

Fast rotational matching

Julio A. Kovacs* and Willy
Wriggers

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

Correspondence e-mail: jkovacs@scripps.edu

A computationally efficient method is presented – ‘fast rotational matching’ or FRM – that significantly accelerates the search of the three rotational degrees of freedom (DOF) in biomolecular matching problems. This method uses a suitable parametrization of the three-dimensional rotation group along with spherical harmonics, which allows efficient computation of the Fourier Transform of the rotational correlation function. Previous methods have used Fourier techniques only for two of the rotational DOFs, leaving the remaining angle to be determined by an exhaustive search. Here for the first time a formulation is presented that makes it possible to Fourier transform all three rotational DOFs, resulting in notable improvements in speed. Applications to the docking of atomic structures into electron-microscopy maps and the molecular-replacement problem in X-ray crystallography are considered.

Received 19 March 2002

Accepted 30 May 2002

1. Introduction

In various scientific computing situations the problem arises of finding a best rotation to match two given volumetric density objects. Examples have been given in diverse fields such as pattern recognition (Shams *et al.*, 2001) and image processing (Russ, 1998), as well as engineering (Paquet *et al.*, 2000), machine vision (Siddiqi *et al.*, 1999) and biophysics (Wriggers *et al.*, 1999). Our analysis below is restricted to examples in concurrent structural biology research where three-dimensional rotational matching is the limiting step, such as biomolecular docking (Wriggers & Chacón, 2001) and the use of the molecular-replacement method (Drenth, 1999), but generalizations are possible to other disciplines.

We discuss two structural biology applications: molecular docking (better known as ‘real-space molecular replacement’ in the crystallography community) and molecular replacement. The latter application arises in X-ray crystallography where the goal is to estimate the missing phases from the observed structure-factor amplitudes (Drenth, 1999) and a model structure. Since the structure-factor amplitudes and the Fourier transform of their squares – the Patterson function – are translation-invariant, it is often simpler to first obtain the rotational match of the Patterson function of the probe molecule with the observed data before identifying its position in the unit cell. The correlation coefficient of Patterson functions (as a function of rotational space) is typically called ‘rotation function’ in crystallography. The aim is to find the angles that maximize that function.

Originally, the rotational match between two objects was computed directly by an exhaustive scan of the three rota-

tional DOFs (typically expressed in Euler angles; Weisstein, 1999). This slow approach is still sometimes used, but a faster alternative was developed by Crowther, who first realised that the cross-correlation of the objects could be computed efficiently with Fast Fourier Transforms (FFTs) if the objects were expressed in spherical harmonics (Crowther, 1972). This rotational-space method is analogous to the better-known acceleration of translational search using FFTs on the basis of the convolution theorem (Bracewell, 1986). The calculation of the correlation in direct space is very expensive, since it requires $O(M^2)$ operations, where M is the number of sampling points, whereas by means of FFTs the number of operations scales as $M \log M$ (Katchalski-Katzir *et al.*, 1992). Crowther's 'fast rotation function' method succeeded in FFT-accelerating two of the three rotational DOFs, but one angle was still subject to the traditional exhaustive search. The main innovation presented in this work is a full FFT in all three rotational DOFs. Comparison timings show that FRM performs substantially faster than previously known methods.

2. Preliminaries

Let two finite-size objects in 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 – that is, to find a rigid motion of space that produces the best overlap between the two functions – 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.

First we introduce some notation. For a rotation R in the three-dimensional rotation group $SO(3)$, let Λ_R be the rotation operator defined by

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

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: $S^2 = \{|\mathbf{u}| \in \mathbb{R}^3 \mid |\mathbf{u}| = 1\}$.

Let $Y_{lm}: S^2 \rightarrow \mathbb{C}$ be the spherical harmonic functions, where $l \geq 0$ and $-l \leq m \leq l$. These can be written in terms of the associated Legendre functions P_l^m ,

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

Details about the spherical harmonics can be found in Hobson (1931).

Square-integrable functions on the sphere can be expanded in series of spherical harmonics (Hobson, 1931), since these form an orthonormal basis for the space of those functions. Therefore, our functions f and g can be approximated by finite sums,

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

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

where $\hat{f}_{lm}(r)$ and $\hat{g}_{lm}(r)$ are the spherical harmonic coefficients of the restrictions of f and g to the sphere of radius r and the number B of terms is related to the angular sampling for the Euler angles (see later).

