Acta Cryst. (1995). D**51**, 740–748

The Direct Rotation Function: Rotational Patterson Correlation Search Applied to Molecular Replacement

BY WARREN L. DELANO AND AXEL T. BRÜNGER

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

(*Received* 23 *September* 1994; *accepted* 25 *January* 1995)

Abstract

Most rotation functions try to achieve maximal correlation between two Patterson functions by systematically rotating one and computing the overlap with the other. In contrast, the direct rotation function rotates a search model relative to the crystal unit cell and evaluates the linear correlation coefficient (Patterson correlation, PC) between squared normalized structure-factor amplitudes of the observed and calculated diffraction data. Structure factors are calculated from the rotated search model in a $P1$ unit cell identical to that of the target crystal. PC makes use of all self-Patterson vectors of the search model. A comparison of the direct rotation function, a real-space rotation function, and a fast rotation function suggests that the direct rotation function provides a considerable enhancement of the signal-to-noise ratio compared to other two. Combined with PC refinement, the direct rotation function was successful in solving multidomain macromolecular crystal structures.

Introduction

Patterson searches or molecular replacement have proved to be an effective means of solving macromolecular crystal structures using knowledge of similar structures (Hoppe, 1957; Rossmann & Blow, 1962). A model is obtained by sequence-similarity searches using the database of known crystal and solution nuclear magnetic resonance structures. The three-dimensional structure of the model must be close to that of the unknown crystal structure, or at least to parts of it. This search model is first rotated and then translated in the unit cell of the target crystal in order to obtain a maximum fit between observed and calculated diffraction data. If the correct orientation and position are found, the model can then serve as a starting point for model rebuilding and refinement.

Considerable effort has been devoted towards developing efficient algorithms for computing rotation functions (Rossmann & Blow, 1962; Huber, 1965; Crowther, 1972; Navaza, 1994). However, the increased computational efficiency comes at the cost of numerical approximations that may limit the success of a rotation search. A rotation function must produce an orientation as close as possible to the correct one in order for the subsequent translation function to be successful. As a result of the steady increase in available computational resources, efficiency is not necessarily a prime concern anymore. Therefore, more exact and computationally demanding approaches can be realistically used.

The criteria of fit of the rotation function is commonly approximated by a product between Patterson functions integrated over a specified volume. Furthermore, most rotation functions rotate Patterson functions against each other in order to achieve maximal correlation between them. Consequently, the integration volume needs to be restricted in order to avoid inclusion of artificial cross-vectors in the Patterson map of the search model.

Hauptman (1982) showed that the linear correlation coefficient between squared normalized structure factors, also referred to as 'Patterson correlation' (PC), is a measure of the phase accuracy of a partial atomic model. PC, therefore, has a natural application in molecular replacement where partial models are often used (Fujinaga & Read, 1987; Brünger, 1990). PC correlates observed and calculated structure factors where the latter are calculated for each sampled orientation of the search model in a $P1$ unit cell with dimensions identical to those of the target crystal (Fig. 1). PC is used for the direct rotation function (Brünger, 1992), refinement of the orientation and interdomain relationships (Brünger, 1990), and translation searches (Fujinaga & Read, 1987). PC makes use of all Patterson self-vectors of the search model and, thus, the integration volume does not have to be limited. The direct rotation function has been already used to solve two difficult molecular-replacement problems (Berchtold *et al.*, 1993; Gewirth & Sigler, 1995).

Here we present a comparison of the direct rotation function with two other rotation functions: the *AMoRE* fast rotation function (Navaza, 1994), and a real-space rotation function (Huber, 1965; Steigemann, 1974). The rotation functions are evaluated under a series of controlled conditions where the correct orientations are known. The test systems studied are an Fab fragment of a murine antibody complexed with digoxin and a mutant of the DNA-binding domain of the glucocorticoid receptor complexed with its cognate DNA. Several tests are performed using observed or simulated diffraction

ISSN 0907-4449

data, partial models, different resolution ranges and thermal B factors of the search model. From a peak analysis of the grid points produced by each rotation function, the signal-to-noise ratio is determined. The direct rotation function consistently provides a significantly better signal-to-noise ratio than the other two rotation functions. Finally, we propose a molecular-replacement strategy for solving multisubunit crystal structures with greater confidence.

Theory

Conventional rotation functions

All rotation functions involve the computation of the Patterson function or its reciprocal space analog for the search model (P_m) and the observed diffraction data (P_x). In conventional rotation functions, the search model is placed in a large orthogonal unit cell and the Patterson function P_m is rotated in order to achieve maximal overlap with P_x (Fig. 1). Therefore, lattice-translation vectors between neighboring unit cells need to be removed from the rotation function by restricting the integration volume (Rossman & Blow, 1962; Huber, 1965).

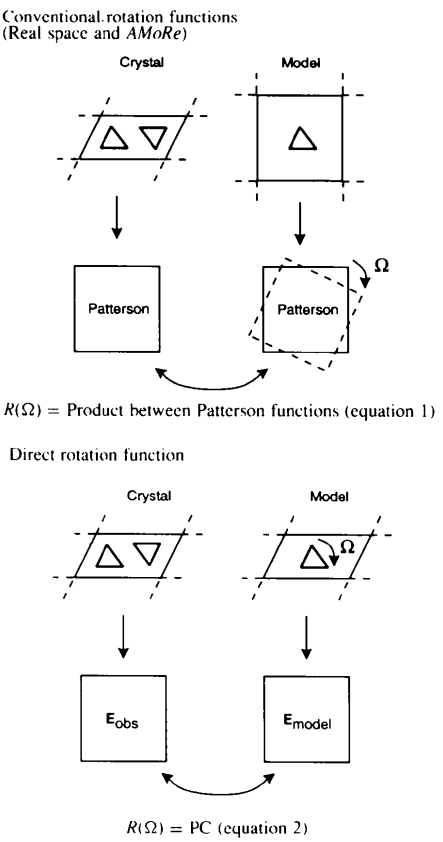

Fig. 1. Comparison of the direct rotation function and conventional rotation functions. Instead of rotating a Patterson map computed from the model, in the direct rotation function, the model itself is rotated, and structure factors are recomputed for each sampled orientation of the model. Instead of using a product of the Patterson functions, the direct rotation function computes the linear correlation coefficient between squared normalized structure factors PC (equation 2).

