

Phasing from an envelope

Quan Hao

Cornell High-Energy Synchrotron Source and
Department of Molecular Biology and Genetics,
Cornell University, Ithaca, NY 14853, USA

Correspondence e-mail: qh22@cornell.edu

Solution of the phase problem is central to crystallographic structure determination. Conventional molecular-replacement methods are ineffective in the absence of knowledge of the structure of a homologous protein. A recent method utilizing the low-resolution molecular shape determined from solution X-ray scattering data has been shown to be successful in locating the molecular shape within the crystallographic unit cell for the cases of the trimeric nitrite reductase (AxNiR, 105 kDa) and the dimeric superoxide dismutase (SOD, 32 kDa). This was achieved by performing a direct real-space search for orientation and translation using the orientation of the non-crystallographic axis obtained by performing a self-rotation on the crystallographic data. This effectively reduces the potential six-dimensional search to a four-dimensional one (Eulerian angle γ and three translational parameters). The program *FSEARCH* incorporating this method has been generalized to handle molecules from all space groups. The program can also be used in general six-dimensional cases for a molecular-replacement solution given a predetermined envelope from any source, such as electron-microscopic images or solution scattering, provided that the envelope can be converted to the standard *CCP4* map format or expressed in terms of spherical harmonics. It is hoped that this method will greatly facilitate the *ab initio* structure determination of proteins and provide a good foundation for further structure refinement.

Received 7 February 2001

Accepted 6 June 2001

1. Introduction

Solution X-ray scattering data obtained using synchrotron-radiation X-rays have proven to be very useful in providing low-resolution structural details of proteins and other macromolecules in solution. Owing to the significant progress that has been made with the development of *ab initio* phasing methods for low-resolution shape restoration in terms of spherical harmonics (Svergun & Stuhrmann, 1991; Svergun *et al.*, 1996), the spatial parameters of a structure's molecular envelope can be determined in a model-independent manner which does not, for example, require the use of crystal structure coordinates for interpretation. This method has been used to analyse scattering data from a nitrogenase protein complex to provide a stable and unique shape restoration at ~ 15 Å resolution (Grossmann *et al.*, 1997). The low-resolution structure of nitrite reductase (105 kDa) from *Alcaligenes xylooxidans* (AxNiR) had been determined from scattering data (Grossmann & Hasnain, 1997). Although the crystal structure was previously solved by the molecular-replacement method at 2.8 Å, the molecular shape of nitrite reductase (AxNiR) determined from solution scattering was successfully

used as a candidate for *ab initio* phasing by locating the molecule within the crystallographic unit cell (Hao *et al.*, 1999). In order to test the generality of this method, the fairly small dimeric molecule superoxide dismutase (SOD) from bovine erythrocytes (32 kDa) was similarly treated (Ockwell *et al.*, 2000).

2. X-ray scattering experiments and molecular-shape determination

In the case of a monodisperse system consisting of randomly orientated particles it is possible to obtain structural information of the particles from small-angle scattering data. Small-angle scattering is performed on a dilute suspension of biological macromolecules whose shape is to be determined. Then, from the one-dimensional scattering profile, the three-dimensional shape of the particles can be characterized in terms of a series of multipole coefficients. Knowledge of these multipole coefficients allows one to generate an approximate model of the shape of the macromolecules being examined (Stuhrmann, 1970; Svergun *et al.*, 1996).

If we assume that the scattering is caused by a globular homogeneous molecule, one can define its molecular envelope by a two-dimensional angular function $F(\theta, \varphi)$ describing the molecular boundary such that the particle density $\rho(r)$ is unity inside and vanishes elsewhere. The function $F(\theta, \varphi)$ can conveniently be expanded into a series of spherical harmonics $Y_{lm}(\theta, \varphi)$ according to (Stuhrmann, 1970),

$$F(\theta, \varphi) = R_0 \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\theta, \varphi),$$

with f_{lm} being complex multipole coefficients and L representing the multipole order. R_0 is a scale factor [$\simeq (3V/4\pi)^{1/3}$], where V is the volume of the particle. Furthermore,

