

Acta Crystallographica Section D

**Biological  
Crystallography**

ISSN 0907-4449

## **Fast rotational matching of rigid bodies by fast Fourier transform acceleration of five degrees of freedom**

**Julio A. Kovacs, Pablo Chacón, Yao Cong, Essam Metwally and Willy Wriggers**

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

# Fast rotational matching of rigid bodies by fast Fourier transform acceleration of five degrees of freedom

**Julio A. Kovacs, Pablo Chacón,<sup>†</sup>  
 Yao Cong,<sup>‡</sup> Essam Metwally<sup>§</sup>  
 and Willy Wriggers<sup>\*‡</sup>**

Department of Molecular Biology, The Scripps  
 Research Institute, 10550 North Torrey Pines  
 Road, La Jolla, CA 92037, USA

<sup>†</sup> Present address: Centro de Investigaciones  
 Biológicas, Calle Velázquez, 144, 28006  
 Madrid, Spain.

<sup>‡</sup> Present address: School of Health Information  
 Sciences, University of Texas-Houston,  
 7000 Fannin, Houston, TX 77030, USA.

<sup>§</sup> Present address: Tripos Inc., 1699 South  
 Hanley Rd, St Louis, MO 63144, USA.

Correspondence e-mail:  
 wriggers@biomachina.org

Received 12 March 2003  
 Accepted 20 May 2003

The 'fast rotational matching' method (an approach to find the three rotational degrees of freedom in matching problems using just one three-dimensional FFT) is extended to the full six-dimensional (rotation and translation) matching scenario between two three-dimensional objects. By recasting this problem into a formulation involving five angles and just one translational parameter, it was possible to accelerate, by means of fast Fourier transforms, five of the six degrees of freedom of the problem. This method was successfully applied to the docking of atomic structures of components into three-dimensional low-resolution density maps. Timing comparisons performed with our method and with 'fast translational matching' (the standard way to accelerate the translational parameters utilizing fast Fourier transforms) demonstrates that the performance gain can reach several orders of magnitude, especially for large map sizes. This gain can be particularly advantageous for spherical- and toroidal-shaped maps, since the scanning range of the translational parameter would be significantly constrained in these cases. The method can also be harnessed to the complementary surface (or 'exterior docking') problem and to pattern recognition in image processing.

## 1. Introduction

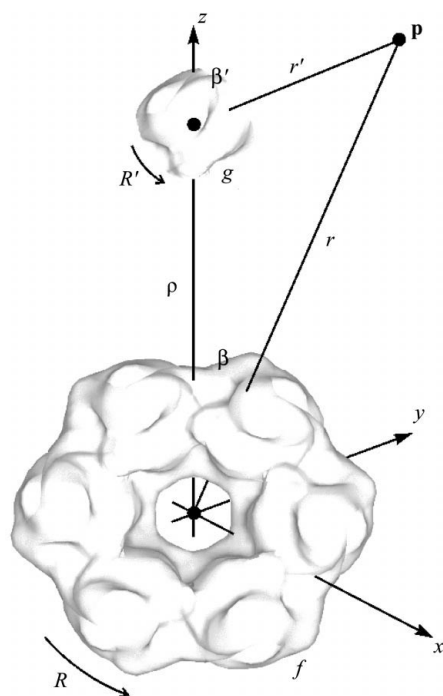
Numerous scientific disciplines encounter at some point the problem of finding the best rotational and/or translational match between two given objects in three-dimensional space. Examples can be found in fields such as pattern recognition (Shams *et al.*, 2001), engineering (Paquet *et al.*, 2000), machine vision (Siddiqi *et al.*, 1999) and biophysics (Wriggers *et al.*, 1999). In the latter field, we find it in several areas of molecular and structural biology, for instance biomolecular docking (Ritchie & Kemp, 2000; Wriggers & Chacón, 2001), image processing and reconstruction in electron microscopy (EM; Frank, 1996) and the molecular-replacement method used in X-ray crystallography (Navaza, 1994). A fast method would be specially useful in the protein–protein or ligand–receptor docking problem, of importance in drug development, and also in the problem of structure determination of large biomolecular machines.

Although the applicability of our method is general, the motivation for our work arose from the problem of the docking of atomic structures of components into low-resolution EM maps of macromolecular complexes (Frank, 1996). A number of computational strategies have been proposed to bridge this resolution gap. In wide use are methods whose fitting criterion is the maximization of the cross-correlation function between the two three-dimensional

objects (see, for example, Wriggers & Chacón, 2001, and references therein).