For a rotation R we have

$$(\Lambda_R g)(r\mathbf{u}) = g[R^{-1}(r\mathbf{u})] = g[rR^{-1}(\mathbf{u})]$$

$$= \sum_{l,m} \hat{g}_{lm}(r) Y_{lm}[R^{-1}(\mathbf{u})].$$

Under rotations, spherical harmonics transform into linear combinations of others of the same degree,

$$Y_{lm}[R^{-1}(\mathbf{u})] = \sum_n D_{nm}^l(R) Y_{ln}(\mathbf{u}),$$

where the $D_{nm}^l(R)$ are the matrix elements of the irreducible representations of $SO(3)$ (Brink & Satchler, 1993). In order to represent rotations, we use Euler angles with the 'ZYZ' convention (Brink & Satchler, 1993) or 'y' convention (Weisstein, 1999); namely, $R(\varphi, \theta, \psi)$ will mean rotate by ψ about the original z axis, then by θ about the original y axis and finally by φ about the original z axis.¹ Using these Euler angles, the D_{nm}^l can be written as

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

where the functions d_{mn}^l are real. An explicit expression for these functions can be found in Brink & Satchler (1993), but since their direct computation is rather expensive, we use an efficient recursive procedure given in Risbo (1996). Therefore, the rotated version of g is written as

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

3. Purely rotational search

Here, we assume that the objects do not need to be translated with respect to one another, so what is sought for is the rotation of one of them that maximizes its overlap with the other. The correlation in this case will be a function of a rotation only.

3.1. Volumetric approach

In this case, the correlation function $c: SO(3) \rightarrow \mathbb{R}$ is written as

$$c(R) = \int_{\mathbb{R}^3} f \cdot \overline{\Lambda_R g}.$$

¹ This convention is the same as that used by Crowther (1972) and in programs in the CCP4 suite such as *AMoRe* (Navaza, 1994) and *MOLREP* (Vagin & Teplyakov, 1997).

(The overline indicates complex conjugation; we choose to formally conjugate the real argument so that later we will not need to make an additional change of variables.)

By substituting the expansions for f and Λ_{Rg} , we obtain

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

Owing to the orthogonality of the spherical harmonics, the integral reduces to

$$\delta_{l'l'} \delta_{mm'} I_{mm'}^l,$$

where δ denotes the Kronecker delta function and

$$I_{mm'}^l = \int_0^\infty \hat{f}_{lm}(r) \overline{\hat{g}_{l'm'}(r)} r^2 dr.$$

We will now factorize the rotation R . Say that, in terms of Euler angles, $R = (\varphi, \theta, \psi)$. It is easy to check that $R = R_1 \cdot R_2$, with

$$R_1 = (\xi, \pi/2, 0), \quad R_2 = (\eta, \pi/2, \omega),$$

where the new parameters are²

$$\xi = \varphi - \pi/2, \quad \eta = \pi - \theta, \quad \omega = \psi - \pi/2.$$

Using (1) and the fact that

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

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

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

Thus, the correlation function becomes

$$c(R) = \sum_{lmhnm'} d_{mh}^l d_{hm'}^l I_{mm'}^l \exp[i(m\xi + h\eta + m'\omega)] \\ =: T(\xi, \eta, \omega).$$

We call T , as a function of the angles ξ, η, ω , the *rotational correlation function* (RCF).

Hence, we see that the Fourier transform of T is

$$\hat{T}(m, h, m') = \sum_l d_{mh}^l d_{hm'}^l I_{mm'}^l. \quad (3)$$

An inverse Fast Fourier Transform then yields the RCF on a grid in (ξ, η, ω) space.

We applied this approach to the well known molecular-replacement problem. Crowther (1972) presented an efficient way to compute the correlation function of two Patterson densities. This matter falls exactly in the case considered here. Our approach thus takes advantage of the factorization of the rotations [or equivalently, the change of parameters $(\varphi, \theta, \psi) \rightarrow (\xi, \eta, \omega)$] in order to express the rotation function as a full Fourier series in the three angular variables (3), rather

² Since $(\varphi, \theta, \varphi)$ is equivalent to $(\psi + \pi, -\theta, \varphi + \pi)$, this transformation can also be written as an origin shift: $\xi = \varphi + \pi/2, \eta = \theta + \pi, \omega = \psi + \pi/2$.

