X-ray crystallography at subatomic resolution

Alberto Podjarny, Eduardo Howard, Andre Mitschler and Bernard Chevrier
  UPR de Biologie Structurale, IGBMC, 1 rue Laurent Fries, 67404 Illkirch, France
Claude Lecomte, Benoit Guillot, Virginie Pichon-Pesme and Christian Jelsch
  LCM³B, UMR 7036 CNRS, Faculté des Sciences, Université Nancy I, 54506, Vandoeuvre-lès-Nancy, France

The diffraction of X-rays by molecular crystals is the technique of reference for obtaining three-dimensional information about atomic positions and interactions, information essential for the comprehension of the function and the molecular mechanisms. In the case of small molecules, very precise high resolution measurements allowed the observation of hydrogen atoms of and bond electronic densities. Thus, relations could be established between the deviations from standard stereochemistry of spherical atomic models and the chemical reactivity. In the case of biological macromolecules, one could correlate the spatial arrangement of the components of proteins and nucleic acids to their biological function.

These two types of studies progressed independently during the two last decades, primarily because of the limited resolution of the macromolecular crystallographic results, 2 to 3 Å in the majority of the cases, against 0.5 Å or better for the small molecules. The resolution, which is the minimal separation of the crystal plans giving place to an observable X-ray diffraction spot, is indeed an essential parameter of a crystallographic study. It is directly related to the minimum distance separating the details of the electronic density. A resolution of 2 Å is sufficient to distinguish peptides from a protein or the bases of a nucleic acid, but not the individual atoms, and even less the bond densities.

In the last ten years, various technical improvements, ranging from better techniques of expression and crystallisation to the use of synchrotron sources for measurements of diffraction and algorithms of multipolar and quantum modelling, made it possible to improve considerably the resolution and the quality of the macromolecular models [1]. Biological structural studies with resolutions between 1.5 and 0.9 Å became more current. In this range of resolution, the individual atoms can be clearly distinguished and the hydrogen atoms start to appear. As the errors of atomic position are reduced of an order of magnitude (typically from 0.2 to 0.03 Å), the variations observed from standard stereochemistry start to be significant.

Since 1997, several structures were solved with a resolution better than 0.9 Å, in particular crambin [2], subtilisin [3] and aldose reductase [4]. With such a resolution, the level of the details observed in the best ordered areas approaches that of the small molecules studies. The hydrogen atoms and the bond densities are clearly visible, and the atomic errors of co-ordinates are reduced another order of magnitude (~0.003 Å), which makes the stereochemical differences highly significant. Estimation of the atomic charges starts to be possible.

In what follows, we will discuss two of these cases, crambin and aldose reductase, from the crystallisation and the determination of the structure to the relations between structural details and the reactivity.

Experimental methodology in high resolution studies of macromolecules

Crystal Growth

The growth of the crystals of a biological macromolecule is one of the principal stumbling blocks in the process of determination of the structure, because there is no obvious relation between the nature of the macromolecule and the optimum conditions for crystallisation. The method of crystallisation most frequently used is dialysis in vapour phase, during which a drop containing protein and a stabilising solution (mother liquor) is dyalised against a container containing a precipitating agent. The role of this agent is to establish a competition for water inside the drop, so that the concentration of protein in the drop will increase gradually until the limit of solubility. The protein then precipitates, and, in the successful outcomes, there is nucleation and appearance of crystals.

The characteristic of the macromolecular crystals is that each protein (or nucleic acid) is mainly surrounded by water, and that the contacts with the other macromolecules include only a fraction of total surface.

Very high resolution diffraction comes from highly ordered macromolecular crystals. This implies a very ordered macromolecule