The standard way to exhaustively compute the cross-correlation function in the rigid-body search space is by means of what we call 'Fast Translational Matching' (FTM; Chacón & Wriggers, 2002). This consists of making an exhaustive scan of the rotations (by sampling the Euler angles, for instance) and, for each rotation, applying the convolution theorem and fast Fourier transforms to rapidly compute the correlation function for all the translations. Alternatively, we demonstrated that one could perform the exhaustive search in reverse order: scan the translations and, for each of them, perform a rapid computation of the correlation function for all the rotations using Fast Rotational Matching (FRM; Kovacs & Wriggers, 2002), a novel method developed by us that FFT-accelerates all three rotational degrees of freedom (DOFs). These approaches deal with just half of the DOFs, either the translational or the rotational ones.

Here, we present a method that allows the handling of more DOFs in a similar fashion and that, therefore, significantly speeds up the matching process. We show that it is possible to FFT-accelerate five of the DOFs (which are angular parameters), while the remaining (linear) parameter is the only one that needs to be scanned. Roughly speaking, our method views the correlation as a function of two rotations and one displacement, allowing us to write the correlation function as a Fourier expansion in terms of five angular variables (representing the rotations) and one linear variable.



**Figure 1**

Six degrees of freedom matching setup.  $r$  is the distance of a generic point  $\mathbf{p}$  from the reference point of object  $f$  (which is at the origin of coordinates) and  $\beta$  is its colatitude.  $r'$  and  $\beta'$  are the analogous quantities but relative to the reference point of object  $g$  [located at  $(0, 0, \rho)$ ]. See text for details.

We have applied this novel approach to the docking of high-resolution structures of subunits into low-resolution EM maps. Comparison with FTM shows that our method gives always timings of the order of minutes, as opposed to the FTM times which could be several hours, depending on the grid size.

## 2. Preliminaries

Let two objects in three-dimensional space  $\mathbb{R}^3$  be given by the density functions

$$f : \mathbb{R}^3 \rightarrow \mathbb{R} \quad \text{and} \quad g : \mathbb{R}^3 \rightarrow \mathbb{R}.$$

These functions are assumed to be bounded and of 'compact support', i.e. they vanish outside a bounded set. The criterion to perform the matching of the objects (that is, to find a rigid motion of space that produces the best overlap between the functions  $f$  and  $g$ ) is to maximize the correlation between one of them and a rotated and translated version of the other. This criterion allows not only for the case in which both objects are 'the same' (modulo a rigid motion), but also for the cases in which one of them is a part of the other or in which part of one is part of the other.

In order to describe the relative positions and orientations of both objects, we follow Ritchie & Kemp (2000) and rotate both objects while translating one of them along the positive  $z$  axis only. In this way, the six-dimensional search is performed over five angular parameters and one linear parameter. In Fig. 1,  $R$  and  $R'$  are the respective rotations and  $\rho$  is the distance between the two objects. Note that the reference point on each object is in principle arbitrary, but it is convenient to adopt the center of mass (COM).

In the next few paragraphs, we define the notation to be used later. For details, the reader is referred to Kovacs & Wriggers (2002) and references therein.

Points  $\mathbf{p}$  in  $\mathbb{R}^3$  will be written as  $\mathbf{p} = r\mathbf{u}$ , where  $r = |\mathbf{p}|$  and  $|\mathbf{u}| = 1$ . We denote with  $S^2$  the unit sphere:  $S^2 = \{\mathbf{u} \text{ in } \mathbb{R}^3 \text{ such that } |\mathbf{u}| = 1\}$ . For a rotation  $R$  in the three-dimensional rotation group  $SO(3)$ , let  $\Lambda_R$  be the rotation operator defined by

$$(\Lambda_R g)(\mathbf{p}) := g[R^{-1}(\mathbf{p})] \quad \text{for all } \mathbf{p} \text{ in } \mathbb{R}^3.$$

We also define, for  $\rho \geq 0$ , the translation operator  $T_\rho$ ,

$$(T_\rho g)(x, y, z) := g(x, y, z - \rho) \quad \text{for all } (x, y, z) \text{ in } \mathbb{R}^3.$$

Let us introduce the following shorthand: given a point  $\mathbf{p} = r\mathbf{u}$ , let  $(\beta, \lambda)$  be the spherical coordinates of  $\mathbf{u}$ . Namely,  $\beta$  is the colatitude (angle from the north pole) and  $\lambda$  is the longitude (angle about the  $z$  axis). Let  $\mathbf{p}' = \mathbf{p} - (0, 0, \rho) = r'\mathbf{u}'$  with  $|\mathbf{u}'| = 1$  and let  $(\beta', \lambda')$  be the spherical coordinates of  $\mathbf{u}'$ . Then  $\lambda' = \lambda$ ,  $r'^2 = r^2 - 2r\rho\cos\beta + \rho^2$  and  $\cos\beta' = (r\cos\beta - \rho)/r'$  (Fig. 1).