P_x and P_m are rotated with respect to each other and a product between the Patterson functions is computed,

$$\mathrm{Rot}(\Omega) = \int_U P_x(r)P_m(\Omega r)\mathrm{d}V\,, \qquad (1)$$

where the 3×3 rotation matrix Ω is described by three angles (*e.g.* Euler angles), r is the integration variable, and U is the volume of integration, usually spherical, centered at the origin.

Formulations of the rotation function in real and reciprocal space are in principle equivalent, although there are differences in the nuerical approximations employed and in the way certain weighting and cut-off schemes can be applied.

In the real-space formulation, only the strongest grid points of the model Patterson function P_m are used in the integration and the value of the crystal Patterson function P_x at these grid points are computed by interpolation (Huber, 1965; Steigemann, 1974). The integration is typically carried out over a volume that eliminates all Patterson vectors beyond a certain radius. The effect of the origin peak can be removed by subtracting its contribution from the Patterson function (Lipson & Cochran, 1966; Nordman, 1966; Müller, Oehlenschläger & Buehner, 1995).

Crowther's (1972) fast rotation function uses a spherical harmonic expansion of (1) and a fast Fourier transform in order to compute the rotation function in a highly efficient manner. Navaza's (1993, 1994) implementation of the fast rotation function in *AMoRe* differs from Crowther's original work in two respects. First, instead of using a Fourier–Bessel expansion, *AMoRE* achieves better accuracy with numerical intergration (Navaza, 1987, 1993). Second, *AMoRe* employs a more stable recurrence relation to compute the reduced rotation matrices (Navaza, 1990). This permits the inclusion of higher order spherical harmonics allowing more accurate expansions of the Patterson functions over a larger volume U.

The direct rotation function

The direct rotation function differs from conventional rotation functions by placing the model in a cell of $P1$ symmetry having cell dimensions and angles equal to that of the target crystal structure (Fig. 1). Thus, the Patterson function computed from the model can be directly compared against the observed one without having to restrict the integration volume. Instead of rotating the Patterson function, the model itself is rotated by a matrix operator Ω.

The target of the direct rotation function is defined as the linear correlation coefficient between observed

(E_{obs}) and calculated [$E_m(\Omega)$] normalized structure factors (Hauptman, 1982),

$$R_{direct}(\Omega) = PC = \langle |E_{obs}|^2 |E_{m(\Omega)}|^2 - \langle |E_{obs}|^2 \rangle \langle |E_{m(\Omega)}|^2 \rangle \rangle \div \{ [\langle |E_{obs}|^4 - \langle |E_{obs}|^2 \rangle^2 \rangle \times \langle |E_{m(\Omega)}|^4 - \langle |E_{m(\Omega)}|^2 \rangle^2 \rangle]^{1/2} \} , \quad (2)$$

where the symbols $\langle\rangle$ denote an average computed over the observed reflections expanded to P_1. The normalization of the structure factors is carried out numerically in equal-volume resolution shells.

Peak analysis

A cluster analysis was carried out for the grid points of both the direct and the real-space rotation functions (Brünger, 1992). The top 6000 grid points of the rotation functions were sorted by value. The highest grid point was marked. All other grid points were considered one at a time in decreasing order. A grid point was marked if no other grid point within 10° was already marked. The marked grid points are referred to as 'points'. It should be noted that a peak of the rotation function would show up as point but the converse is not necessarily true. The cluster analysis produces exactly one point for a peak that is narrower than 10°. Multiple points are produced for a peak that is broader than 10°. Although this may seem somewhat redundant it offers an advantage for PC refinement by reducing the impact of the multiple maxima problem. Increasing the number of starting orientations around a given peak increases the chances of convergence to the global minimum which may not exactly coincide with the peak, especially when PC refinement is carried out at a higher resolution than the rotation function.

The *AMoRe* rotation function performs a different peak analysis by identifying all grid points whose rotation-function values are higher than any of their nearest neighbors. We refer to these peaks also as points of the *AMoRe* rotation function. The slightly different definition of points for the different rotation functions did not significantly affect the comparison of the signal-to-noise ratios of the rotation functions (as defined in the *Results* section).

Materials and methods

Test case: (26–10) *Fab fragment*

The structure of the murine antidigoxin (26–10) Fab fragment complexed with digoxin was solved (Brünger, 1991) by using a conventional rotation function, PC refinement, and a translation function with the Fab portion of the HyHEL-5 Fab–lysozyme complex determined by Sheriff *et al.* (1987). The complex crystallized in space group $P2_1$, a = 44.144, b = 164.69, c = 70.17 Å, γ = 108.50°, with two Fab molecules in the asymmetric unit related by non-crystallographic symmetry. The diffraction data were twinned with partially overlapping reflections from the twins. A procedure was employed to reduce the residual amplitudes from the weaker of the two twins (Strong, 1990). The data set was 78% complete at 15–4 Å resolution after application of the detwinning procedure (data set 3r9-det using the notation of Brünger, 1991). The structure was eventually refined to an R value of 17.4% at 8–2.5 Å resolution (Jeffrey *et al.*, 1993).