▶ **Fig. 1:** Effect of seeding in the crystal growth: a) Dilution 1/10, too many points of nucleation, too many crystals b)Dilution 1/100, few points of nucleation, a crystal of very high quality
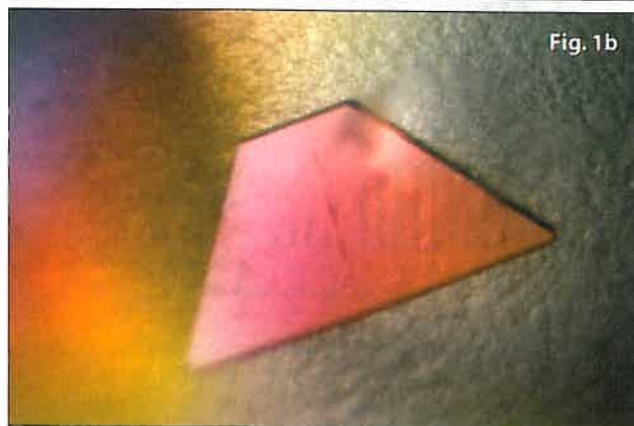
itself and a large crystal packing interface, and therefore a relatively small proportion of solvent. A very soluble, compact and monomeric molecule will tend to be ordered and to form large interfaces. In general the proteins present a great number of polymorphic varieties. Nevertheless, obtaining a crystal form adapted to high resolution remains largely a question of trial and error.
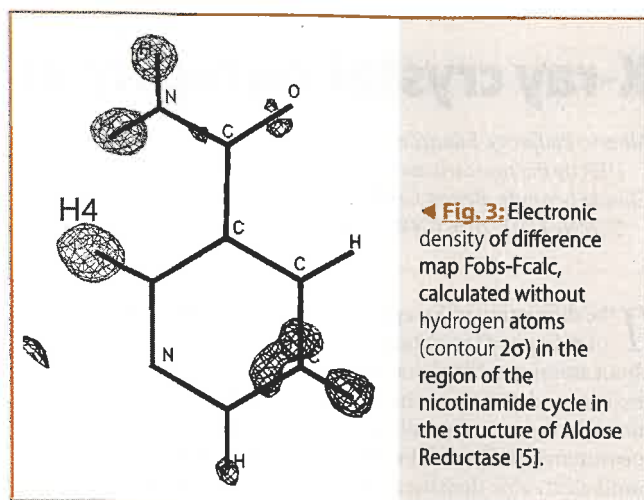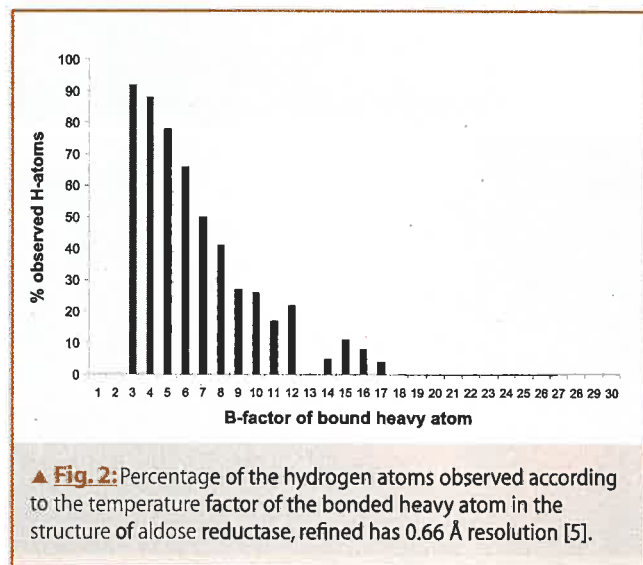
Once one form diffracting to high resolution is found, it is possible to improve it. In the case of aldose reductase [4], several parameters were adjusted to optimise the quality of the crystals, namely:

- crystallisation device: suspended drop or sitting drips;
- the volume of the drop (from 12 to 40 μl);
- the creation of points of nucleation by seeding with microcrystals;
- the temperature (4 or 24°C);
- the pH (from 5.0 to 6.5);
- the co-crystallisation with ligands.

An example of the effect of seeding is given in figure 1.

X-rays are obtained by acceleration (or deceleration) of electrons. The first sources were vacuum tubes, in which X-rays are emitted when a flow of electrons strikes an anode. The rotating anodes, which allow a better dissipation of the heat released during this process, were a first improvement, which allowed a brightness increase of two orders of magnitude. But the most decisive progress came from the synchrotron sources, where the acceleration of an electronic beam at relativistic speed produces highly collimated X-rays. The first synchrotrons offered a gain of brightness of 6 orders of magnitude, and the sources of third generation such as the ESRF (Grenoble) or the APS (Argonne) represent a gain of 14 orders of magnitude.

The brightness of the source is the principal factor in the measurement of the signal of diffraction to high resolution, because the average intensity of diffracted X-rays strongly decreases with the resolution. But it is not the only factor. Other developments were necessary, such as (i) the cryogenic techniques, to avoid or slow down the degradation of the crystal exposed to the beam of X-rays, (ii) the optics of the installation, which ensures a highly parallel and uniform beam, (iii) the technology of the detector, for the fast measurement of a large amount of diffraction data and (iv) the algorithms of reduction of the data. In the case of aldose reductase, the diffraction data were recorded from cooled crystals

◄ **Fig. 3:** Electronic density of difference map Fobs-Fcalc, calculated without hydrogen atoms (contour 2σ) in the region of the nicotinamide cycle in the structure of Aldose Reductase [5].

at 100° K, on a synchrotron of third generation, the Advanced Photon Source in Argonne, Illinois, with an optimised beam optics and a large CCD detector (22.1 × 22.1 cm²), fast and with low background noise.

Computing Methodology for ultra high resolution crystallography

Independent atom model refinement (IAM)

Single crystal X-ray diffraction data lead to structure factor amplitudes. The structure factors are the Fourier transform of the electron density $\rho_{dyn}(\mathbf{r})$ of the unit cell of volume V and parameters a_i, i =1, 3.

$$F(\mathbf{H}) = \int_{unit\,cell} \rho_{dyn}(\mathbf{r}) \exp(2\pi i \mathbf{H}.\mathbf{r})$$

$|\mathbf{H}| = 2\sin\theta/\lambda$, θ is the Bragg angle, λ is the wavelength

$$\rho_{dyn}(\mathbf{r}) = \int_{unit\,cell} \rho_{static}(\mathbf{r}-\mathbf{u}) \, P(\mathbf{u}) \, d^3\mathbf{u}$$