Table 1
Times (s) for rotational matching using Crowther's method and FRM.

B	Crowther		FRM	
	Precomp.	Rest	Precomp.	Rest
32	0.003	1.66	0.03	0.97
64	0.019	19.3	0.24	3.75
128	0.145	337.0	2.60	37.4

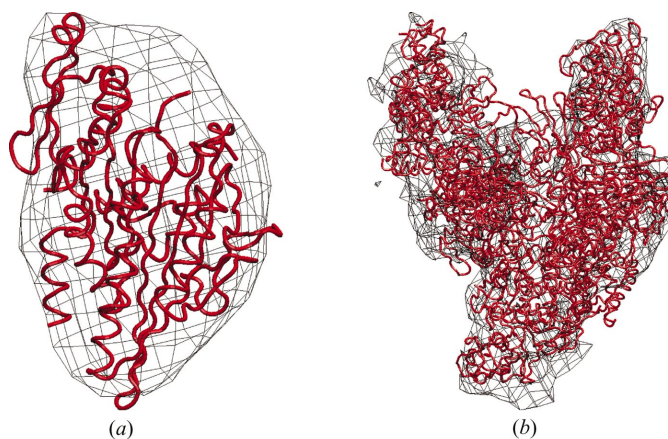


Figure 1
Atomic structures (red) used for the timings and the envelopes of their corresponding electron-microscopy densities (black): (a) ncd (Wriggers *et al.*, 1999) and (b) RNA polymerase (Darst *et al.*, 2002).

than a two-dimensional Fourier series for each value of one of the variables (θ).

We have made a timing comparison between our approach and a 'Crowther-like' modification of it in which the computation of the $d_{mm'}^l(\theta)$ is performed on the fly (Crowther, 1972), as is the case in packages such as *AMoRe* (Navaza, 1994) and *MOLREP* (Vagin & Teplyakov, 1997). Results for the case of the kinesin-related motor protein ncd (Wriggers *et al.*, 1999) are shown in Table 1. Specifically, we searched for the best superposition of two 20 Å resolution density maps given on a $7 \times 9 \times 13$ grid with a voxel size of 6 Å: one of the maps is an experimental electron-microscopy density map, while the other was generated by blurring the atomic structure of ncd (Fig. 1a). The integrals $I_{mm'}^l$ are evaluated numerically, using as upper limit of integration the maximum radius for which the densities have non-zero values. The coefficients d_{mn}^l are precomputed and stored in an array at the beginning of the program. The times shown correspond to a 1 GHz Pentium III Linux PC. We see that the gain in speed ranges from 66% for coarse angular sampling ($6^\circ, B = 32$) through 384% for medium ($3^\circ, B = 64$) to 743% for fine sampling ($1.4^\circ, B = 128$). (Using precomputations, the percentages are 71, 415 and 800%, respectively.)

3.2. Radial approach

In the purely rotational case we are considering here, it is often possible to use radial functions rather than volumetric

functions, especially when dealing with electron-density maps. Namely, if V denotes the boundary of the object represented by f , for instance given by one of its isocontour surfaces,

$$V = \{\mathbf{p} \in \mathbb{R}^3 \mid f(\mathbf{p}) = \alpha\},$$

then the radialized representation of the object (Ritchie & Kemp, 1999) will be given by the function $\tilde{f} : S^2 \rightarrow \mathbb{R}$ defined by

$$\tilde{f}(\mathbf{u}) = \max\{r \geq 0 \mid r\mathbf{u} + \mathbf{C} \in V \cup \{\mathbf{C}\}\},$$

where \mathbf{C} is the center of mass of f .

Clearly, the 'radialization process', illustrated in Fig. 2, may introduce distortions in the shape of the object if the boundary V is not star-shaped with respect to \mathbf{C} . However, roughly the same distortion is produced in both objects, so in principle this should not make the method break down. In fact, we performed tests with nine molecules of diverse shapes (lysozyme, barnase, ribonuclease inhibitor, human serum albumin, lethal factor precursor, haloperoxidase, urease γ subunit, cholera toxin and RNA polymerase) by generating low-resolution maps of the atomic structures and obtained average fitting errors well under 1 Å r.m.s.d. even at resolutions as low as 60 Å.

