n complexes, 91% (2704/2965) of D_n complexes, and 43% (60/139) of the T (43/86), O (12/47) and (non-viral) I (5/6) complexes in the 3D-Complex database have fewer than 500 residues per monomer. In other words, we estimate that *SAM* could be usefully applied in approximately 89% of protein docking problems that involve point group symmetry. One way to circumvent the monomer size limitation would be to use a coarse-grained force-field model to perform the trimeric assembly step, for example. Indeed, since the FFT correlation function used here is based on a simple surface skin density model of protein shape (Ritchie & Kemp, 2000), it would be advisable to refine and rescore the *SAM* models using a conventional molecular mechanics force field if clash-free atomic models are required.

While this article has focused on complexes having point group symmetry, we expect it would be relatively straightforward to extend the *SAM* algorithm to deal with complexes having translational symmetry, such as cylindrical and helical structures. Cylindrical structures could be made in the same way that we make a D_n complex from two C_n systems, but without applying a flip $[\hat{R}_y(\tau_1)]$ in equation (31). Helical structures could be made by introducing an additional translational DOF in our C_n assembly algorithm. This would correspond to replacing $\hat{R}_y(\omega)$ with $\hat{T}_y(\eta)\hat{R}_y(\omega)$ throughout §2.2, where $\hat{T}_y(\eta)$ represents a translation along the major helical axis.

The *SAM* program may be downloaded for academic use at http://sam.loria.fr/.

4. Conclusion

We have presented a novel FFT-based approach called *SAM* for building models of protein complexes with arbitrary point group symmetry. The basic approach relies on a very fast one-dimensional symmetry-constrained spherical polar FFT search to assemble cyclic C_n systems from a given protein monomer. Structures with higher-order (D_n, T, O and I) symmetries may be built by performing a subsequent symmetry-constrained Fourier domain search to assemble trimeric pseudo-molecules. Overall, our results demonstrate that the *SAM* algorithm can correctly and rapidly assemble protein complexes with arbitrary point group symmetry from a given monomer structure in 17 out of 18 test complexes. The main limitation of our approach is that the resolution of the SPF representation begins to degrade with monomers having more than about 500 residues, and this therefore sets a limit on the size of symmetrical complexes that can be modelled. We propose that one way to address this limitation would be to use a residue-based coarse-grained force-field representation in place of the Fourier domain pseudo-molecules during the final trimeric assembly stage.

APPENDIX *A*
Restricting the α range in C_n

Because there exist n C_2 axes perpendicular to the principal C_n axis, the range of α in equation (8) must be restricted to $0 \le \alpha < \pi$, instead of the natural Euler range of $0 \le \alpha < 2\pi$, in order to avoid generating duplicate configurations in the C_n multimer. To prove this mathematically, we need to show that

$$\sum_{j=0}^{n-1} \hat{R}_y(\omega_j)\hat{T}_z(\Delta)\hat{R}(\alpha+\pi,\beta,\gamma)A(\underline{r}) = \hat{R}_z(\pi)\sum_{j=0}^{n-1} \hat{R}_y(\omega_j)\hat{T}_z(\Delta)\hat{R}(\alpha,\beta,\gamma)A(\underline{r}). \tag{44}$$

Because $\hat{R}_z(\pi)$ and $\hat{T}_z(\Delta)$ commute, and using the fact that $\hat{R}_y(\omega_j)\hat{R}_z(\pi) = \hat{R}_z(\pi)R_y(-\omega_j)$, we can rewrite the left-hand side as

$$\sum_{j=0}^{n-1} \hat{R}_y(\omega_j)\hat{R}_z(\pi)\hat{T}_z(\Delta)\hat{R}(\alpha,\beta,\gamma)A(\underline{r}) = \hat{R}_z(\pi)\sum_{j=0}^{n-1} \hat{R}_y(-\omega_j)\hat{T}_z(\Delta)\hat{R}(\alpha,\beta,\gamma)A(\underline{r}). \tag{45}$$

Then, noting that $\hat{R}_y(-\omega_j) = \hat{R}_y(+\omega_{n-j})$ and changing the order of the summation, we obtain

$$\hat{R}_z(\pi)\sum_{j=0}^{n-1} \hat{R}_y(-\omega_j)\hat{T}_z(\Delta)\hat{R}(\alpha,\beta,\gamma)A(\underline{r}) = \hat{R}_z(\pi)\sum_{j=n-1}^{0} \hat{R}_y(+\omega_j)\hat{T}_z(\Delta)\hat{R}(\alpha,\beta,\gamma)A(\underline{r}). \tag{46}$$

This proves equation (44).

