Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Bulk-solvent correction in large macromolecular structures

Bernard Rees,* Lasse Jenner and Marat Yusupov

Laboratoire de Biologie et Génomique Structurales, IGBMC, 1 Rue Laurent Fries, BP 10142, 67400 Illkirch CEDEX, France

Correspondence e-mail: rees@igbmc.u-strasbg.fr

Received 8 April 2005
Accepted 21 June 2005

The estimation of the bulk-solvent contribution to the diffraction of a macromolecular crystal makes use of a solvent mask which delimits the bulk-solvent regions in the crystal. It is shown that the way this mask is usually defined in *CNS* contains a bias which can lead to absurd results in the case of very large structures, where the calculations can only be made on relatively coarse grids. A modified procedure is described and applied to 70S ribosome data at 5.5 Å resolution. The *B* factor affecting the bulk solvent is also discussed. Even in this case of very high and widely variable atomic *B* factors, it seems sufficient to consider a constant and isotropic *B* factor for the bulk solvent. This is initially set to the average value of the atomic *B* factor, but can be refined.

1. Method

It is well known that a sensible modelization of the bulk solvent is very important for the refinement of a macromolecular structure when low-resolution terms are included. The structure factor is usually written

$$\mathbf{F} = k\{\mathbf{F}_{\rm calc} \exp[-\Delta B(\sin\theta/\lambda)^2] + d_{\rm solv}\mathbf{F}_{\rm solv} \exp[-B_{\rm solv}(\sin\theta/\lambda)^2]\}, \quad (1)$$

where k is the scale factor, $\mathbf{F}_{\rm calc}$ the structure factor calculated from the current atomic model, $d_{\rm solv}$ the bulk-solvent electron density (0.34 e Å^{-3} for pure water) and $\mathbf{F}_{\rm solv}$ the scattering of the solvent mask. This mask is a step function with value 1 in the solvent regions of the unit cell and 0 in the molecular regions. It is sometimes more convenient to consider a molecular mask, which is the complement of the solvent mask (value 1 in the molecular regions, 0 outside). By Babinet's theorem, except for H = 0, $\mathbf{F}_{\rm solv}$ is the opposite of the diffraction amplitude of the molecular mask.

At very low resolution, the macromolecule diffracts essentially as its molecular mask affected by the average macromolecular density and since this is usually not very different from $d_{\rm solv}$, $d_{\rm solv}\mathbf{F}_{\rm solv}$ and $\mathbf{F}_{\rm calc}$ are of the same order of magnitude, with a phase difference of π. The solvent contribution decreases faster than $\mathbf{F}_{\rm calc}$ with increasing resolution, but cannot be neglected as long as the resolution is not at least 4–5 Å. Clearly, the correction for bulk solvent becomes particularly important when the resolution limit of the experimental data is low, as is usually the case for very large macromolecules or macromolecular complexes. However, when we tried to apply the standard *CNS* bulk-solvent correction (Brünger *et al.*, 1998), with a reasonable $d_{\rm solv}$, to the *Thermus thermophilus* 70S ribosome, with a resolution limit of 5.5 Å (Jenner *et al.*, 2005), we noticed a large increase in the *R* factor, especially in the lowest resolution bins where it should have improved most. Attempts to optimize the parameters of (1) resulted in an unrealistically small value of $d_{\rm solv}$ (see below).

An essential part of the solvent modelling is the definition of the solvent mask. In *CNS*, this depends on two parameters: $r_{\rm probe}$ and $r_{\rm shrink}$. The first step of the procedure consists in centring spheres of radius $r_{\rm vdW} + r_{\rm probe}$ on all atomic postions of the macromolecular model ($r_{\rm vdW}$ is the van der Waals radius of the atom). This delimits the region that the centre of spherical solvent molecules of radius $r_{\rm probe}$ can occupy. All grid points within the spheres are labelled 0 and the grid points outside are labelled 1 (Fig. 1). The second step defines the regions effectively occupied by the solvent by removing a shell of thickness $r_{\rm shrink}$ from the molecular surface defined in step (i). In the standard *CNS* procedure, this is performed by relabelling as 1 all the points with initial label 0 closer than $r_{\rm shrink}$ from any point with initial

label 1 (Jiang & Brünger, 1994). Ideally, r_{probe} and r_{shrink} should be the average radius of the solvent molecules, but in practice empirical values are used. Jiang & Brünger (1994) proposed r_{probe} = 1.0 Å and r_{shrink} = 1.1 Å.

While this procedure has the advantage of simplicity, it contains an inherent bias which becomes serious whenever the grid step size is not much smaller than r_{shrink}. The reason of this is that the points labelled 1 considered in the second step, even those closest to the molecular surface, are not in general on the surface but already in the solvent region. Obviously, if the grid step is larger than r_{shrink}, no shrinking at all is performed. For a resolution limit of 6 Å, this happens even when a grid step of 1/5 of the resolution limit is chosen, while the usual practice is to choose 1/3 or 1/4.