In the numbering scheme of the HyHEL-5 Fab fragment the variable domain consists of the domains V_L = 1–106 and V_H = 1001–1116, whereas the constant domain consists of residues C_L = 107–212 and C_H1 = 1117–1218. The elbow angle is defined as the angle between the pseudo twofold axes of symmetry of the V_L–V_H and C_L–C_H1 domain pairs. The elbow-angle for the HyHEL-5 search model is 161.1° compared with 171.5° for the (26–10) Fab crystal structure. All atoms of the HyHEL-5 Fab crystal structure were included during the original molecular replacement.

For the rotation-function comparisons presented here, the same Fab search model was used. One series of tests used the original observed diffraction data of the (26–10) Fab–digoxin complex truncated to 4 Å resolution. A second series used simulated structure factors calculated from a partially refined (26–10) Fab–digoxin structure. The latter series was used to assess how the rotation functions would perform under ideal conditions in the absence of experimental error.

Test case: TR_{GR}–GRE complex

TR_{GR} is a mutant of the glucorticoid receptor (GR) which has been modified to adopt the binding specificity of the thyroid receptor (Gewirth & Sigler, 1995). This protein was crystallized in a complex with a glucocorticoid response element (GRE) DNA duplex in space group $P2_12_12_1$, a = 38.744, b = 73.608, c = 116.505 Å. The GRE DNA fragment consisted of the symmetric 18 base-pair duplex,

5′-TTCCAGAACATGTTCTGGA-3′,

with an overhanging 5′ thymine at each end. The duplex contained two adjacent hexameric GRE half sites (5′-TGTTCT-3′) with no spacer between them. One copy of the receptor TR_{GR} was bound to each DNA half site, giving rise to an asymmetric unit containing two proteins and one DNA molecule. The structure was solved by molecular replacement with the direct rotation function using one half of the rat glucocorticoid receptor–GRE DNA complex (Luisi *et al.*, 1991) as a search model. Thus, the search model consisted of a protein monomer and a GRE half site. Residues that differed between GR and TR_{GR} (58, 59, 62, 77–81) were truncated to alanine or glycine in the search model.

The initial diffraction was truncated to 4 Å resolution for solution by molecular replacement using the direct

rotation function. At this resolution the data set was 93% complete and R_{sym} was 5.4%. After collection of a higher resolution data set, the structure was eventually refined to a free R value of 28.8% and an R value of 19.1% at 6–2 Å resolution (Gewirth & Sigler, 1995).

For the rotation-function comparisons presented here, the GR–GRE half complex is again used as a search model with the initial TR_{GR}–GRE diffraction data set truncated to 4 Å resolution. A second set of tests used simulated data calculated from the refined TR_{GR}–GRE crystal structure.

Rotation-function parameters

The fast rotation function was carried out with the *AMoRe* program as implemented in the *CCP*4 package *Version* 2.4 (Collaborative Computational Project, Number 4, 1994). The direct and real-space rotation functions were computed by *X-PLOR, Version* 3.1, (Brünger, 1992) modified to perform the cluster analysis described above for the direct rotation function. Table 1 shows the definition of the signal-to-noise ratios for the different rotation functions.

Unless otherwise noted, all of the rotation searches used reflections from 15 to 4.0 Å resolution and uniform model B factors of 15 Å^2. The direct rotation and the real-space rotation functions used reflections with $F_{obs} > 2\sigma$. The outer radius r of the Patterson integration volume for the real-space and *AMoRE* rotation function was set to the maximum distance from the center of mass of the model to any atom (40 and 27 Å for the Fab and TR_{GR}–GRE test cases, respectively). The inner radius of integration for the real-space rotation function was set to 5.0 Å. The origin peak was subtracted from the Patterson function for the real-space rotation function (Lipson & Cochran, 1966; Nordman, 1966; Müller *et al.*, 1995). The inner radius for *AMoRE* was fixed at zero as required by the program. The dimensions of the orthogonal model unit cell for computing P_m in the *AMoRe* and real-space rotation functions were set to be larger than the extent of the model plus the integration radius r.

The *AMoRE* rotation function used Euler sampling of the rotation function in uniform 2.5° increments. The real-space rotation function used quasi-orthogonal Lattman (1972) sampling with a θ_2 interval of 2.5°. This is an appropriate sampling interval considering the integration radii and the resolution limit (4 Å) that were used in most test cases. The direct rotation function also used quasi-orthogonal Lattman sampling but with a θ_2 interval of 5° in order to reduce the computer time required. Although the direct rotation function samples rotation space on a coarser grid than the *AMoRe* or the real-space rotation function, solution peaks were, in general, broad enough to ensure detection even with the coarser sampling interval (shown in Fig. 2 for the TR_{GR}–GRE test case).

The computation of PC (2) was performed in 20 equal-volume resolution shells. Experience shows that the performance of the direct rotation function is rather insensitive to the bin size.

For both the real-space and direct-rotation functions, structure factors were calculated from electron density in a unit cell sampled with a grid spacing of one third of the high-resolution limit except at low resolution (8–15 and 10–15 Å) where a spacing of 2 Å was used. For the *AMoRe* rotation searches, the grid spacing was fixed at 0.75 Å as required by the program.

Space-group symmetry allowed us to restrict sampling of the rotation function to the asymmetric units shown in Table 2.

