Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ab initio phasing using molecular envelope from solution X-ray scattering

Q. Hao,[a]* F. E. Dodd,[b] J. G. Grossmann[b] and S. S. Hasnain[a,b]

[a]Department of Chemistry and Physics, De Montfort University, Leicester LE1 9BH, England, and [b]CCLRC Daresbury Laboratory, Warrington WA4 4AD, England

Correspondence e-mail: qhao@dmu.ac.uk

Received 29 April 1998
Accepted 27 August 1998

Solving the phase problem is the crucial and quite often the most difficult and time-consuming step in crystallographic structure determination. The traditional methods of isomorphous replacement (MIR or SIR) and molecular replacement require the availability of an isomorphous heavy-atom derivative or the structure of a homologous protein, respectively. Here, a method is presented which utilizes the low-resolution molecular shape determined from solution X-ray scattering data for the molecular search. The molecular shape of a protein is an important structural property and can be determined directly by the small-angle scattering technique. The idea of locating this molecular shape in the crystallographic unit cell has been tested with experimental diffraction data from nitrite reductase (NiR). The conventional Patterson search proved to be unsuccessful, as the intra-envelope vectors are uniformly distributed and do not match those of intra-molecular (atom-to-atom) vectors. A direct real-space search for orientation and translation was then performed. A self-rotation function using 2.8 Å crystallographic data yielded the polar angles of the non-crystallographic threefold axis. Knowledge of the orientation of this axis reduces the potential six-dimensional search to four (Eulerian angle γ and three translational parameters). The direct four-dimensional search within the unit cell produced a clear solution. The electron-density map based on this solution agrees well with the known structure, and the phase error calculated from the map was 61° within 20 Å resolution. It is anticipated that the low-resolution envelope can be used as a starting model for phase extension by the maximum-entropy and density-modification method.

1. Introduction

Solution X-ray scattering is a very effective technique for obtaining low-resolution structural details of proteins and their complexes in solution. However, the application of X-ray scattering to the determination of the molecular structure of biological molecules has remained limited owing to the model-dependent way in which data has to be interpreted. Recently, there has been significant progress with the development of an *ab initio* method for low-resolution shape restoration in terms of spherical harmonics (Svergun & Stuhrmann, 1991; Svergun *et al.*, 1996). This method is model-independent and does not, for example, require the use of crystal structure coordinates for interpretation. Grossmann *et al.* (1997) have used this method to analyze scattering data from a nitrogenase protein complex. The data were obtained using synchrotron radiation X-rays and provided a stable and unique shape restoration at ~15 Å resolution.

Recently, the low-resolution molecular shape of nitrite reductase from *Alcaligenes xylosoxidans* (AxNiR) has been determined from scattering data (Grossmann & Hasnain, 1997) and the crystal structure determined at 2.8 Å (Dodd *et al.*, 1997). The structure was solved by the molecular-replacement method, but here we use it to test the location of the molecular shape in the crystallographic unit cell, followed by *ab initio* phasing; it is a good example in view of the threefold symmetry.

2. Molecular-shape determination using scattering data

Solution-scattering data for AxNiR were collected on Station 8.2 at the Daresbury Synchrotron Radiation Source. The molecular shape of AxNiR was determined directly from the scattering profile alone in a model-independent manner. Full details have been published elsewhere (Grossmann & Hasnain, 1997) and only brief results are presented here.

If we assume that the scattering is caused by a globular homogeneous molecule, its molecular envelope can be defined 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),

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

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

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

where $P_l^m(\cos\theta)$ are the associated Legendre functions (with argument $\cos\theta$) and l and m are integers with $-l \leq m \leq l$. Consequently, the ratio of the quadrupolar term and the zero-order term, $(5^{1/2}|f_{20}|)/f_{00}$, is a good indication of the deviation of the molecular shape from spherical. 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 15–20 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 practice unique shape calculations with the multipole order $L = 4$ 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 rotation function and also the original crystal structure determination) 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 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 co-ordinate system for the trimeric molecule is chosen so that the threefold axis coincides with the z axis). The multipole coefficients with maximum order $L = 7$ are listed in Table 1. The R_0 value is 34.23 and the ratio $(5^{1/2}|f_{20}|)/f_{00}$ is 0.235. The error is estimated to be less than ±10% (essentially arising from the uncertainty in the calculation of the volume from the experimental scattering profile). The restored molecular envelope (Grossmann & Hasnain, 1997) provides excellent agreement with the crystal structure (Dodd *et al.*, 1997). The molecular shape viewed at two orthogonal orientations is shown in Fig. 1. The deviation from a spherical shape is evident and the trimeric nature of the molecule is clearly visible.