We denote by  $Y_{lm} : S^2 \rightarrow \mathbb{C}$  the spherical harmonic functions, where  $l \geq 0$  and  $-l \leq m \leq l$  are the degree and order, respectively (see Hobson, 1931 for details). These can be written in terms of the associated Legendre functions  $P_l^m$  as follows:

$$Y_{lm}(\beta, \lambda) = (-1)^m \left[ \frac{(2l+1)(l-m)!}{4\pi(l+m)!} \right]^{1/2} P_l^m(\cos \beta) \exp(im\lambda).$$

Our density functions  $f$  and  $g$  can be approximated by finite sums of the form

$$f(\mathbf{r}\mathbf{u}) \simeq \sum_{l=0}^{B-1} \sum_{m=-l}^l \hat{f}_{lm}(r) Y_{lm}(\mathbf{u}),$$

(and similarly for  $g$ ), where the  $\hat{f}_{lm}(r)$  are the spherical harmonic coefficients of the restriction of  $f$  to the sphere of radius  $r$  and the 'bandwidth'  $B$  is related to the sampling of the angular parameters; the number of sampling points (in each of  $\beta, \lambda$ ) used to compute the coefficients is, according to the sampling theorem (Healy *et al.*, 1998), equal to  $2B$ . As a consequence, the correlation function to be obtained (the output of a five-dimensional FFT) will also be sampled at  $2B$  points for each of the five angular parameters.

For a rotation  $R$  we have

$$(\Lambda_R f)(\mathbf{r}\mathbf{u}) = \sum_{l,m,n} \hat{f}_{lm}(r) D_{mn}^l(R) Y_{lm}(\mathbf{u}),$$

where the  $D_{mn}^l(R)$  are the matrix elements of the irreducible representations of  $SO(3)$  (Brink & Satchler, 1993). To represent rotations, we use Euler angles with the  $ZYZ$  convention (Brink & Satchler, 1993) (or transposed  $y$  convention; Weisstein, 1999); namely,  $R(\varphi, \theta, \psi)$  means rotate by  $\psi$  about the  $z$  axis, then by  $\theta$  about the  $y$  axis and finally by  $\varphi$  again about the  $z$  axis. Using these Euler angles, the  $D_{mn}^l$  can be written as (Brink & Satchler, 1993),

$$D_{mn}^l(\varphi, \theta, \psi) = \exp(-im\varphi) d_{mn}^l(\theta) \exp(-in\psi), \quad (1)$$

where the functions  $d_{mn}^l$  are real. These are a generalization of the Legendre functions in the sense that (Brink & Satchler, 1993)

$$d_{m0}^l(\beta) = (-1)^m \left[ \frac{(l-m)!}{(l+m)!} \right]^{1/2} P_l^m(\cos \beta). \quad (2)$$

Finally, for the rotated and translated version of  $g$ , we obtain

$$(T_\rho \Lambda_R g)(\mathbf{r}\mathbf{u}) = (\Lambda_R g)(\mathbf{r}'\mathbf{u}') = \sum_{l,m,n} \hat{g}_{ln}(r') D_{mn}^l(R') Y_{lm}(\mathbf{u}').$$

### 3. Methods

The correlation will be a function of two rotations  $R, R'$  and of the distance  $\rho$  that the second object is translated along the positive  $z$  axis (Fig. 1):

$$c(R, R'; \rho) = \int_{\mathbb{R}^3} \overline{\Lambda_R f} \cdot \overline{T_\rho \Lambda_{R'} g}.$$

(In order to obtain the correct sign in the exponential of the Fourier transform, we perform the complex conjugation, denoted by overlines, of the two factors.)