We modified the second step of the procedure in the following way. Instead of the grid points, all intersection points of the sphere of radius $r_{vdW,\,i} + r_{probe}$ of every atom i of the model with the grid lines are considered (point P in Fig. 1). The condition for any such a point to be on the molecular surface and not inside the molecule is that its distance to the other atomic centres j is never less than $r_{vdW,\,j} + r_{probe}$. All grid points inside the sphere of radius r_{shrink} centred on such a point P are labelled 1.

r_{probe} and r_{shrink} were optimized in a way similar as in Jiang & Brünger (1994) and essentially the same values were obtained: r_{probe} = 1.0 Å and r_{shrink} = 1.2 Å

The complete procedure is as follows.

(i) Calculate a solvent mask with $r_{shrink} = 0$. This can be performed using *CNS*.

(ii) Apply the shrinking procedure described above (program *SHRINKMASK*).

(iii) Calculate and extract $\mathbf{F}_{calc}$ and $\mathbf{F}_{solv}$ with the mask determined in step (ii) (*CNS*).

(iv) Determine by least-squares the best values of k, ΔB, d_{solv} and B_{solv} of (1) (program *SOLVPAR*). If $|\Delta B|$ is large, ΔB is added algebraically to the B factors of all atoms of the model. d_{solv} and B_{solv} will be used in conjunction with the mask in further *CNS* refinements.[1]

2. Thermal motion

As mentioned above, $\mathbf{F}$ in (1) can be considered as the diffraction of the macromolecules minus the diffraction of the molecular masks. (1) implicitly assumes a rigid-body and isotropic thermal motion. The diffraction amplitude at a given time t is the sum of the amplitudes from the macromolecules of all the unit cells of the crystal, each one in its instantaneous position at t, displaced from the equilibrium position owing to thermal motion or local disorder. The contribution of the molecular mask at the same position is subtracted from the amplitude scattered by each macromolecule. As a result, B_{solv} in (1) should normally be equal to the B factor of the macromolecule. By the same argument, if different parts of the macromolecule have different B factors, the corresponding parts of the molecular mask should be affected by the same B factors. $\mathbf{F}_{solv}$ in (1) should therefore be replaced by

$$\mathbf{F}_{solv,B}(\mathbf{H}) = \sum_j \exp[-B_{av,j}(\sin\theta/\lambda)^2] \exp(2\pi i \mathbf{H}\cdot\mathbf{r}_j)\Delta V, \qquad (2)$$

where the sum is over all grid points j within the molecular mask. $\mathbf{r}_j$ is the position of a grid point and ΔV is the associated volume element.

[1] It is, in principle, possible to determine d_{solv} and B_{solv} using *CNS*. However, we have been unable to use this option successfully with our ribosome data.

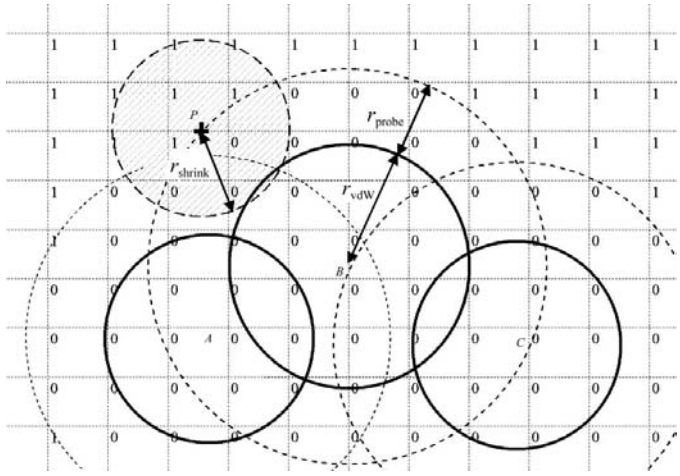

Figure 1
Definition of the solvent mask. The figure shows the result of the first step of the mask definition, where all grid points within the spheres of radius $r_{vdW} + r_{probe}$ centred on the atoms A, B, C, ... of the current model are labelled 0, while the points outside are labelled 1. In the second step, in the standard *CNS* procedure, any point 0 closer than r_{shrink} to a point 1 has its label changed to 1. In the procedure proposed here, all points P at the intersection of the surface defined in the first step with a grid line (in direction x, y or z) are considered. All grid points within a sphere of radius r_{shrink} centred on a point P are relabelled 1. r_{vdW} is the van der Waals radius of the atom type and r_{probe} and r_{shrink} are related to the radius of the solvent molecules, but are considered as empirical parameters (r_{probe} = 1 Å, r_{shrink} = 1.2 Å in this work).

