Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 13 May 2001
Accepted 19 July 2001

A simple inverse method for calculating electron-density maps

John R. Somoza,[a]* Abraham Szöke[b] and Hanna Szöke[b]

[a]Axys Pharmaceuticals, Inc., 385 Oyster Point Boulevard, Suite 1, S. San Francisco, CA 94080, USA, and [b]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. Correspondence e-mail: john@somoza.net

Electron-density maps are generally prepared by Fourier transforming a set of complex structure factors. However, a map can also be obtained through a real-space reconstruction method. Starting from an empty unit cell, the map can be iteratively modified until it agrees with the given structure factors. In this paper, a simple method is described for preparing electron-density maps using this technique and two examples of its application are given.

1. Introduction

In the past, we have described the development of a computational approach termed the 'holographic method' (Szöke, 1993; Maalouf *et al.*, 1993; Somoza *et al.*, 1995; Szöke *et al.*, 1997*a*,*b*). The holographic method provides a way of iteratively changing an electron-density map until it simultaneously agrees with both the diffraction data and any available real-space information.

There are several advantages to creating a map in this way. The biggest advantage is that this method provides an effective framework for incorporating real-space information (positivity, knowledge of the solvent region *etc*.) into the calculation of an electron-density map. The real-space information is incorporated by directly restraining the electron density.

Another advantage of the holographic method is that the resulting maps are less likely to include Fourier artifacts, such as those that result from an incomplete or truncated data set. Electron-density maps are usually obtained by Fourier transforming the structure factors. The holographic method, in contrast, uses a minimization scheme to modify the map until it agrees with the diffraction data. At each step, structure factors are calculated from the map and compared to the observed structure factors. A calculated structure factor for which there is no corresponding observed structure factor is simply ignored. Thus, missing diffraction data have a much smaller impact on the resulting electron density when the maps are generated by this method than they do in conventional difference Fourier maps.

This paper describes a technique that is a small subset of the holographic method, and which will be referred to as real-space reconstruction (RSR). The RSR technique shares some of the advantages of the holographic method, including the reduced susceptibility to Fourier artifacts. Starting from an empty unit cell, we describe a way of iteratively modifying electron density within the cell until a map is identified that is consistent with the given structure-factor data.

2. Implementation of the real-space reconstruction method

The goal of this method is to find the best (and everywhere non-negative) electron-density map compatible with a given set of complex structure factors, $\mathbf{F}_o(\mathbf{h})$. To implement this procedure, several things are needed. First, we need a way of representing arbitrary electron-density maps. Second, we need a way of evaluating how well the electron density agrees with $\mathbf{F}_o(\mathbf{h})$. Finally, we need a method of changing the map to best fit the given structure factors.

2.1. Description of the electron-density map as a sum of Gaussians

The first step in describing the density is to subdivide the unit cell into a regular grid. At each grid point, $\mathbf{r}_p$, a Gaussian is placed that is described by the expression

$$\rho(\mathbf{r}_p) = n(p)(\pi\eta\Delta r^2)^{-3/2}\exp(-|\mathbf{r}-\mathbf{r}_p|^2/\eta\Delta r^2), \qquad (1)$$

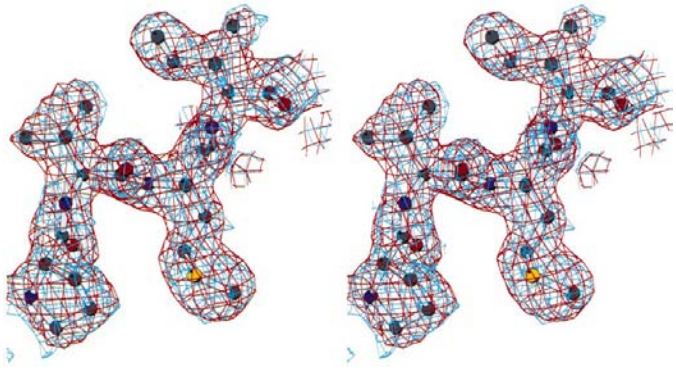

Figure 1
Stereo figure showing residues 40–43 of HGXPRTase along with the corresponding $2F_o - F_c$ (blue) and RSR (red) electron-density maps. These maps were prepared after omitting residues 1–50 as well as all of the solute molecules.

where Δr is the mean grid spacing and η (typically 0.8) specifies the width of the Gaussian with respect to the grid spacing. The Gaussians are spherical, of fixed width, and contain a variable number of electrons, $n(p)$. The electron density for the entire unit cell is described by the expression

$$\rho(\mathbf{r}) = (\pi\eta\Delta r^2)^{-3/2} \sum_{p=1}^{P} n(p) \exp(-|\mathbf{r} - \mathbf{r}_p|^2/\eta\Delta r^2), \qquad (2)$$

where P is the total number of grid points in the unit cell. It should be noted that the description of the electron density depends on the maximum resolution of the data being modeled. As the maximum resolution increases, the spacing between adjacent grid points (Δr) decreases, as does the width of each Gaussian.