Table 1. *Definition of the signal-to-noise ratio for rotation functions*

The definition is illustrated for the TR_{GR}–GRE diffraction data using the GR protein without the DNA as the search model. Only the top points and the point corresponding to the correct orientation are shown. Signal points (*i.e.* within 15° around one of the correct orientations) are shown in bold. For each point, Euler angles are given along with the rotation function value, the rotation function value in σ units (standard deviations above the mean), and the error from the known orientation (°). The top two points of the direct rotation function correspond to the two solutions related by non-crystallographic symmetry. The real-space rotation function produced one of the solutions as point 17. The *AMoRe* rotation function produced one of the correct orientations as point 72. For the direct rotation function., the signal-to-noise ratio is $4.39\sigma/3.55\sigma = 1.24$ (point 1 *versus* point 3). For the real-space rotation function the signal-to-noise ratio is $3.63\sigma/4.4\sigma = 0.83$ (point 17 *versus* point 1). For the *AMoRe* rotation function the signal-to-noise ratio is $17.73\sigma/3.64\sigma = 0.48$ (point 72 *versus* point 1).

Direct rotation function for TR_{GR}–GRE

Point No.	θ_1	θ_2	θ_3	RF	σ	Error (°)
1	**123.68**	**45.00**	**137.52**	**0.068**	**4.39**	**9.557**
2	**121.58**	**55.00**	**42.83**	**0.059**	**3.82**	**3.642**
3	143.77	45.00	116.08	0.055	3.55	23.529
4	**115.65**	**55.00**	**149.40**	**0.054**	**3.46**	**5.094**
5	**135.66**	**55.00**	**34.41**	**0.052**	**3.31**	**10.883**
6	139.23	80.00	37.49	0.051	3.27	28.681
7	135.19	70.00	0.19	0.050	3.22	37.273
8	125.33	40.00	125.33	0.050	3.19	19.357
9	127.35	90.00	38.25	0.048	3.06	34.257
10	**134.31**	**50.00**	**26.31**	**0.048**	**3.05**	**12.190**

Real-space rotation function for TR_{GR}–GRE

Point No.	θ_1	θ_2	θ_3	RF	σ	Error (°)
1	107.00	25.00	335.00	0.820	4.40	28.979
2	127.84	30.00	107.84	0.806	4.35	36.885
3	106.12	25.00	346.12	0.791	4.28	31.724
4	127.55	30.00	127.55	0.771	4.20	25.483
5	111.59	25.00	39.59	0.763	4.16	32.761
6	122.83	22.50	45.68	0.717	3.98	34.170
7	135.74	35.00	92.88	0.716	3.97	43.423
8	105.25	25.00	357.25	0.704	3.92	37.230
17	**122.25**	**40.00**	**309.75**	**0.633**	**3.63**	**17.350**

AMoRe rotation function for TR_{GR}–GRE

Point No.	θ_1	θ_2	θ_3	RF	σ	Error (°)
1	35.06	63.02	85.26	10.900	3.64	73.898
2	92.82	30.83	13.76	10.500	3.50	41.226
3	255.61	80.39	48.69	9.600	3.20	65.053
4	343.27	36.00	170.00	9.600	3.20	85.555
5	230.83	39.45	152.65	6.500	3.17	76.573
6	33.12	67.98	139.49	9.400	3.13	87.847
7	253.40	78.77	30.00	9.300	3.10	63.824
8	309.20	78.52	55.80	9.100	3.03	52.105
72	**121.73**	**55.95**	**43.88**	**5.200**	**1.73**	**4.570**

Table 2. *Asymmetric units for rotation functions and Euler angle conventions*

The Euler angles $\theta_1, \theta_2, \theta_3$ are defined by counter-clockwise rotation of the model by θ_1 around the z axis, rotation by θ_2 around the new x axis, and rotation by θ_3 around the new z axis. The quasi-orthogonal Euler angles $(\theta_+, \theta_2, \theta_-)$ are defined through $\theta_+ = \theta_1 + \theta_3$ and $\theta_- = \theta_1 - \theta_3$ (Lattman, 1972). The Euler angles are related to *AMoRe*'s (α, β, γ) angles by the following equations: $\theta_1 = -[\gamma - (\pi/2)]$, $\theta_2 = -\beta$, $\theta_3 = -[\alpha + (\pi/2)]$.

Program	Space group $P2_1$, b axis unique	Space group $P2_12_12_1$
	$0 < \theta_- < 2\pi$	$0 < \theta_- < \pi$
X-PLOR	$0 < \theta_2 < (\pi/2)$	$0 < \theta_2 < (\pi/2)$
	$0 < \theta_+ < 4\pi$	$0 < \theta_+ < 4\pi$
	$0 < \alpha < 2\pi$	$0 < \alpha < \pi$
AMoRe	$0 < \beta < (\pi/2)$	$0 < \beta < (\pi/2)$
	$0 < \gamma < 2\pi$	$0 < \gamma < 2\pi$

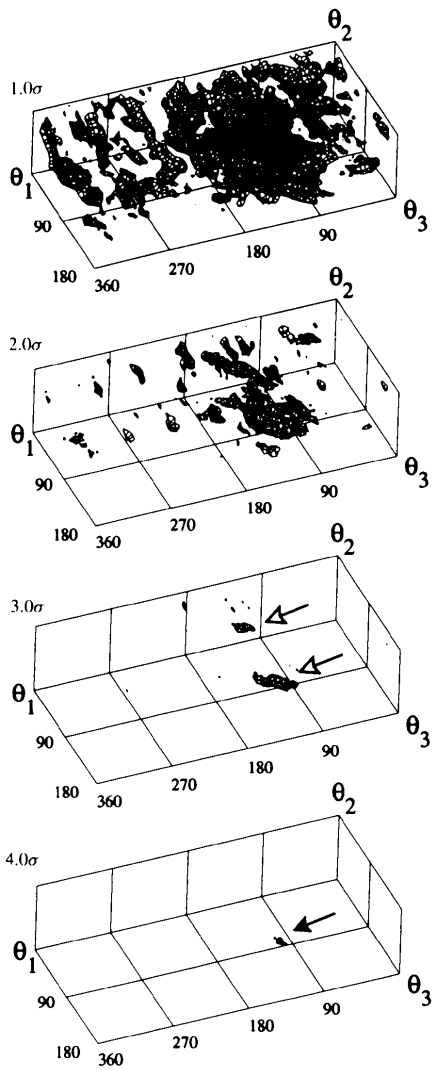

Fig. 2. Three-dimensional maps of the direct rotation function (in Euler angle space) at various contour levels. The GR protein was used as a search model with the $TR_{GR}GRE$ diffraction data. At a level of 3σ, the two correct orientations related by non-crystallographic symmetry are indicated by hollow arrows. At 4σ, only one of the two points shows up (indicated by the solid arrow).