Table 1
Multipole coefficients (real and imaginary components of f_{lm} values) evaluated for the shape restoration of NiR (maximum order $L = 7$).

l	m	Real	Imaginary
0	0	3.449	—
1	0	0.013	—
2	0	−0.362	—
3	0	0.020	—
3	3	0.152	−0.156
4	0	0.158	—
4	3	−0.023	0.052
5	0	−0.097	—
5	3	−0.005	0.020
6	0	0.085	—
6	3	−0.089	0.082
6	6	−0.007	−0.008
7	0	0.104	—
7	3	−0.038	0.093
7	6	−0.011	0.045

3. Locating the molecular shape in the crystallographic unit cell

One crystalline form of AxNiR has space group $P2_12_12_1$, with unit-cell parameters $a = 67.89$, $b = 102.20$, $c = 151.88$ Å. There is a single AxNiR trimer in the asymmetric unit. X-ray diffraction data extending to 2.8 Å were collected on Station 9.5 at Daresbury Synchrotron Radiation Source (Dodd *et al.*, 1997). The low-resolution limit was 53 Å. The data have a completeness of 91% to 2.8 Å and 84% to 12 Å. The crystal structure was originally solved by the conventional molecular-replacement method with *AMoRe* (Navaza, 1994) using the

NiR of *Alcaligenes faecalis* (PDB code 1AFN) as the search model (Dodd *et al.*, 1997).

In general, the cross-rotation function (Rossmann & Blow, 1962) is the main tool for orienting a known search molecule in a crystal unit cell, a key step in solving macromolecular structures by the molecular-replacement method. The first attempt to locate the molecular shape determined by solution scattering was the conventional Patterson search at different resolutions using *AMoRe* (Navaza, 1994). This was unsuccessful, as the intra-envelope vectors are uniformly distributed and do not match the intra-molecular (atom-to-atom) vectors represented by the Patterson function. A six-dimensional search (three orientational and three translational) in real space appeared to be the only option. However, a self-rotation function using the program *ALMN* in the *CCP*4 suite (Collaborative Computing Project, Number 4, 1994) with 2.8 Å crystallographic data yielded polar angles for the non-crystallographic threefold axis (at $\omega = 107.7$, $\varphi = 0.0^\circ$) of the molecular shape. The knowledge of the orientation of this axis has reduced the potential six-dimensional search to four (Eulerian angle γ and three translational parameters). A computer program, *FSEARCH*, was written for simultaneous rotational (around the centre of the search model) and translational search to find the best match between $F_{\rm obs}$ and F_c. The four-dimensional search within the unit cell using crystallographic data (∞–12 Å) has produced a clear solution: $\alpha = 0$, $\beta = 107$, $\gamma = 95^\circ$, $x = 20$, $y = 42$ and $z = 22$ Å. Details are shown in Table 2 and Fig. 2, which illustrates the quality of fit between the molecular shape found here and the molecular structure as found by conventional crystal structure determination. *FSEARCH* also calculates the amount of overlap between symmetry-related copies of the search model which should, of course, be zero or very small in the correct solution; in our solution, there is no overlap. The solvent content of the crystal is 47%.

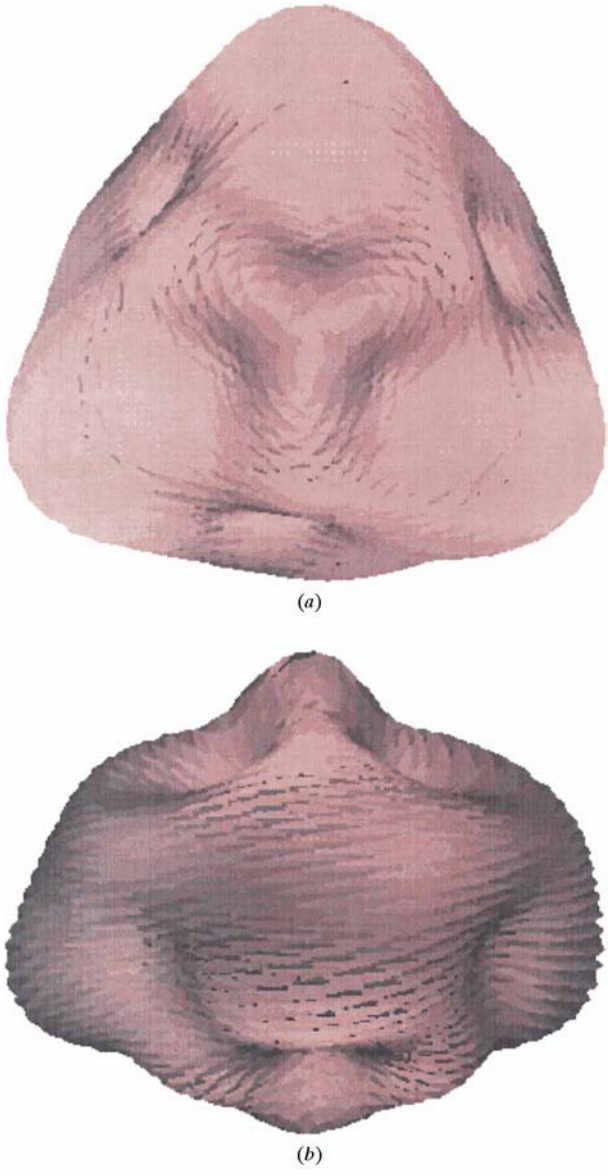

Figure 1
Molecular shape of AxNiR deduced from solution scattering for the multipole order $L = 7$. The mean radius is 34 Å. (*a*) View looking down the threefold-symmetry axis. (*b*) Model rotated by 90° around the horizontal axis.