By using the expressions given in the previous section, we obtain,

$$c(R, R'; \rho) = \sum_{ll'mm'nn'} \overline{D_{nn'}^{l'}(R) D_{mm'}^{l'}(R')} \int_{\mathbb{R}^3} \hat{f}_{lm}(r) \hat{g}_{l'm'}(r') Y_{ln}(\mathbf{u}) Y_{l'n'}(\mathbf{u}').$$

Taking into account (2), the integral reduces to  $(-1)^n \delta_{-n,n'} I_{nnmm'}^{ll'}(\rho)$ , where  $\delta$  denotes the Kronecker delta function and

$$I_{nnmm'}^{ll'}(\rho) = \left[ (l + \frac{1}{2})(l' + \frac{1}{2}) \right]^{1/2} \cdot \int_0^\pi \int_0^\infty \overline{\hat{f}_{lm}(r) \hat{g}_{l'm'}(r')} d_{n0}^{l'}(\beta') r^2 dr \cdot d_{n0}^{l'}(\beta) \sin \beta d\beta.$$

We now factorize the rotations  $R$  and  $R'$  (cf. Kovacs & Wriggers, 2002). If, in terms of Euler angles,  $R = (\varphi, \theta, \psi)$  and  $R' = (\varphi', \theta', \psi')$ , then  $R = R_1 \cdot R_2$  and  $R' = R'_1 \cdot R'_2$ , with

$$R_1 = (\xi, \frac{\pi}{2}, 0), \quad R_2 = (\eta, \frac{\pi}{2}, \omega), \\ R'_1 = (\xi', \frac{\pi}{2}, 0), \quad R'_2 = (\eta', \frac{\pi}{2}, \omega'),$$

where

$$\xi = \varphi - \frac{\pi}{2}, \quad \eta = \pi - \theta, \quad \omega = \psi - \frac{\pi}{2}, \\ \xi' = \varphi' - \frac{\pi}{2}, \quad \eta' = \pi - \theta', \quad \omega' = \psi' - \frac{\pi}{2}.$$

Using (1) and the identity

$$D_{nm}^l(R_1 \cdot R_2) = \sum_h D_{nh}^l(R_1) D_{hm}^l(R_2),$$

(which expresses the homomorphic nature of representations) we obtain

$$D_{nm}^l(R) = \sum_h d_{nh}^l d_{hm}^l \exp[-i(n\xi + h\eta + m\omega)],$$

where for brevity we denote

$$d_{mn}^l := d_{mn}^l(\frac{\pi}{2}). \quad (3)$$

Similarly,  $D_{n'm'}^{l'}(R') = \sum_{h'} d_{n'h'}^{l'} d_{h'm'}^{l'} \exp[-i(n'\xi' + h'\eta' + m'\omega')]$ . Hence, on letting  $\sigma = \xi - \xi'$ , the correlation function becomes

$$c(R, R'; \rho) = \sum_{ll'mm'nn'} (-1)^n d_{nh}^l d_{hm}^l d_{-nh'}^{l'} d_{h'm'}^{l'} \\ \times \exp[i(n\sigma + h\eta + m\omega + h'\eta' + m'\omega')] I_{nnmm'}^{ll'}(\rho) \\ =: T(\sigma, \eta, \omega, \eta', \omega'; \rho).$$

We call  $T$ , as a function of the five angles  $\sigma, \eta, \omega, \eta', \omega'$  and the parameter  $\rho$ , the rotational correlation function (RCF).

From this, it is apparent that the Fourier transform of the RCF is

$$\hat{T}(n, h, m, h', m'; \rho) = (-1)^n \sum_{l,l'} d_{nh}^l d_{hm}^l d_{-nh'}^{l'} d_{h'm'}^{l'} I_{nnmm'}^{ll'}(\rho). \quad (4)$$