Results

We defined the signal-to-noise ratio of rotation functions as the ratio of the value of the highest signal point to that of the highest noise point measured in σ units above the mean. The criterion by which we distinguished signal from noise was determined by the radius of convergence of PC refinement. Generally, a point that is within about 15° of one of the correct orientations will converge to it by PC refinement (Brünger, 1993) and is thus considered a signal point. A point that is more than 20° away from the correct orientation is outside the radius of convergence of PC refinement and is thus considered to be noise. Artificial specification of a cutoff of 20° did not significantly change the results and, thus, our conclusions are independent of the precise definition of signal and noise.

A signal-to-noise ratio greater than 1 indicates that the highest point is within 15° to the correct orientation. A signal-to-noise ratio less than one indicates that the highest point is noise. Since only the top 99 points obtained from each of the rotation functions were analyzed, a zero signal-to-noise ratio was assigned if the correct orientation was not among the top 99 points.

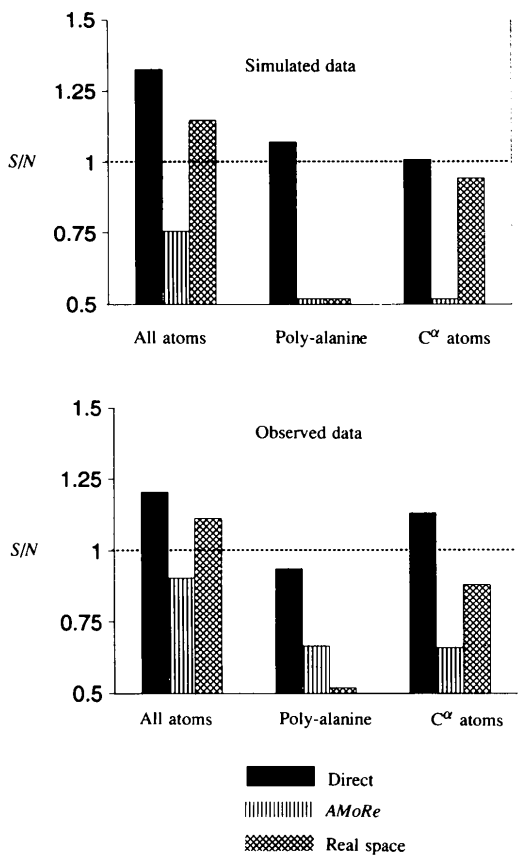

Fig. 3. Signal-to-noise ratio for the three rotation functions using the HyHEL-5 Fab structure with the (26–10) Fab–digoxin complex diffraction data.

Because of the presence of twofold non-crystallographic symmetry in both test cases, two correct orientations for the search model exist for both the Fab and TR_{GR}–GRE test cases. If one of the orientations was found by the particular rotation function, the other one was generally also present among the top points. Interestingly, the strength of both signals is not necessarily identical (Fig. 2).

Fig. 3 compares the performance of the direct rotation function, the *AMoRe* fast rotation function, and the real-space rotation function when applied to the (26–10) Fab test case using the HyHEL-5 Fab search model. With the all-atom model, the direct rotation function successfully identifies the two correct orientations related by non-crystallographic symmetry for both observed and simulated diffraction data as the two top points. The highest points of the other two rotation functions are incorrect, although the correct orientations are among the top 99 points. The strength of the signal falls off as a partial model is used consisting of backbone and C^{β} atoms. The signal-to-noise ratio slightly improves for observed data when reducing the polyalanine model to a C^{α} model. Apparently, the C^{α} model is more precise and, therefore, produces a stronger signal although the number of scattering atoms is significantly reduced.

Figs. 4 and 5 compare the performance of the rotation functions when using search models consisting of the constant or the variable domain of the Fab, respectively. The direct rotation function correctly identifies the solutions as the highest points while this is not the case for the conventional rotation functions. Despite having half as many X-ray scatterers in the search model, which amount to less than 25% of the asymmetric unit, the solutions produced by the direct rotation function for the individual domains always have a higher signal-to-noise ratio than those for the complete Fab molecule (Fig. 3). Thus, the individual domains are better search models than the complete Fab which is probably a consequence of the 10° difference in elbow angle between the search model and the crystal structure (Brünger, 1991). The direct rotation function correctly identifies the orientations of the individual domains as the hightest point when using C^{α} models consisting of as little as 5% of the X-ray scatterers in the asymmetric unit (Figs. 4 and 5).