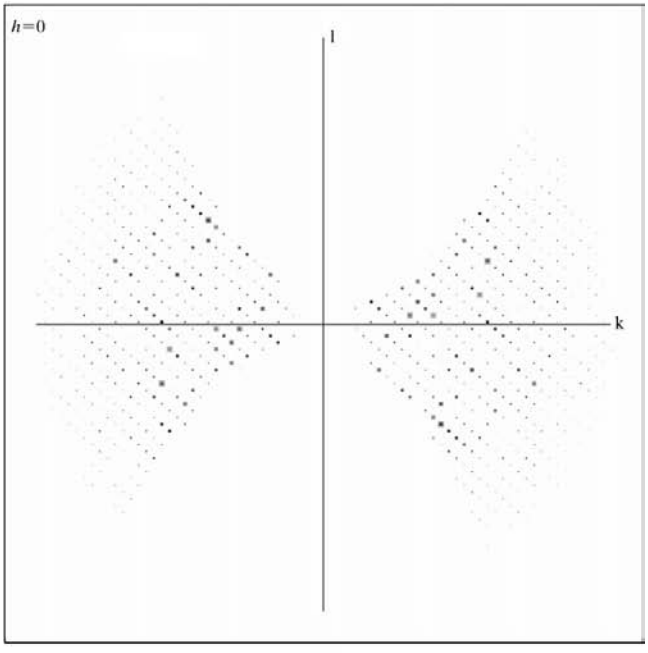

Figure 2
Display of the reflections in the (*a*) 0*kl* and (*b*) *h*0*l* planes of the cathepsin S data. Note the large wedge of data missing along the *l* axis.

2.2. The cost function and the minimizer

In order to identify an electron-density map that is consistent with a given set of structure factors, we transform the map to obtain a set of calculated structure factors, $F_c(\mathbf{h})$, and compare those structure factors to the $F_o(\mathbf{h})$ using the following cost function:

$$f_{\rm RSR} = \sum_h |\mathbf{F}_o(\mathbf{h}) - \mathbf{F}_c(\mathbf{h})|^2. \qquad (3)$$

The cost function defined in (3) is minimized using a conjugate-gradient algorithm designed by Goodman *et al.* (1993). The map is altered by changing the number of electrons, $n(p)$, in the Gaussians centered at each point r_p until the minimization algorithm converges.

3. Examples

Two examples of using the RSR method to prepare electron-density maps will be given. In each case, the starting point was a set of experimentally determined structure-factor amplitudes and a coordinate file that partially models the observed data and that is used to provide phases. For each example, two types of map were prepared: a sigmaa-weighted $2F_o - F_c$ map and a RSR map. In each case, the *SFALL* program (Collaborative Computational Project, Number 4, 1994) was used to calculate structure factors from the partial model and to place the $\mathbf{F}_o$'s on an (approximately) absolute scale. This was followed by sigmaa weighting using the *SIGMAA* program (Read, 1986). A $2F_o - F_c$ Fourier map was calculated using the program *FFT* (Ten Eyck, 1977). The RSR map was prepared using the program *EDEN* (Maalouf *et al.*, 1993; Somoza *et al.*, 1995; Szöke *et al.*, 1997*a*,*b*).

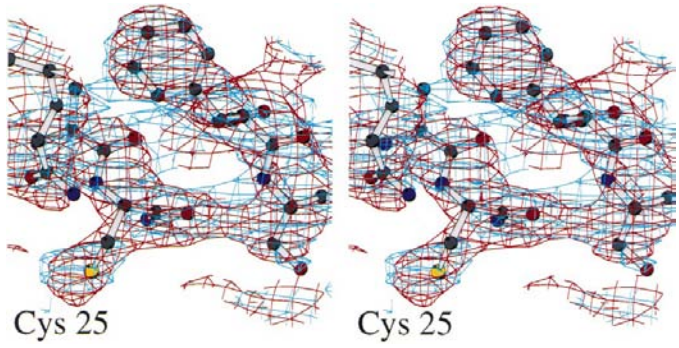

Figure 3
Stereo figure showing the vicinity of the active site cysteine of cathepsin S (Cys 25), along with the corresponding $2F_o - F_c$ (blue) and RSR (red) electron-density maps. The $2F_o - F_c$ map shows a break in density for the sidechain of Cys25, along with several instances of false connectivity. This region of the RSR map has no breaks in density and fewer instances of false connectivity. The *Z* axis roughly coincides with the horizontal axis of the figure.

Table 1
Data statistics.