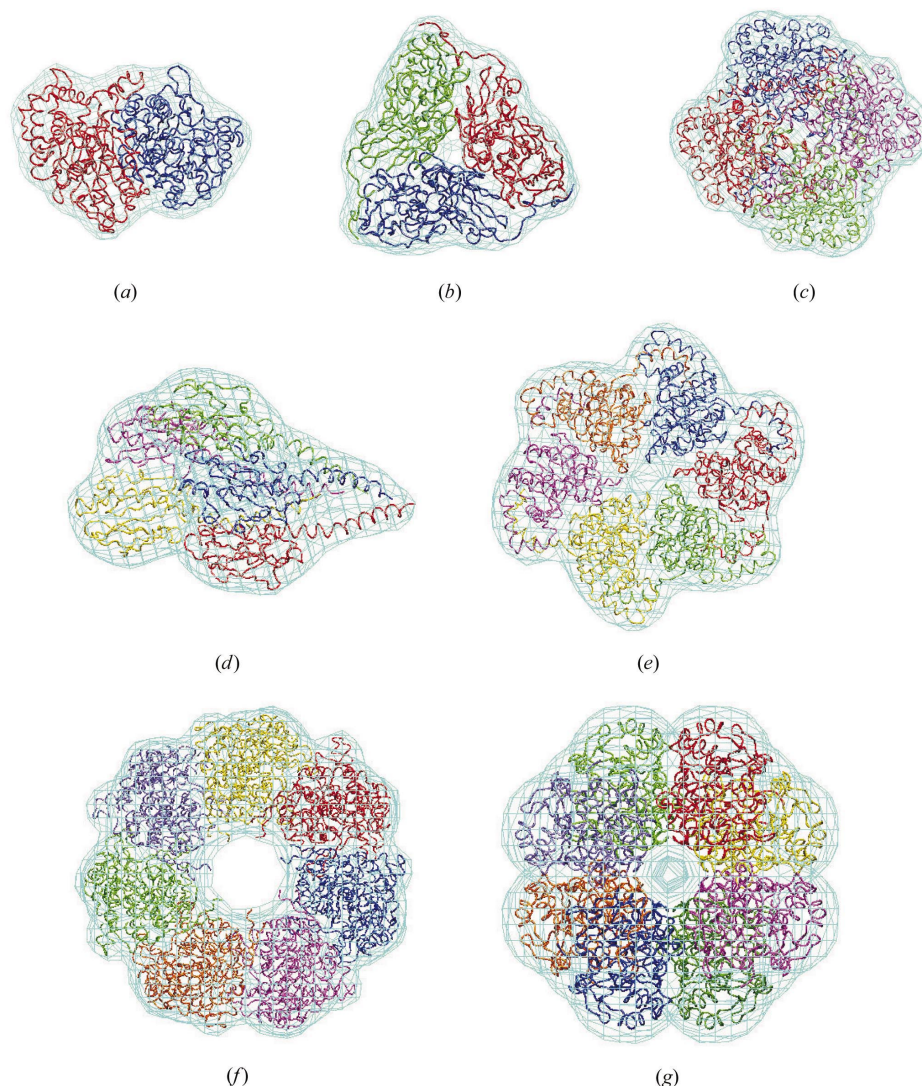
For each  $\rho$ , a five-dimensional inverse Fast Fourier Transform then yields the RCF on a grid in  $(\sigma, \eta, \omega, \eta', \omega')$  space. Implementation details are given in Appendix A.

**Table 1**

Comparison between FRM and FTM for a particular set of parameters:  $B = 16$ , resolution = 15 Å and voxel size = 3 Å.

The first column indicates the PDB codes of the molecules considered (Fig. 2 gives details). The second column shows the numbers of voxels of the corresponding low-resolution maps. The third column gives the number of cycles (that is, of  $\rho$  values) that FRM performed in each case. Times shown correspond to the exhaustive search only; they do not include any subsequent off-lattice refinement of the solutions.

PDB code	No. of voxels	FRM		FTM time	Time ratio
		Cycles	Time (mins)		
1afw	75465	6	11:0	59 min 41 s	5.43
1nic	103823	7	12:41	30 min 37 s	2.41
2pil	114165	7	12:56	1 h 49 min	8.45
7cat	137445	8	13:59	1 h 57 min	8.36
1aw5	148225	9	14:55	1 h 9 min	4.60
1e0j	171349	12	22:8	1 h 9 min	3.14
1der	357627	10	19:7	5 h 21 min	16.8

**Figure 2**

Atomic structures used for the timings and the envelopes of their corresponding synthetic EM densities (in light blue): (a) 1afw (peroxisomal thiolase, dimer); (b) 1nic (copper-nitrite reductase, trimer); (c) 7cat (catalase, tetramer); (d) 2pil (type 4 pilin, pentamer); (e) 1e0j (Gp4D helicase, hexamer); (f) 1der (chaperonin GroEL, heptamer); (g) 1aw5 (5-aminolevulinate dehydratase, octamer).

## 4. Results

We have tested our method by applying it to the docking of atomic structures of monomeric subunits inside EM maps of the corresponding oligomers. These EM maps were synthetically generated by lowering the resolution of the atomic structures of the oligomers (see Chacón & Wriggers, 2002) so that the method described in the previous section can be applied. The docking of the monomers was performed for seven different oligomeric molecules (listed in Table 1 and shown in Fig. 2), for three values of resolution (10, 15 and 20 Å) and for five voxel sizes (2, 3, 4, 5 and 6 Å). Owing to the use of the Laplacian filter, the results in all cases were found to lie within one grid unit from the theoretical solutions, as was found in earlier accuracy evaluations with the FTM algorithm (Chacón & Wriggers, 2002).