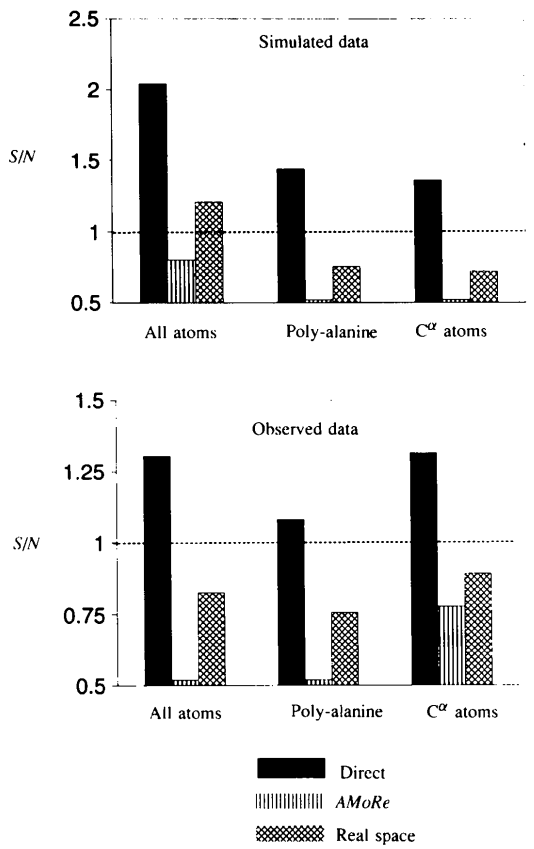

Fig. 4. Signal-to-noise ratio for the three rotation functions using the constant domain of the HyHEL-5 Fab structure with the (26–10) Fab-digoxin complex diffraction data.

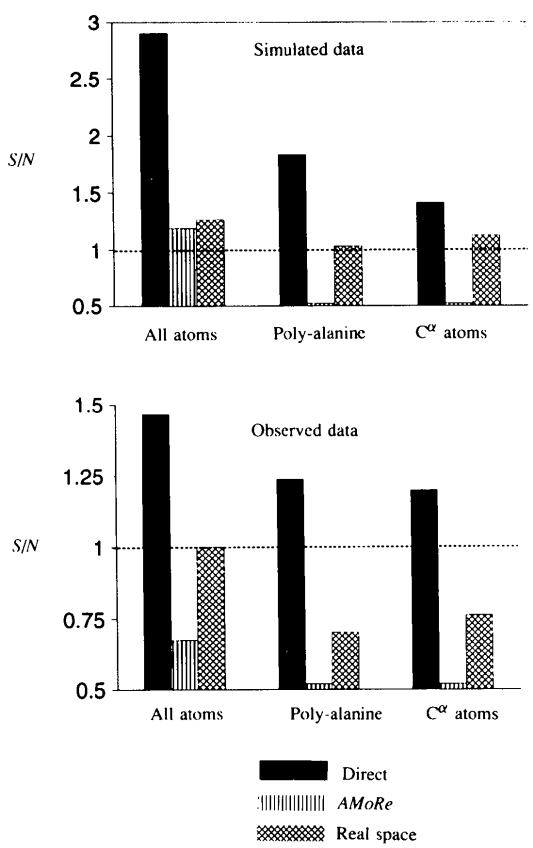

Fig. 5. Signal-to-noise ratio for the three rotation functions using the variable domain of the HyHEL-5 Fab structure with the (26–10) Fab–digoxin complex diffraction data.

Fig. 6 compares the rotation functions obtained for the TR_{GR}–GRE test case. The top point of the direct rotation function corresponds to the correct orientation when using all-atom models for the protein–DNA half complex or the protein by itself. The other two rotation functions do not produce the correct orientation as the highest point.

Fig. 7 shows the effect of resolution on the rotation functions when applied to the (26–10) Fab–digoxin complex using the variable domain as a search model. The signal disappears as the resolution is lowered and is improved as the resolution increased. Together with previous results (Brünger, 1993), we conclude that at lower resolution, false maxima of the rotation function beome dominant. In fact, using data in the ranges 8–15 and 10–15 Å, the top points of the direct rotation function are far (30–90°) from the correct orientations. Over the large resolution range tested, the direct rotation function always had a higher signal-to-noise ratio than the other two rotation functions.

Fig. 8 shows the effect of changing the uniform model *B* factors when applied to the (26–10) Fab–digoxin complex. Modification of the uniform *B* factors in the search model has very little effect on any of the rotation functions.

Discussion

The direct rotation function showed a higher signal-to-noise ratio compared to the *AMoRe* and real-space rotation functions. The underlying reason can be found in the inclusion of all self-Patterson vectors of the search model. In contrast, conventional rotation functions are typically limited to a subset of self vectors wihin the radius of integration.

The increased signal-to-noise ratio provided by the direct rotation function comes at some computational cost. Table 3 compares the time required for computing the three rotation functions. A direct rotation search can take two orders of magnitude longer than an *AMoRE* fast rotation search. Yet with a system such as the TR_{GR}–GRE complex which has a very weak solution signal, only the direct rotation function identified the correct orientation as the highest point (Fig. 6). Systems

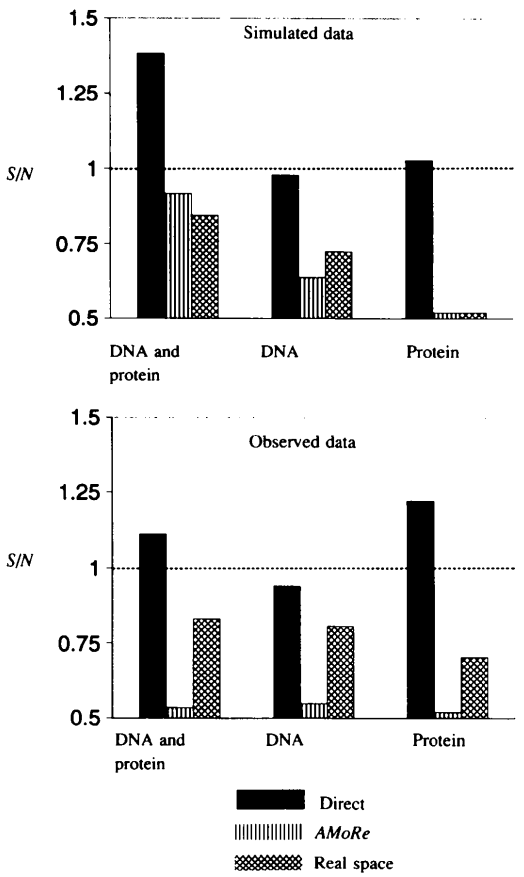

Fig. 6. Signal-to-noise ratio for the three rotation functions using the structure of the GR–GRE half complex as a search model with the TR_{GR}–GRE complex diffraction data.