For brevity we shall omit the $\tilde{}$ and denote with f and g the radial functions corresponding to the given objects. The correlation is now

$$c(R) = \int_{S^2} f \cdot \overline{\Lambda_R g}.$$

Expanding f and g in spherical harmonics,

$$f = \sum_{lm} \hat{f}_{lm} Y_{lm}, \quad g = \sum_{lm} \hat{g}_{lm} Y_{lm},$$

the correlation becomes

$$c(R) = \sum_{lmhm'} d_{mh}^l d_{hm'}^l \hat{f}_{lm} \overline{\hat{g}_{lm'}} \exp[i(m\xi + h\eta + m'\omega)] =: T(\xi, \eta, \omega),$$

where, as before, $\xi = \varphi - \pi/2$, $\eta = \pi - \theta$ and $\omega = \psi - \pi/2$. Hence,

$$\hat{T}(m, h, m') = \sum_l d_{mh}^l d_{hm'}^l \hat{f}_{lm} \overline{\hat{g}_{lm'}}.$$

In Table 2 we show the timings of this method when applied to the two molecules presented in Fig. 1, for which experimental electron-microscopy maps were available to us: ncd (described previously: Wriggers *et al.*, 1999) and RNA polymerase (Darst *et al.*, 2002). The latter is a 15 Å resolution map given on a $43 \times 49 \times 45$ grid with a voxel size of 3 Å. The times correspond to a 1 GHz Pentium III Linux PC. $B = 16$ and $B = 32$ correspond, approximately, to angular sampling steps (Table 2) of 12 and 6°, respectively. The difference between the FRM matching results corresponding to both samplings is negligible, showing that $B = 16$ is a sufficient number of spherical harmonics for this application. We also compared the performance of our method to an exhaustive search of all Euler angles and to a docking based on landmark pointers (Wriggers *et al.*, 1999). In the latter procedure, the objects are represented at reduced complexity using $n = 7$ and $n = 8$ landmarks;

Table 2

Comparison times for radial FRM, for landmark matching (LM) and for standard exhaustive search (ES).

	$B = 32$		$B = 16$	
	ncd	RNAp	ncd	RNAp
FRM (ms)	390	560	54	117
LM (s)	55	76	45	68
ES	13 min	7 h	96 s	54 min

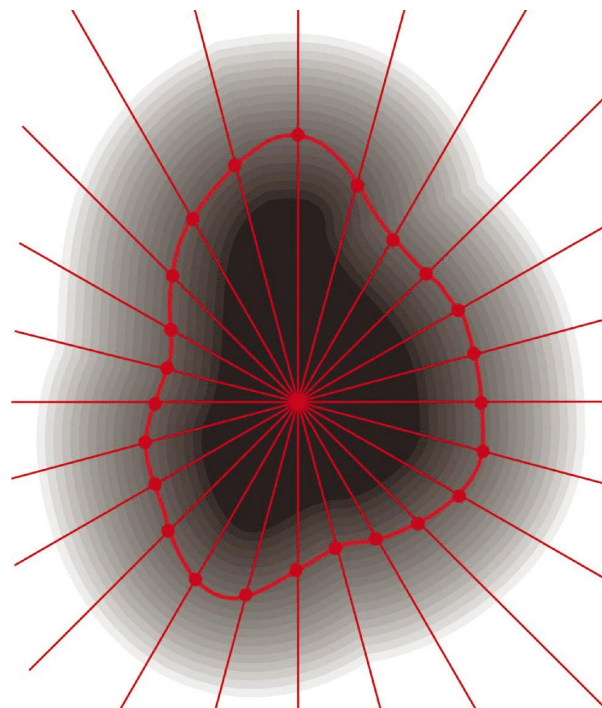


Figure 2

Schematic picture depicting the radialization process.

the subsequent exhaustive search of $n!$ possible combinations corresponds to a sampling roughly comparable to that achieved with $B = 16$ and $B = 32$ harmonics, respectively.

The calculation of the spherical harmonic coefficients has been implemented using *SpharmonicKit* (Healy *et al.*, 1998), a collection of C source code that computes discrete Legendre transforms and full spherical harmonic transforms. The algorithm developed by these authors builds on previous work of Driscoll & Healy (1994).

4. Conclusions

As shown above, the optimal rotational superposition of two molecular-density objects can be found by expressing the objects as series of spherical harmonics and by expressing the rotation function as a Fourier series in the three angular variables. A novel factorization of the rotations reduces the computation load of the d -functions to the single argument $\pi/2$

(2) and yields a simple formula for the Fourier transform of the rotation function itself (3). The efficiency gained thereby is almost one order of magnitude relative to the Crowther method for fine angular sampling ($B = 128$) (Table 1).