	HGXPRTase	Cathepsin S
Space group	$P2_1$	$R3$
Cell parameters [a, b, c (Å)]	48.8, 74.7, 55.3, $\beta = 110.8^\circ$	107.61, 107.61, 105.19
Resolution range (Å)	100–1.9	100.0–2.5
Number of observations	97259	23957
Number of unique reflections	27314	10991
Completeness (%)		
All data	91.2	71.0
Highest resolution shell	55.6 (1.96–1.9Å)	65.0 (2.6–2.5Å)
$R_{sym}(I)$* (%)	6.3	9.7

HGXPRTase. The first example used diffraction data to 1.9 Å resolution that were collected on HGXPRTase (Somoza *et al.*, 1996; Table 1), a purine salvage enzyme. The data are reasonably complete (Table 1) and represent a typical situation in macromolecular crystallography. In this situation, one would not expect significant problems with truncation artifacts and would expect the difference Fourier and the RSR maps to be fairly similar. An examination of the difference Fourier and the RSR maps (Fig. 1) shows that, indeed, the two methods lead to similar maps, and that they are each clear and easy to interpret.

Cathepsin S. The second example used data from cathepsin S, a cysteine protease that plays an important role in the processing of MHC class II (McGrath *et al.*, 1998). Only one data-quality crystal of cathepsin S was obtained and the resulting 2.5 Å data set was significantly and systematically incomplete, with a large cone of data missing around the l axis (Figs. 2*a*,*b*) (McGrath *et al.*, 1998; Table 1). The corresponding difference Fourier maps are very streaky along the Z direction and have areas of missing and/or spurious electron density (Fig. 3). The elongated density along Z is an inevitable consequence of the missing data. However, the missing density as well as some of the spurious density are probably the consequence of Fourier artifacts.

Fig. 3 shows an example of both the Fourier and RSR maps in the area of the cathepsin S active site. There are clear problems with both maps, including both missing and spurious densities. As mentioned above, most of these errors are intrinsically linked to the missing cone of data. Although a comparison of the maps is necessarily subjective, the RSR map seems better in that there are fewer spurious connections, leading to a map that is easier to trace. Interestingly, the errors in the two maps often occur in different places. This suggests that the crystallographer fitting the density would benefit from looking at both maps.

4. Discussion

As shown above, it is possible to calculate a map by searching for electron density that agrees with the diffraction data. Under certain circumstances, this inverse method has advantages over a direct transform of the data; the biggest advantage lies in its ability to handle problems with the data that would ordinarily give rise to Fourier artifacts. When Fourier transforming structure factors, any missing data will give rise to spurious electron density. These artifacts are especially noticeable when blocks of data are missing, which happens when an incomplete data set is collected or when low-resolution data are not included in the Fourier transform calculation.

The root of these artifacts is that, when carrying out a Fourier transform, there is no difference between a missing structure factor and a structure factor with an amplitude of 0. This is not the case when the maps are computed using the inverse method. In the inverse case, the maps are prepared by minimizing a cost function obtained by comparing $\mathbf{F}_c$'s from the model with a given set of $\mathbf{F}_o$'s. If a subset of the $\mathbf{F}_o$'s is missing, neither it nor the corresponding $\mathbf{F}_c$'s is included in the calculation. Thus, the maps are not biased by the missing data.

A second, probably less significant, benefit of the inverse method as we have implemented it is that non-negativity of the electron density is enforced throughout the map. The incorporation of this constraint acts as a density-modification procedure, effectively improving the phases.

A possible caveat to the RSR method is that it relies on a minimization procedure that could suffer from convergence problems or that could converge on a false minimum. However, monitoring the cost function provides a way of determining whether the minimization procedure has led to a solution that is far from the global minimum. We have never seen any evidence of convergence problems or of convergence on local minima.

In summary, we have presented an inverse method of preparing electron-density maps. Under certain conditions, there is reason to believe that this method is less susceptible to artifacts than the more conventional difference Fourier method. Preparing an RSR map can be performed in an almost automated way with the use of a script that calls both *EDEN* and the CCP4 programs. *EDEN* can be obtained from the authors.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Goodman, D. M., Johansson, E. M. & Lawrence, T. W. (1993). *Multivariate Analysis: Future Directions*, edited by C. R. Rao, ch. 11. Amsterdam: Elsevier.

Maalouf, G. J., Hoch, J. C., Stern, A. S., Szöke, H. & Szöke, A. (1993). *Acta Cryst.* A**49**, 866–871.

McGrath, M. E., Palmer, J. T., Brömme, D. & Somoza, J. R. (1998). *Protein Sci.* 7, 1294–1302.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Somoza, J. R., Chin, M., Focia, P. J., Fletterick, R. J. & Wang, C. C. (1996). *Biochemistry*, **35**, 7032–7040.

Somoza, J. R., Goodman, D. M., Beran, P., Szöke, H., Truckses, D. M., Kim, S.-H. & Szöke, A. (1995). *Acta Cryst.* A**51**, 691–708.

Szöke, A. (1993). *Acta Cryst.* A**49**, 853–866.

Szöke, A., Szöke, H. & Somoza, J. R. (1997*a*). *Acta Cryst.* A**53**, 291–313.

Szöke, A., Szöke, H. & Somoza, J. R. (1997*b*). Proceedings of the CCP4 Study Weekend, January 1997.

Ten Eyck, L. F. (1977). *Acta Cryst.* A**33**, 486–492.