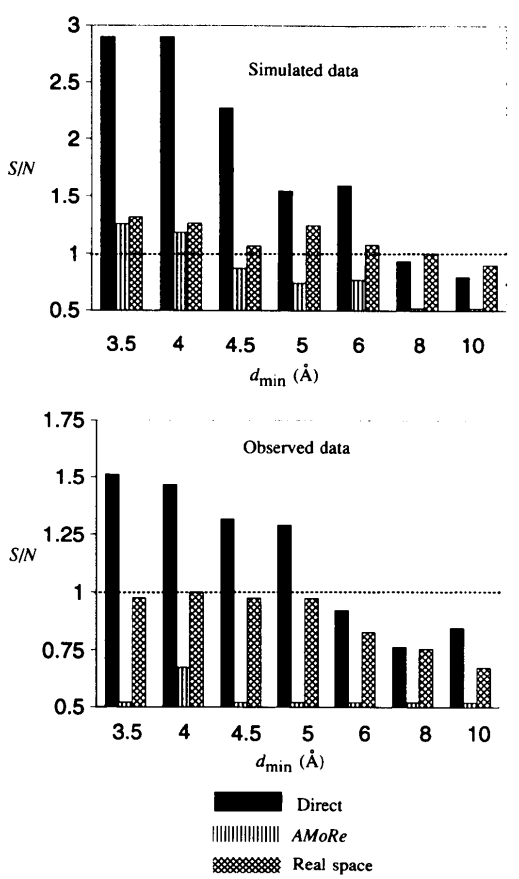

Fig. 7. Signal-to-noise ratio as a function of resolution for the three rotation functions using the atoms of the variable domain of the HyHEL-5 Fab structure with the (26–10) Fab–digoxin complex diffraction data.

that have stronger solution signals will be amenable to the use of the other rotation functions, but the same solution produced by the direct rotation function is likely to be more clearly distinguished from noise. In many cases, such added confidence may be worth the computational effort required by the direct rotation function. The increased sensitivity of the direct rotation function enables the use of search models that constitute a smaller portion of the asymmetric unit, such as C^{α} models (Figs. 4 and 5).

The broad peaks belonging to the correct orientations (Fig. 2) suggest that it may be possible to conduct direct rotation searches on a very coarse grid followed by optimization of selected peaks using simulated annealing or genetic algorithms (Mitchell Lewis, personal communication).

Concluding remarks

Under the conditions tested, the direct rotation function shows a considerable enhancement of the signal-to-noise ratio relative to the other rotation functions. However, the application of PC refinement (Brünger, 1990, 1991, 1993) to the highest points of a conventional rotation function may have also identified the correct orientation, an approach that is now widely used for solving crystal structures by molecular replacement. The question arises what advantage the direct rotation function offers compared to this approach.

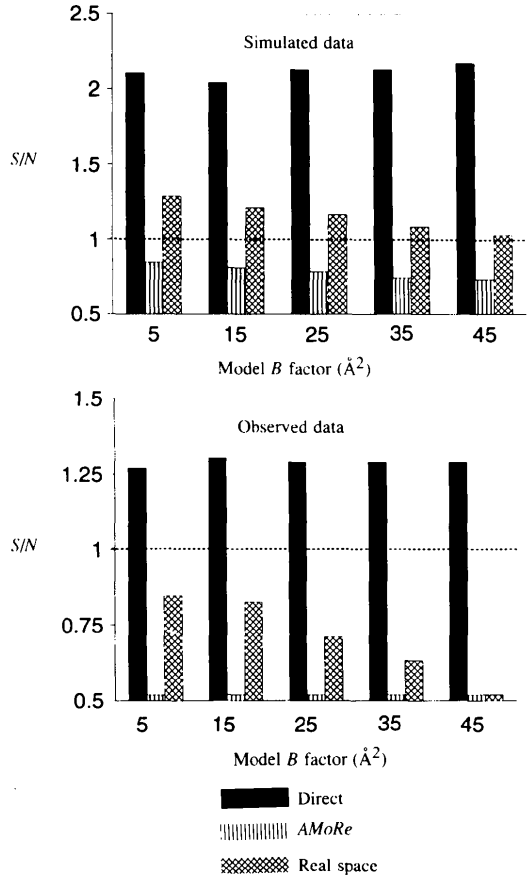

Fig. 8. Signal-to-noise ratio as a function of uniform model B factors for the three rotation functions using the atoms of the constant domain of the HyHEL-5 Fab structure with the (26–10) Fab–digoxin complex diffraction data.

Table 3. *Rotation-function computer times*

Central processing unit times (in hours on an Hewlett Packard HP735 workstation) of the rotation searches for the (26–10) Fab–digoxin diffraction data using the HyHEL-5 Fab as a search model and for the TR_{GR}–GRE diffraction data using the GR–GRE half complex as a search model.