The matching on the millisecond timescale with the radial approach makes it possible to compute rotational fits instantaneously in interactive modeling sessions. Our earlier work on interactive docking prompted us to reduce the complexity of the search space using clustering techniques (Wriggers *et al.*, 1998) that encode original data by representative landmarks. In comparison, radial FRM matches the isocontour levels of the original data directly and is two to three orders of magnitudes faster than landmark matching (Table 2), possibly at the expense of a reduced fitting contrast owing to the use of the less stringent correlation criterion (Wriggers *et al.*, 1999). However, the contrast can be recovered by pre-filtering the maps, for instance with a Laplacian filter (Chacón & Wriggers, 2002).

Since FRM relies on the superposition of the centers of mass, it will be the subject of future research to validate the robustness of the method under experimental noise and positional misalignments. This problem will be most efficiently addressed in the context of an FFT-accelerated six-dimensional search (J. Kovacs, Y. Cong, P. Chacón, E. Metwally & W. Wriggers, in preparation).

When limited to rotational-space search only, we see advantages of our method, in particular for very fine sampling of the angular variables, where computation of the d -functions for the Crowther method is unfavorable owing to prohibitive computational costs (on the fly) or prohibitive memory requirements (when precomputed).³ The improvement in speed could permit some progress on the problem of solving for the best match when there are conformational changes in the fitted structures (*e.g.* hinge angles in proteins with flexible domains) or incomplete models – cases that can be notoriously difficult at present in biomolecular docking and molecular-replacement applications. In particular, one could envision approaches in which every protein in the PDB is used in succession to attempt to solve the structure of a new protein, especially in the context of structural genomics.

³ In fact, Crowther's 'precomputed' approach could not be carried out for $B = 128$ on a 512 MB memory machine.

We thank Yao Cong, Essam Metwally, Pablo Chacón and Lynn F. Ten Eyck for discussions. Support by the La Jolla Interfaces in Science Program/Burroughs Wellcome Fund (JK) and NIH grant 1R01 GM62968 (WW) is gratefully acknowledged.

References

- Bracewell, R. N. (1986). *The Fourier Transform and its Applications*, 2nd ed. New York: McGraw-Hill.
- Brink, D. M. & Satchler, G. R. (1993). *Angular Momentum*. Oxford: Clarendon Press.
- Chacón, P. & Wriggers, W. (2002). *J. Mol. Biol.* **317**, 375–384.
- Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. Gordon & Breach.
- Darst, S., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G. & Wriggers, W. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 4296–4301.
- Drenth, J. (1999). *Principles of Protein X-ray Crystallography*, 2nd ed. New York: Springer Verlag.
- Driscoll, J. R. & Healy, D. M. Jr (1994). *Adv. Appl. Math.* **15**, 202–250.
- Healy, D. Jr, Rockmore, D., Kostelec, P. & Moore, S. (1998). *Fast Spherical Transforms: Spharmonic Kit*. <http://www.cs.dartmouth.edu/~geelong/sphere>.
- Hobson, E. W. (1931). *The Theory of Spherical and Ellipsoidal Harmonics*. Cambridge University Press.
- Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Freisem, A. A., Aflalo, C. & Vakser, I. A. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 2195–2199.
- 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.
- Risbo, T. (1996). *J. Geodesy*, **70**, 383–396.
- Ritchie, D. W. & Kemp, G. J. L. (1999). *J. Comput. Chem.* **20**, 383–395.
- Russ, J. C. (1998). *The Image Processing Handbook*, 3rd ed. Boca Raton, FL, USA: CRC Press.
- Shams, L. B., Brady, M. J. & Schaal, S. (2001). *Neural Networks*, **14**, 345–354.
- Siddiqi, K., Shoukroufandeh, A., Dickinson, S. J. & Zucker, S. W. (1999). *Int. J. Comput. Vis.* **35**, 13–32.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Weisstein, E. W. (1999). *CRC Concise Encyclopedia of Mathematics*. Boca Raton, FL, USA: 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.
- Wriggers, W., Milligan, R. A., Schulten, K. & McCammon, J. A. (1998). *J. Mol. Biol.* **284**, 1247–1254.