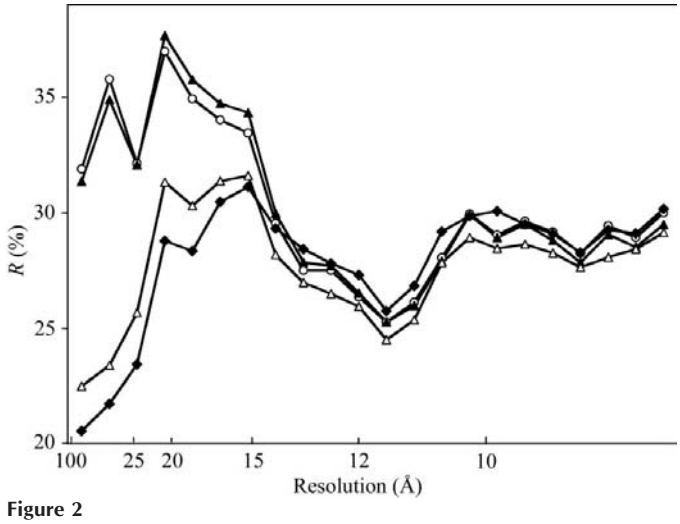

Figure 2
R factor at low resolution for the *T. thermophilus* 70S ribosome complexed with tRNAs and *thrS* mRNA (Jenner *et al.*, 2005). All reflections with $I > 2\sigma(I)$ in the 300–5.5 Å resolution range were used in the structure-factor calculations. Except in case (*c*), the grid-step size was 1/5 of the high-resolution limit. (*a*) No solvent correction (open circles). (*b*) Standard *CNS* correction, with solvent density refined to 0.053 e Å^{-3} (black triangles). (*c*) Standard *CNS* correction, with grid step size 1/10 of the resolution. The solvent density refined to 0.287 Å^{-3} (open triangles). (*d*) Modified procedure. The refined solvent density is 0.336 e Å^{-3} (black diamonds).

$B_{av,\,j}$ is defined for each grid point as the average of the B factors of the nearest atoms.

3. Results

The bulk-solvent contribution has been calculated for the 70S ribosome from *T. thermophilus* complexed with tRNAs and the *thrS* mRNA operator. The experimental data were collected at the SLS synchrotron source in Villigen, Switzerland (Jenner *et al.*, 2005). All

165 895 independent reflections with $I > 2\sigma(I)$ in the 300–5.5 Å resolution range were used in the structure-factor calculations.

The grid step size was set to 1/5 of the resolution. The *R* factors in the low-resolution range are shown in Fig. 2. With the standard *CNS* procedure, when $d_{\rm solv}$ is set to 0.34 e Å^{-3} and $B_{\rm solv}$ to the average atomic *B*, the agreement between the calculated structure factors and the experimental value is much worse than without solvent correction, with an *R* factor as high as 62% in the lowest-resolution shell (300–39 Å) (not shown). When $d_{\rm solv}$ is refined as in step (iv) of the procedure above, *R* values similar to those without correction are calculated, but the solvent density is now 0.053 e Å^{-3}, an unrealistically low value, which shows that the calculation tends to eliminate the solvent contribution.

With the modified procedure and the same grid step, the *R* values at low resolution are much smaller. Furthermore, $d_{\rm solv}$ refines to 0.336 e Å^{-3}, a value very close to the electron density of water.

To obtain more reasonable results with the standard procedure of *CNS*, a very fine grid has to be used. Fig. 2 shows the low-resolution *R* factors for a grid step of 1/10 of the resolution. However, even with such a fine grid, the refined value of $d_{\rm solv}$, 0.287 e Å^{-3}, is still somewhat too small and the *R*-factor values are consistently higher than those of the modified procedure for resolutions lower than 15 Å. The results with the modified procedure are much less sensitive to the grid size.

We also calculated the solvent contribution to the diffraction assuming a variable solvent *B* factor, as in (2). After the calculation of $\mathbf{F}_{\rm solv,\ B}$ and to be consistent with the empirical approach followed above, the same four quantities of (1) were refined. $B_{\rm solv}$ is now an additional positive or negative isotropic solvent *B* factor. Even though the atomic *B* factors of the ribosome model are very high (average B = 280 Å^2) and show large variations (r.m.s. variation 75 Å^2), the results are practically identical to those obtained with a constant isotropic *B*. The reason is that the difference between the solvent contribution in the two calculations becomes significant only at relatively high resolution, where the solvent contribution to the diffraction is smallest. The conclusion is that a solvent correction with a constant isotropic *B* is probably sufficient in most cases.

4. Availability

The Fortran programs *SHRINKMASK* and *SOLVPAR* and a UNIX shell script running the mask defining procedure are available from rees@igbmc.u-strasbg.fr.

We are grateful to Clemens Schulze-Briese for his help in collecting the ribosome low-resolution diffraction data at the Swiss Light Source.

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Jenner, L., Romby, P., Rees, B., Schulze-Briese, C., Springer, M., Ehresmann, C., Ehresmann, B., Moras, D., Yusupova, G. & Yusupov, M. (2005). *Science*, **308**, 120–123.

Jiang, J. S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.