Table 1 shows timings (for typical values of resolution and voxel size) corresponding to the above experiment performed with FRM and with FTM. This table shows a clear relationship between the number of cycles ( $\rho$  values) that FRM performed and the number of voxels of the maps. This trend has an exception in the case of PDB entry 1der, because this molecule has a relatively large hole in the center which reduces the number of  $\rho$  values to be scanned. The FTM times are variable because they depend on the prime factorization of the grid size, which influences the performance of the FFTW algorithm (Frigo & Johnson, 1998). The runs were performed on a 1.6 GHz AMD Athlon 1900+ PC under Linux with 512 MB RAM.

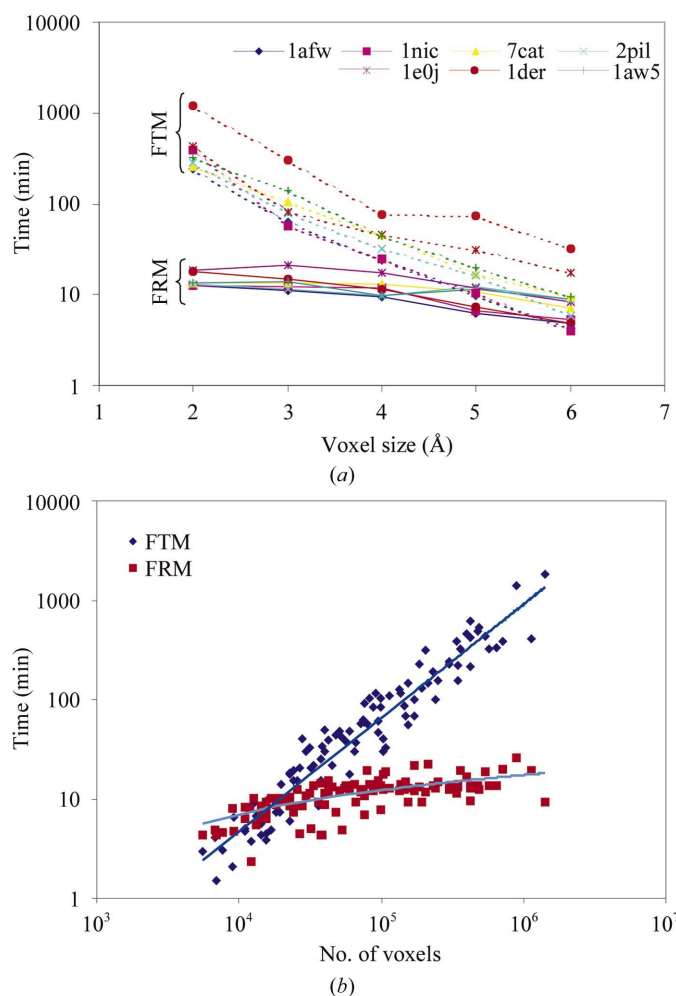
In Fig. 3(a) we collect timing data for the different molecules as functions of the voxel size. For this plot the three values of resolution (10, 15 and 20 Å) were averaged out to make the figure more readable. In all cases we worked with bandwidth  $B = 16$  (i.e. 32 sampling points for each angular parameter or  $\sim 11^\circ$  sampling). We can see that for small voxel sizes FRM is advantageous and also that its timings are much less dependent on the voxel size than the FTM timings are.

Fig. 3(b) is a scatter plot where each dot corresponds to one of the 105 cases tested (seven mole-

cules, three resolutions, five voxel sizes). We fitted regression curves of the form  $t \simeq N^\alpha \log N$  to both the FTM and the FRM plots (where  $t$  is the time and  $N$  is the number of voxels). We obtained  $\alpha_{\text{FTM}} = 1.1$  and  $\alpha_{\text{FRM}} = 0.09$ . Again, we see a stronger dependence of FTM than FRM time on the number of voxels of the maps. In practice, we could say that in 10–15 min FRM can handle most six-dimensional docking problems.

## 5. Conclusions

By recasting the six-dimensional matching problem into a formulation involving five angular DOFs and just one linear DOF and by applying the idea of the original three-dimensional FRM method (Kovacs & Wriggers, 2002), we are able to FFT-accelerate five out of six DOFs of the problem. As we saw in the previous section, this approach results in the running times being much more independent of the problem sizes, making FRM a method with a wide range of applicability. This is related to the convenient fact that our FFTs are performed over angular variables rather than linear ones,



**Figure 3**  
(a) Times for FRM and FTM for various molecules *versus* the voxel size of the low-resolution maps. (b) Timing trends for FRM and FTM *versus* the number of voxels of the low-resolution maps.

eliminating the need for ‘padding’ the grid with zeros to avoid boundary effects (see Press *et al.*, 1992).