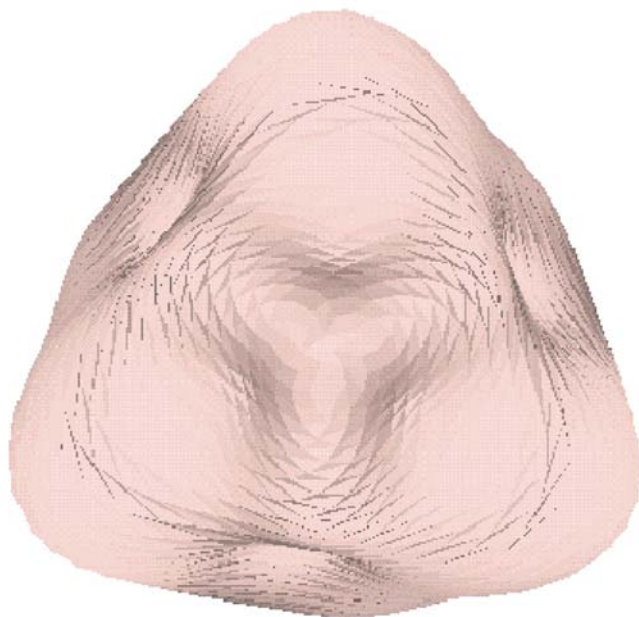


Figure 1

Nitrite reductase from *A. denitrificans*; molecular envelope obtained directly from solution X-ray scattering data at $\sim 15 \text{ \AA}$ resolution.

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

where $P_l^m(\cos \theta)$ are associated Legendre functions (with argument $\cos \theta$) and l and m are integers with $-l \leq m \leq l$. Consequently, the ratio of quadrupolar term and zero-order term $5^{1/2}|f_{20}|/f_{00}$ is a good indication of the deviation of the molecular shape from sphericity. A computational procedure to evaluate the multipole coefficients from the experimental scattering curve by minimizing a residual R was developed by Svergun & Stuhrmann (1991). Details of the algorithm are presented for example in Svergun *et al.* (1996).

The range of experimentally available scattering data generally allows the determination of 15–25 variables in the shape description. This imposes an upper limit for the multipole resolution L , since the number of independent parameters in the above series is equal to $(L+1)^2 - 6$ (arbitrary rotations and translations of the molecule do not alter the scattering curve and therefore lead to a reduction of six variables). Consequently, in general unique shape calculations with the multipole order of $L = 4$ are possible. In addition, molecular symmetry imposes restrictions on the multipole coefficients f_{lm} which can improve the reliability of the shape restoration by reducing the number of parameters to be calculated. The higher the symmetry, the more multipole coefficients can be omitted, which results in an enhanced resolution (*i.e.* multipole expansions with $L = 6$ or 7 are achievable). AxNiR contains three chemically identical subunits and is known to be a trimer in solution; assuming the trimer has threefold symmetry (which is shown by the self-rotation function) there are additional constraints on the multipole coefficients. The multipole expansion up to $L = 7$ for this symmetry group requires only 22 free parameters, of which 19 parameters are found to have values larger than 0.01 (*i.e.* all coefficients other than $m = 3n$, where n is an integer, should vanish provided that the Cartesian coordinate system for the trimeric molecule is chosen so that the threefold axis coincides with the z axis). The restored envelope of AxNiR is displayed in Fig. 1. The shape of the protein neatly reproduces the molecular features when compared with the overall details from the crystal structure (Grossmann *et al.*, 1997). In the case of the dimeric SOD, the multipole expansion up to order $L = 6$ requires 25 parameters (all coefficients other than $m = 2n$, where n is an integer, should vanish) to be determined from the scattering profile (Ockwell *et al.*, 2000).

3. FSEARCH program: locating the molecular shape in the crystallographic unit cell

The conventional method for correctly positioning a known search molecule in a crystallographic unit cell – an important first step in solving macromolecular structures by the molecular-replacement method – is by the use of the cross-rotation function (Rossmann & Blow, 1962). However, attempts to locate the molecular shape determined from solution scattering by performing a Patterson search at different resolutions using *AMoRe* (Navaza, 1994) were not

successful as the density inside the envelope is uniform, hence there is no discrimination between intra-envelope Patterson vectors.