APPENDIX *B*
Real and complex SPF basis functions

Here, we let $y_{lm}(\theta,\varphi)$ and $Y_{lm}(\theta,\varphi)$ represent real and complex spherical harmonic basis functions, respectively, where $Y_{lm}(\theta,\varphi)$ are defined by

$$Y_{lm}(\theta,\varphi) = \left[\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}\right]^{1/2} P_{lm}(\cos\theta)\exp(im\varphi) \tag{47}$$

and where $P_{lm}(\cos\theta)$ are the Legendre polynomials (Hobson, 1931). The real and complex basis are functions related by a unitary transformation matrix $U^{(l)}$ as

$$y_{lm}(\theta,\varphi) = \sum_{m'} U^{(l)}_{mm'} Y_{lm'}(\theta,\varphi), \tag{48}$$

where the matrix elements of $U^{(l)}$ have the form (Biedenharn & Louck, 1981)

$$\begin{pmatrix} y_{ll} \\ y_{lm} \\ y_{l0} \\ y_{l\overline{m}} \\ y_{l\bar{l}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2^{1/2}} & 0 & 0 & 0 & \frac{(-1)^l}{2^{1/2}} \\ 0 & \frac{1}{2^{1/2}} & 0 & \frac{(-1)^m}{2^{1/2}} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{i(-1)^m}{2^{1/2}} & 0 & \frac{-i}{2^{1/2}} & 0 \\ \frac{i(-1)^l}{2^{1/2}} & 0 & 0 & 0 & \frac{-i}{2^{1/2}} \end{pmatrix} \begin{pmatrix} Y_{ll} \\ Y_{lm} \\ Y_{l0} \\ Y_{l\overline{m}} \\ Y_{l\bar{l}} \end{pmatrix}. \tag{49}$$

Together with the spherical harmonics, we use real orthonormal Gauss–Laguerre radial basis functions in order to provide a complete orthonormal three-dimensional basis set

$$R_{nl}(r) = \left[\frac{2}{\lambda^{3/2}}\frac{(n-l-1)!}{\Gamma(n+1/2)}\right]^{1/2} \exp(-\rho^2/2)\rho^l L^{(l+1/2)}_{n-l-1}(\rho^2), \tag{50}$$

where $\rho^2 = r^2/\lambda$ and λ is a radial distance scale factor. In this work, we set λ = 40 Å.

Although it is convenient to calculate a polar Fourier expansion of protein shape initially in real space with real expansion coefficients, a_{nlm} (Ritchie & Kemp, 2000),

$$A(\underline{r}) = \sum_{nlm} a_{nlm} R_{nl}(r) y_{lm}(\theta,\varphi), \tag{51}$$

subsequently it is often more convenient to work in the complex basis using

$$A(\underline{r}) = \sum_{nlm} A_{nlm} R_{nl}(r) Y_{lm}(\theta,\varphi), \tag{52}$$

where the complex expansion coefficients, A_{nlm}, are related to the original real coefficients by

$$A_{nlm} = \sum_{m'} U^{(l)}_{m'm} a_{nlm'}. \tag{53}$$

In the complex basis, the rotation operator is represented as the complex Wigner rotation matrix:

$$D^{(l)}_{mm'}(\alpha,\beta,\gamma) = \exp(-im\alpha)\, d^l_{mm'}(\beta)\exp(-im'\gamma). \tag{54}$$

Note that a pure rotation about the y axis may be expanded as

$$\hat{R}_y(\beta) \equiv \hat{R}_z(-\pi/2)\hat{R}_y(-\pi/2)\hat{R}_z(\beta)\hat{R}_y(\pi/2)\hat{R}_z(\pi/2). \tag{55}$$

This expansion is useful in the polar Fourier representation because it allows $d^l_{mm'}(\beta)$ to be expanded as a product of exponentials (Edmonds, 1957):

$$d^l_{mm'}(\beta) = \sum_t \exp(im\pi/2)\, d^l_{mt}(-\pi/2)\exp(-it\beta) \times d^l_{tm'}(\pi/2)\exp(-im'\pi/2). \tag{56}$$