Rotation function	(26–10) Fab–digoxin complex	TR_{GR}–GRE
AMoRe	0.1	0.02
Real space	2.0	0.91
Direct	31.2	14.0

One of the test cases presented here, the (26–10) Fab fragment, was originally solved by a combination of a conventional rotation function combined with PC refinement (Brünger, 1991). It was difficult to achieve convergence of PC refinement for the 24 parameters describing the positions and orientations of the four subunits of the Fab fragment (constant and variable domains of the heavy and light chains). By trial and error, two different conditions were found under which either one of the orientations related by non-crystallographic symmetry emerged. No single condition was found under which both orientations could be simultaneously identified. Thus, the success of PC refinement was to some degree determined by chance. Furthermore, one of the solutions was far down the list of highest points of the conventional rotation function. This example illustrates two situations where PC refinement fails: if the correct orientation is not in the selected list of highest points or if PC refinement does not converge. Although the first situation can in principle be avoided by a comprehensive PC-refinement search (Evans, Rose, To, Young & Bundle, 1994) this approach is in most cases impractical because of the extremely large amount of computer time required.

The direct rotation function can improve the performance of PC refinement by using the strategy outlined in Table 4. First, the direct rotation function has a higher chance finding the correct orientation as the highest point (Figs. 3–8). Second, it can be applied to the subunits of the search model, such as the constant and variable domains of an Fab fragment, and still produce a significant signal whereas conventional rotation functions may miss the signal altogether (Figs. 4 and 5). After finding the orientations of the individual subunits, the search model can be assembled taking into account reasonable geometric and packing criteria. PC refinement of the orientations and positions of the subunits will further improve the search model which can then be used for the subsequent translation search.

Table 4. *Combined direct rotation function and PC-refinement strategy*

- Direct rotation searches of model domains
- Selection of the highest points for further analysis
- Assembly of the domains for good packing and steric arrangements
- PC refinement of the interdomain angles, positions, and the overall orientation
- Translation search

This approach was used in the structure solution of the TR_{GR}–GRE mutant of the glucorticoid receptor DNA complex (Gewirth & Sigler, 1995). It should be noted that translation searches of the individual half complexes were unsuccessful, only the assembled full complex produced a significant translation-function solution. Thus, the combination of the direct rotation function using the half complex and PC refinement of the interdomain positions and angles of the full complex were needed to unambiguously solve the structure. The use of the direct rotation function significantly increased our confidence at the early stages of the structure-solution process because the highest point represented the same orientation regardless of whether the protein–DNA half complex or the protein alone was used (Fig. 6). From this and other examples one can expect the direct rotation to become a useful method to address difficult molecular-replacement cases.

We thank M. Büehner for useful discussions, D. Gewirth for providing unpublished data, T. Fischmann for help with the *AMoRe* program, and P. Adams, L. Rice, and S. Shah for careful reading of the manuscript. This work was funded in part by a grant from the National Science Foundation to ATB (DIR 9021975).

References

BERCHTOLD, H., RESHETNIKOVA, L., REISER, C. O. A., SCHIRMER, N. K., SPRINZL, M. & HILGENFELD, R. (1993). *Nature (London)*, **365**, 126–132.
BRÜNGER, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
BRÜNGER, A. T. (1991). *Acta Cryst.* A**47**, 195–204.
BRÜNGER, A. T. (1992). *X-PLOR, Version* 3.1, *A System for X-ray Crystallography and NMR*. New Haven, CT: Yale Univ. Press.
BRÜNGER, A. T. (1993). *Immunomethods*, Vol. 3, edited by I. A. WILSON, pp. 180–190. New York: Academic Press.
COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* D**50**, 760–763.
CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.
EVANS, S. V., ROSE, D. R., TO, R., YOUNG, N. M. & BUNDLE, D. R. (1994). *J. Mol. Biol.* **241**, 691–705.
FUJINAGA , M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
GEWIRTH, D. & SIGLER, P. B. (1995). *Nature Struct. Biol.* In the press.
HAUPTMAN, H. (1982). *Acta Cryst.* A**38**, 289–294.
HOPPE, E. (1957). *Acta Cryst.* **10**, 750–751.
HUBER, R. (1965). *Acta Cryst.* A**19**, 353–356.
JEFFREY, P. D., STRONG, R. K., SIEKER, L. C., CHANG, C. Y. Y., CAMPBELL, R. L., PETSKO, G. A., HABER, E., MARGOLIES, M. N. & SHERIFF, S. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 10310–10314.
LATTMAN, E. E. (1972). *Acta Cryst.* B**28**, 1065–1068.
LIPSON, A. & COCHRAN, W. (1966). *The Determination of Crystal Structures*, 3rd ed., pp 169–170. New York: Cornell Univ. Press.
LUISI, B. F., XU, W. X., OTWINOWSKI, Z., FREEDMAN, L. P., YAMAMOTO, K. R. & SIGLER, P. B. (1991). *Nature (London)*, **352**, 497–505.
MÜLLER, T., OEHLENSCHLÄGER, F. & BUEHNER, M. (1995). *J. Mol. Biol.* **247**, 360–372.
NAVAZA, J. (1987). *Acta Cryst.* A**43**, 645–653.
NAVAZA, J. (1990). *Acta Cryst.* A**46**, 619–620.
NAVAZA, J. (1993). *Acta Cryst.* D**49**, 588–591.
NAVAZA, J. (1994). *Acta Cryst.* A**50**, 157–163.
NORDMAN, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29–38.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.
SHERIFF, S., SILVERTON, E. W., PADLAN, E. A., COHEN, G. H., SMITH-GILL, S. J., FINZEL, B. & DAVIES, D. R. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 8075–8079.
STEIGEMANN, W. (1974). PhD thesis, Technische Univ., München, Germany.
STRONG, R. K. (1990). PhD thesis, Harvard Univ., USA.