In the case of a macromolecule lacking non-crystallographic symmetry, a full six-dimensional search (three rotational and three translational) to find the best match between F_{obs} and F_c appears necessary. For a molecule in possession of non-crystallographic symmetry, the search may be performed in two separate stages. Utilizing a self-rotation function using *ALMN* in the *CCP4* suite (Collaborative Computational Project, Number 4, 1994) with crystallographic data can yield two Eulerian angles, α and β , for the non-crystallographic (NCS) axis of the molecular shape. Once the orientation of this NCS axis is known, the potential six-dimensional search is reduced to four (Eulerian angle γ and three translational

Table 1
Molecular-replacement solutions for AxNiR and SOD.

The Eulerian angle γ and three translational parameters, X , Y and Z were searched by using the *FSEARCH* program and the results are given in ascending order of the R factor. The correct solutions are indicated by bold values. The nearest noise peaks are also listed.

	γ	X	Y	Z	R
AxNiR	95	20	42	22	0.601
	115	20	42	22	0.615
SOD	60	18	24	54	0.471
	45	6	24	57	0.503

parameters), resulting in a significant reduction in calculation time when locating the molecular shape within the crystallographic unit cell. It is worth noting that for any given orientation of a non-crystallographic axis, it is possible that the macromolecule can take either of two orientations with respect to the axis; hence, a search based on the original and the 'flipped' molecule is necessary.

The program *FSEARCH* can be used to conduct a simultaneous (one- to six-dimensional) search on orientation and translation to find the best match between F_{obs} and F_c . It has been generalized to handle molecules from all space groups and, in particular, those in possession of non-crystallographic symmetry. A flow chart of the program *FSEARCH* is shown in Fig. 2.

The *FSEARCH* program has been tested with the solution scattering and crystallographic data from the proteins AxNiR and SOD. In both cases, the correct orientation and translation of the molecular mask which had been determined from solution scattering profile was clearly identified by the program. Full details of the solutions have been published elsewhere (Hao *et al.*, 1999; Ockwell *et al.*, 2000) and a summary of the results is shown in Table 1 and Figs. 3 and 4.

4. Concluding remarks

It has been demonstrated that the molecular shape determined from solution scattering can be located in the crystallographic unit cell using the program *FSEARCH*. The program has been generalized to handle molecules from all space groups and adapted to be compatible with the standard *CCP4* libraries and file formats. *FSEARCH* can also be used in general six-dimensional cases for a molecular-replacement solution given a predetermined envelope from any source, such as electron-microscopic images or solution scattering, provided that the envelope can be converted to the standard *CCP4* map format or expressed in terms of spherical harmonics. The knowledge of the orientation of a non-crystallographic symmetry axis (conveniently determined by a self-rotation function) can reduce the potential six-dimensional search to four (Eulerian angle γ and three translational parameters). The actual CPU time consumed by a four-dimensional search on an SGI Origin 200 server (one 175 MHz R10000 processor) was about 1 h. However, if no such axis exists in the molecule, a six-dimensional search would be necessary and a time frame in the order of 1000 CPU hours would be expected.