Then, by writing

$$\Delta^l_{tm} = d^l_{tm}(\pi/2) = d^l_{mt}(-\pi/2), \tag{57}$$

and by collecting constant terms

$$\Gamma^l_{mtm'} = \exp[i(m-m')\pi/2]\Delta^l_{tm}\Delta^l_{tm'} = i^{m-m'}\Delta^l_{tm}\Delta^l_{tm'}, \qquad (58)$$

the Wigner rotation matrix elements may be written in a completely exponential form:

$$D^{(l)}_{mm'}(\alpha,\beta,\gamma) = \sum_t \Gamma^l_{mtm'} \exp(-im\alpha)\exp(-it\beta)\exp(-im'\gamma). \qquad (59)$$

In order to perform three-dimensional rotational FFT searches over the Wigner rotations, it is necessary to scale the β rotation onto the natural domain of the FFT. This may be achieved by putting $\beta' = 2\beta$ and then writing

$$\exp(-it\beta) = \sum_j \lambda_{tj}\exp(-ij\beta'). \qquad (60)$$

It can be shown that the coefficients λ_{tj} may be determined to reproduce exactly a finite set of M_β rotational samples by treating equation (60) as a discrete Fourier transform (DFT) analysis equation (Ritchie *et al.*, 2008):

$$\lambda^{(\beta_{\max})}_{tj} = \frac{1}{M_\beta}\sum_{n=0}^{M_\beta-1}\exp(-itn\beta_{\max}/M_\beta)\exp(2\pi ijn/M_\beta), \qquad (61)$$

where $\beta_{\max} = \pi$ here. Hence, we can collect coefficients as

$$\Lambda^{um}_{lv} = \sum_t \Gamma^{tm}_{lv}\lambda^{(\pi)}_{tu} \qquad (62)$$

to obtain

$$D^{(l)}_{mm'}(\alpha,\beta,\gamma) = \sum_t \Lambda^l_{mtm'}\exp(-im\alpha)\exp(-it\beta')\exp(-im'\gamma). \qquad (63)$$

We use this scaling technique in a restricted C_n FFT search (Appendix *A*) in order to scale the allowed α angle range $(0:\pi)$ back to the natural range of the FFT $(0:2\pi)$.

Acknowledgements

This work was funded by Agence Nationale de la Recherche, grant reference ANR-11-MONU-006.

References

André, I., Bradley, P., Wang, C. & Baker, D. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 17656–17661.

Berchanski, A. & Eisenstein, M. (2003). *Proteins Struct. Funct. Genet.* **53**, 817–829.

Berman, H. M. *et al.* (2002). *Acta Cryst.* D**58**, 899–907.

Biedenharn, L. C. & Louck, J. C. (1981). *Angular Momentum in Quantum Physics.* Reading: Addison-Wesley.

Comeau, S. R. & Camacho, C. J. (2004). *J. Struct. Biol.* **150**, 233–244.

Edmonds, A. R. (1957). *Angular Momentum in Quantum Physics.* New Jersey: Princeton University Press.

Goodsell, D. S. & Olsen, A. J. (2000). *Ann. Rev. Biophys. Biomol. Struct.* **29**, 105–153.

Hobson, E. W. (1931). *The Theory of Spherical and Ellipsoidal Harmonics.* London: Cambridge University Press.

Huang, P.-S., Love, J. J. & Mayo, S. L. (2005). *J. Comput. Chem.* **26**, 1222–1232.

Karaca, E., Melquiond, A. S. J., de Vries, S. J., Kastritis, P. L. & Bonvin, A. M. J. J. (2010). *Mol. Cell. Proteomics*, **9**, 1784–1794.

Levy, E. D., Boeri Erba, E., Robinson, C. V. & Teichmann, S. (2008). *Nature*, **453**, 1262–1266.

Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. (2006). *PLoS Comput. Biol.* **2**, e155.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Pierce, B., Tong, W. & Weng, Z. (2005). *Bioinformatics*, **21**, 1472–1478.

Ritchie, D. W. (2005). *J. Appl. Cryst.* **38**, 808–818.

Ritchie, D. W. & Kemp, G. J. L. (2000). *Proteins Struct. Funct. Genet.* **39**, 178–194.

Ritchie, D. W., Kozakov, D. & Vajda, S. (2008). *Bioinformatics*, **24**, 810–823.

Roseman, A. M. (2000). *Acta Cryst.* D**56**, 1332–1340.

Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. (2005). *Nucleic Acids Res.* **33**, W363–W367.

Tong, L. (2001). *Acta Cryst.* D**57**, 1383–1389.