We see in Fig. 3 that FRM has a better performance than FTM for small voxel size (or large number of voxels), as is typical in commonplace EM maps. The variability of the FTM points in Fig. 3(b) arises from the variation in the performance of the FFTW (Frigo & Johnson, 1998) algorithm used, which depends on the prime decomposition of the array size. On the other hand, the variability of the FRM points is mainly a consequence of how far the initial  $\rho$  value lies from the best solution.

We have made all tests of our algorithm using 32 samplings (*i.e.*  $\sim 11^\circ$  sampling) for each of the angular parameters, which means that the number of harmonics is  $B = 16$ . The results of our tests show that the accuracy for EM docking applications is comparable to that of FTM. (This is to be expected owing to the low resolution of the maps,  $\geq 10$  Å.) A higher  $B$  value would be needed in cases where the sizes of the two molecules are very different from each other.

The reason for not taking a higher value of  $B$  in our tests is memory limitations in our machine. The storage required by the code is  $\sim 128(B^5 + B^4)$  bytes (neglecting contributions of lower order). This is used to store the correlation function and its Fourier transform for each  $\rho$ . For  $B = 16$  this is about 150 MB. Now, we are restricted to using a power of 2 for  $B$ , since the algorithm that we use to compute the spherical harmonic coefficients, *SpharmonicKit* (Healy *et al.*, 1998), works only under this condition. The next power of 2,  $B = 32$ , would require about 4.5 GB. Since memory costs are dropping steadily, a value of  $B = 32$  will soon come within reach of standard personal computers.

Having the possibility to go to higher values of  $B$ , it would then be sensible to let  $B$  be a function of  $r$  when the spherical samplings are performed. A linear function would be reasonable, so that the spatial density of the sampling points would be roughly uniform. This would imply more spherical coefficients  $\hat{f}_{lm}(r)$  and  $\hat{g}_{lm}(r)$  for larger  $r$ . This, in turn, would make the two-center integrals  $I_{mm'}^{ll'}(\rho)$  faster to compute than if one used the maximum  $B$  for all  $r$ . In addition, the smaller  $\rho$  is, the smaller the range of  $l$  and  $l'$  (otherwise the integrals vanish), hence the smaller the range of the indices in  $\hat{T}(n, h, m, h', m'; \rho)$  and the faster the computation of the inverse FFT. Thus, a variable  $B$  would yield a more cost-effective method.

There are important special cases where the gain in performance by FRM over FTM would be particularly high: for spherical molecular shapes (*e.g.* viruses) and for toroidal molecular shapes (*cf.* the case of PDB entry 1der, for which the speed-up is almost 17 times). This is so because in these cases the range of  $\rho$  is small relative to the size of the molecules.

We note that FRM can be adapted to the ‘exterior docking’ (or ‘complementary surface’) problem, specifically applicable to protein–protein or ligand–receptor binding-site determination, which is of importance in drug design. For this application FRM would also have a noticeable advantage over FTM, since the latter would need a large grid to accommodate both molecules. Finally, we note that our three-dimensional

approach can also be adapted to the two-dimensional matching problem in image processing by replacing the spherical harmonic expansions by Fourier expansions. The resulting 'two-dimensional FRM' method might prove advantageous in applications that require each image to be matched many times, such as the alignment problem in EM (Cong, Kovacs & Wriggers, work to be published).

## APPENDIX A

### Implementation details

Here we describe how we implemented (4) efficiently by scanning the linear parameter  $\rho$ .

The scanning range of  $\rho$  is restricted according to the sizes of the objects. We define, for each object, the following 'radii':

$r_{\min,e}$	} min and max distances from COM to star
$r_{\max,e}$	
$r_{\min,i}$	} min and max distances from COM to
$r_{\max,i}$	
	points of zero density
$r'_{\min,i}$	} min and max distances from COM to
$r'_{\max,i}$	
	points of nonzero density.