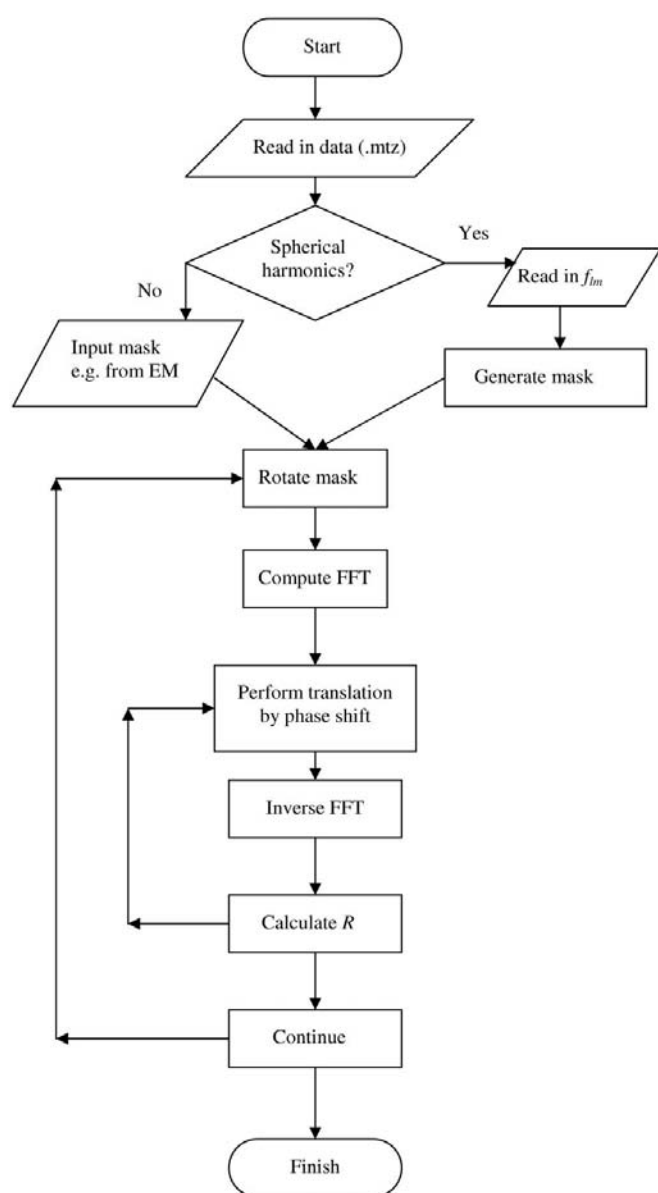


Figure 2
Flow chart of the program *FSEARCH*. The input envelope must be in the standard *CCP4* map format or expressed in terms of spherical harmonics.

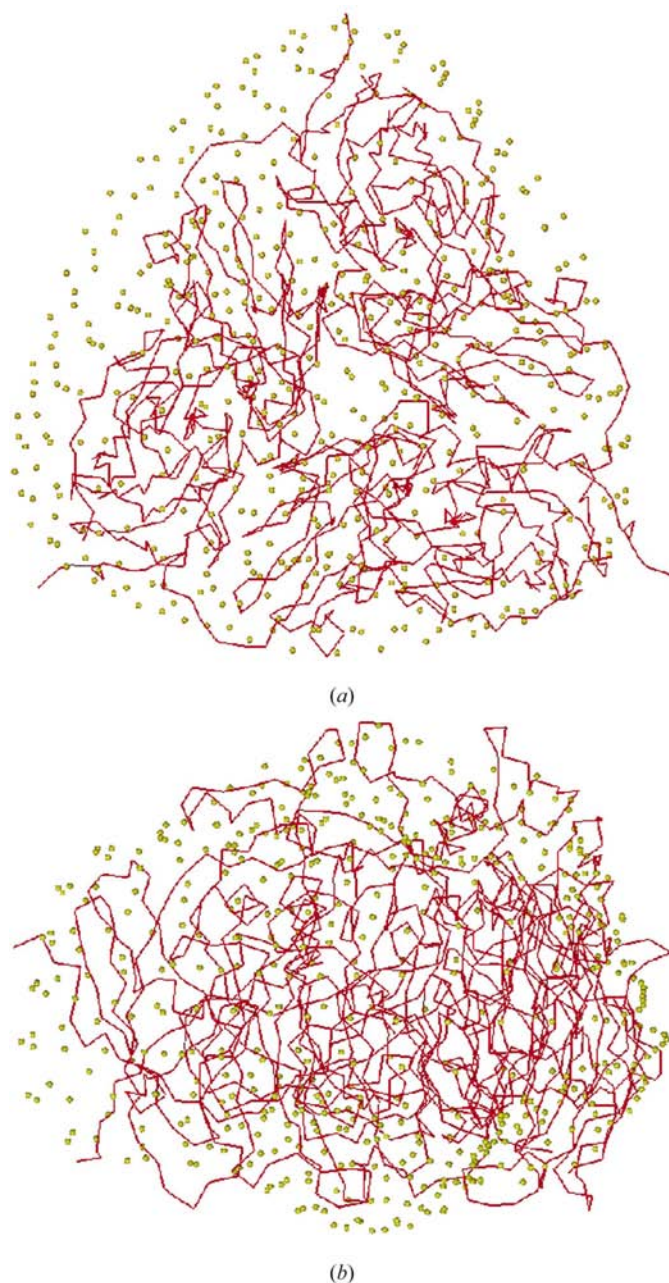


Figure 3

Molecular shape of AxNiR (represented by yellow dots) found from solution scattering and located in the unit cell by using the *FSEARCH* program. The crystal structure (Dodd *et al.*, 1997) is superimposed and represented by red chains (a) looking down the threefold NCS axis and (b) 90° to the first view.