Table 2
The positions with lowest R values found in the real-space search..

Molecular shape from solution scattering and crystallographic data within 12 Å were used and data are given in ascending order of R, $R = \sum |F_c - F_o| / \sum F_o$. The first position was chosen to be the solution.

(*a*) Initial four-dimensional (γ, x, y and z) search within the whole unit cell with 5° steps in γ and 2 Å steps in x, y and z

α	β	γ	x	y	z	R
0	108	95	20	42	22	0.601
0	108	95	20	44	22	0.615
0	108	115	20	42	22	0.615
0	108	75	24	36	30	0.616
0	108	110	20	42	24	0.618

(*b*) Finer search around the solution found from the initial search with 1° steps in α, β and γ and 1 Å steps in x, y and z

α	β	γ	x	y	z	R
0	107	95	20	42	22	0.587
0	106	95	20	42	22	0.590
0	106	96	20	42	22	0.590
0	105	95	20	42	22	0.591

Phases were then calculated from this solution and compared with phases of the known structure (Table 3). As expected, the low-resolution reflections have lower phase errors. The 67 reflections within 20 Å resolution have an average phase error of 61° and can be used as good starting phases for further phase extension to higher resolutions.

4. Discussion

We have demonstrated that the molecular shape determined from solution scattering can be located in the crystallographic unit cell. The knowledge of the orientation of a non-crystallographic symmetry axis (conveniently determined by a self-

Table 3
Average phase errors of the molecular-shape replacement result against the refined crystallographic model at 28 Å (Dodd *et al.*, 1997) in descending order of resolution.

Resolution (Å)	Data completeness (%)	Number of reflections	Mean phase error (°)
40	36	4	28
30	56	16	32
25	60	29	41
20	72	67	61
15	79	163	78
12	84	325	87

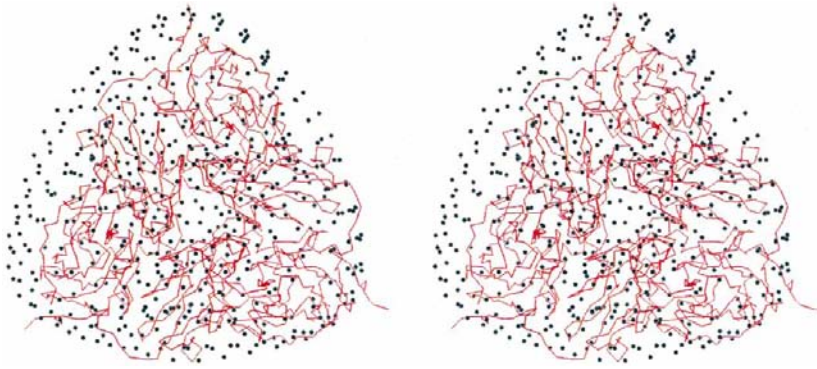

Figure 2
Stereo pair showing the molecular shape found from solution scattering and located in the unit cell by a real-space search, superimposed on the 2.8 Å crystal structure model (Dodd *et al.*, 1997). The molecular shape is represented by black dots uniformly distributed within its outline and the crystal structure model by red chains.

rotation function) has reduced the potential six-dimensional search to four dimensions (Eulerian angle γ and three translational parameters). However, if no such axis exists in the molecule, a six-dimensional search would be necessary which is very time-consuming. The low-resolution phases calculated from the correctly positioned molecular shape can be used for phase extension to higher resolution by the maximum-entropy and density-modification method (*e.g.* solvent flattening, histogram matching, NCS averaging). In addition, scattering instruments on more powerful synchrotron radiation sources would allow collection of weaker higher angle solution-scattering data, enabling the resolution of the molecular mask to improve to 10–15 Å. The homogenous nature (for purposes of simplicity) of the mask in the current study could be modified to incorporate the internal mass distribution, which would provide obvious improvement. Thus, the method has much promise for *ab initio* structure determination of large proteins and is likely to become an important tool in the macromolecular crystallographer's tool-bag.

We would like to thank CCLRC and BBSRC for their funding and support of this project. This work was supported by the BBSRC/EPSRC Biomolecular Sciences programme (Grant 719/B06916 to SSH). We would like to also thank our colleagues including Drs Zelda Abrahams and Bob Eady, and Professor Barry Smith for their interest and support.

References

Collaborative Computing Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dodd, F. E., Hasnain, S. S., Abraham, Z. H. L., Eady, R. R. & Smith, B. E. (1997). *Acta Cryst.* D**53**, 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.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
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.