In terms of these, we define  $\rho_{\min}$ ,  $\rho_{\max}$  (minimum and maximum values allowed) and  $\rho_{\text{ini}}$  (initial value):

$$\rho_{\min} = \begin{cases} \max\{r_{\min,i}^{(f)} - r_{\max,i}^{(g)}, 0\} & \text{if } r_{\min,i}^{(f)} > 0 \text{ and } r_{\min,i}^{(g)} > 0 \\ r_{\min,i}^{(f)} + r_{\min,i}^{(g)} & \text{if } r_{\min,i}^{(f)} > 0 \text{ and } r_{\min,i}^{(g)} = 0 \\ \max\{r_{\min,i}^{(f)} - r_{\max,e}^{(g)}, 0\} & \text{if } r_{\min,i}^{(f)} = 0 \end{cases}$$

$$\rho_{\max} = r_{\max,e}^{(f)} - r_{\min,e}^{(g)}$$

$$\rho_{\text{ini}} = \begin{cases} 0 & \text{if } r_{\min,i}^{(f)} > 0 \text{ and } r_{\min,i}^{(g)} > 0 \\ (\rho_{\min} + \rho_{\max})/2 & \text{if } r_{\min,i}^{(f)} > 0 \text{ and } r_{\min,i}^{(g)} = 0 \\ \max\{[r_{\max,e}^{(f)} + r_{\min,e}^{(g)} - r_{\max,e}^{(g)} - r_{\min,e}^{(f)}]/2, 0\} & \text{if } r_{\min,i}^{(f)} = 0. \end{cases}$$

The process starts by computing the correlation function for  $\rho = \rho_{\text{ini}}$  and then for values of  $\rho$  to the left and to the right of it, storing the peaks of the correlation function and the corresponding  $\rho$  values. The step size for  $\rho$  is taken to be  $h = \max\{1, \min\{r_{\max,e}^{(f)}, r_{\max,e}^{(g)}\}/B\}$ , so that it is consistent with the angular spacing. This is also the step size used in the numerical evaluation of the two-center integrals which, for this purpose, are written in the following way:

$$I_{mm'}^{ll'}(\rho) = \left[ (l + \frac{1}{2})(l' + \frac{1}{2}) \right]^{1/2} \cdot \int_0^\infty \left[ \int_{-\infty}^\infty \hat{f}_{lm}(r) \hat{g}_{l'm'}(r') d_{n0}^l(\beta) d_{n0}^{l'}(\beta') dz \right] \cdot s ds,$$

where the variables  $s$ ,  $z$  are related to  $r$ ,  $\beta$  by  $s = r \sin \beta$ ,  $z = r \cos \beta$ .

During this process the program keeps track of the maximum correlation value that has been encountered and the corresponding  $\rho$ , continuing to compute to the left and to the right of that  $\rho$  until the highest peaks at both sides are smaller than 90% of the maximum.

In order to produce accurate and high-contrast results, we pre-filter the maps with a Laplacian kernel, as is performed in Chacón & Wriggers (2002).

Support by the La Jolla Interfaces in Science Program/Burroughs Wellcome Fund and NIH grant 1R01 GM62968 is gratefully acknowledged.

## References

- Brink, D. M. & Satchler, G. R. (1993). *Angular Momentum*. Clarendon Press.
- Chacón, P. & Wriggers, W. (2002). *J. Mol. Biol.* **317**, 375–384.
- Frank, J. (1996). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. San Diego: Academic Press.
- Frigo, M. & Johnson, S. G. (1998). *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1381–1384. Seattle, Washington: IEEE.
- Healy, D. Jr, Rockmore, D., Kostelec, P. & Moore, S. (1998). *FFTs for the Two-sphere – Improvements and Variations*. Preprint and software available at <http://www.cs.dartmouth.edu/~geelong/sphere/>.
- Hobson, E. W. (1931). *The Theory of Spherical and Ellipsoidal Harmonics*. Cambridge University Press.
- Kovacs, J. A. & Wriggers, W. (2002). *Acta Cryst.* **D58**, 1282–1286.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Paquet, E., Rioux, M., Murching, A., Naveen, T. & Tabatabai, A. (2000). *Signal Proc. Image Commun.* **16**, 103–122.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C. The Art of Scientific Computing*, 2nd ed, ch. 13. New York: Cambridge University Press.
- Ritchie, D. W. & Kemp, G. J. L. (2000). *Proteins Struct. Funct. Genet.* **39**, 178–194.
- Shams, L. B., Brady, M. J. & Schaal, S. (2001). *Neural Networks*, **14**, 345–354.
- Siddiqi, K., Shokoufandeh, A., Dickinson, S. J. & Zucker, S. W. (1999). *Int. J. Comput. Vision*, **35**, 13–32.
- Weisstein, E. W. (1999). *CRC Concise Encyclopedia of Mathematics*. Boca Raton, FL: CRC Press.
- Wriggers, W. & Chacón, P. (2001). *Structure*, **9**, 779–788.
- Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). *J. Struct. Biol.* **125**, 185–195.