It is anticipated that the low-resolution phases calculated from the correctly positioned molecular shape can be used as a good starting point for phase extension through the use of genetic algorithms, whereby the mask would be used as the arena for ascertaining a macromolecule's internal structure. Some preliminary tests are promising and full results will be reported in due course. Once the resolution of the structure has been improved to ~ 5 Å using this method, phase extension to higher resolutions may be achieved by maximum

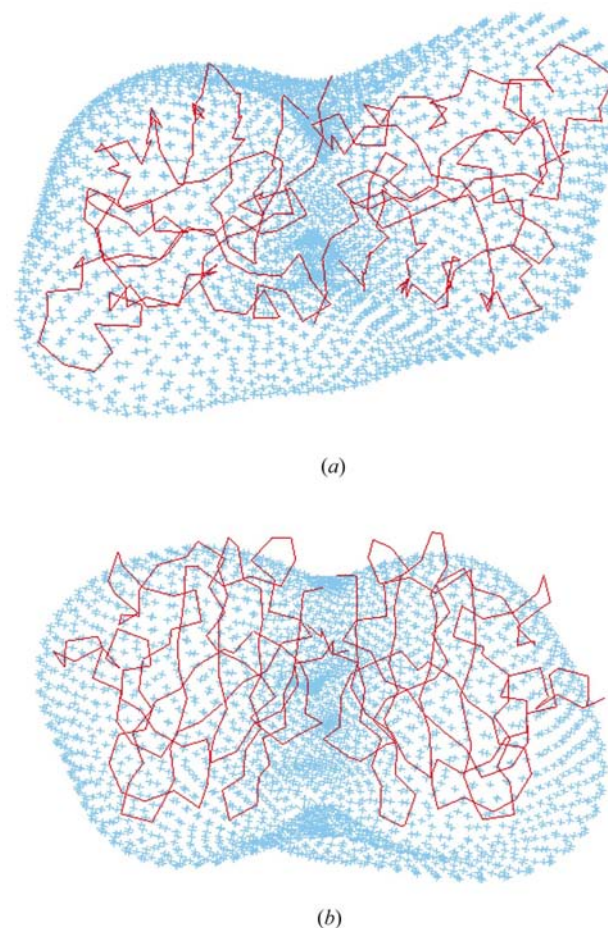


Figure 4

Molecular shape of SOD (represented by blue crosses) found from solution scattering and located in the unit cell by using the *FSEARCH* program. The crystal structure (Hough & Hasnain, 1999) is superimposed and represented by red chains (a) looking down the twofold NCS axis and (b) 90° to the first view.

entropy and density-modification methods (*e.g.* solvent flattening, histogram matching, NCS averaging). Thus, it is hoped that this method will greatly facilitate the *ab initio* structure determination of proteins and provide a good foundation for further structure refinement.

The *FSEARCH* program can be obtained by contacting the author of this paper.

I am grateful to Drs J. Grossmann, F. Dodd, M. Hough and Professor S. Hasnain for providing the test data and useful discussions. I would like to thank D. Ockwell for improving and testing the *FSEARCH* program.

References

- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
 Dodd, F. E., Hasnain, S. S., Abraham, Z. H. L., Eady, R. R. & Smith, B. E. (1997). *Acta Cryst.* **D53**, 406–418.

- Grossmann, J. G. & Hasnain, S. S. (1997). *J. Appl. Cryst.* **30**, 770–775.
- Grossmann, J. G., Hasnain, S. S., Yousafzai, F. K., Smith, B. E. & Eady, R. R. (1997). *J. Mol. Biol.* **266**, 642–648.
- Hao, Q., Dodd, F. E., Grossmann, J. G. & Hasnain, S. S. (1999). *Acta Cryst. D* **55**, 243–246.
- Hough, M. & Hasnain, S. S. (1999). *J. Mol. Biol.* **287**, 579–592.
- Navaza, J. (1994). *Acta Cryst. A* **50**, 157–163.
- Ockwell, D. M., Hough, M., Grossmann, J. G., Hasnain, S. S. & Hao, Q. (2000). *Acta Cryst. D* **56**, 1002–1006.
- Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
- Stuhrmann, H. B. (1970). *Acta Cryst. A* **26**, 297–306.
- Svergun, D. I. & Stuhrmann, H. B. (1991). *Acta Cryst. A* **47**, 736–744.
- Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). *Acta Cryst. A* **52**, 419–426.