$P(\mathbf{u})$ is an atomic probability distribution function which Fourier transform is the Debye Waller factor. The crystal is triply periodic, therefore the Fourier transform has non zero values only on reciprocal lattice points defined by the reciprocal vectors

$$\mathbf{a*}_i = (\mathbf{a}_j \wedge \mathbf{a}_k) \, V^{-1}$$
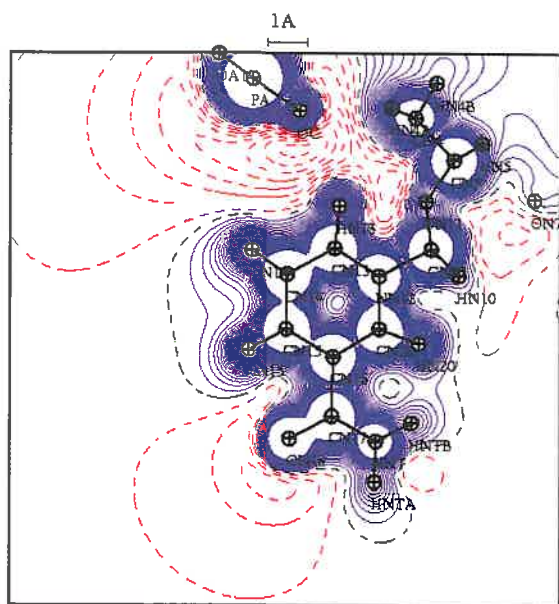
$F(\mathbf{H})$ are complex quantities and we must know both their amplitude and phase for directly calculating $\rho_{dyn}(\mathbf{r})$ by inverse Fourier transform. Methods for ab initio phase determination are still under development but they are not the aim of this paper.

As the electron density is mainly concentrated around atomic positions, the structure factor may be expressed as:

$$F(\mathbf{H}) = \sum_j f_j(|\mathbf{H}|) \exp(2\pi i \mathbf{H}.\mathbf{r}_j) \exp(-0.25 B_j|\mathbf{H}|^2)$$

Where **r** are the atomic positions, B_j the Debye Waller factors and f_j the atomic scattering factors; these latter the are Fourier transform of the electron density of the free neutral atom (Independent Atom Model, I.A.M.).

This equation is the basic one for most macromolecular crystallographic refinements which fit the observed |F|'s values using a model with four parameters per atom, i.e. atomic coordinates and isotropic Debye Waller factor. At the usual resolution for macromolecular crystallography (2Å<d<3Å), the resolution and

▲ **Fig. 2:** Percentage of the hydrogen atoms observed according to the temperature factor of the bonded heavy atom in the structure of aldose reductase, refined has 0.66 Å resolution [5].

▲ Fig. 4: Electrostatic potential generated by NADP⁺ (calculated from subatomic resolution diffraction data). Contours 0.1 e Å⁻¹, continuous line V > 0, dotted line V < 0.

therefore the number of observations |F| is not enough for determining the total number of parameters and it has to be completed with relations imposing a standard stereochemistry for the polypeptidic chain. At higher resolution one can observe deviations to the standard geometry. If the data are at atomic resolution (d<1.2Å), the isotropic temperature factor may be replaced by an anisotropic factor and the resulting accuracy of atomic positions is good enough to validate shifts from the standard geometry.
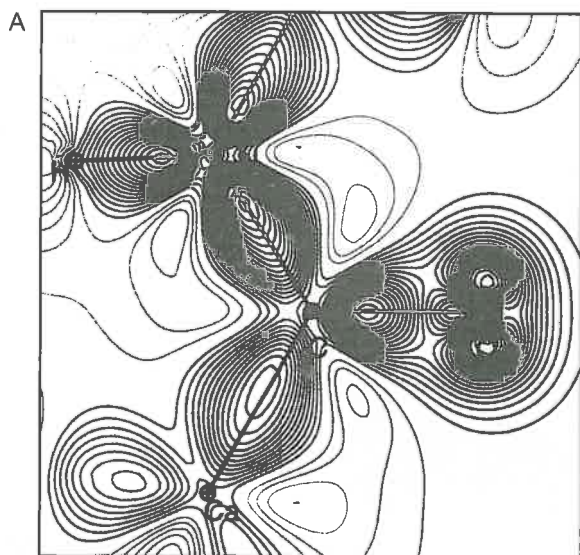
▲ Fig. 5: Model deformation density of the peptide group calculated from the electron density data base [12]. Contours 0.05 e Å⁻³, continuous ρ > 0, dotted ρ< 0.

Subatomic resolution (d<0.8 Å): hydrogen atoms

At subatomic resolution, d<0.8Å, informations on valence electron density distribution may be obtained when the B Debye Waller factor is lower than $4Å^2$. Hydrogen atoms also clearly show up. Deviations from the spherical atom model appear as electron density peaks in the bonds on deformation electron density maps (calculated by the difference between the observed electron density and the IAM density).

In aldose reductase 54% of hydrogen atoms were identified as well as most of the bonding density in the bonds of the active site of the protein. The probability to observe these features is directly related to the B factor [2, 5] (figure 2).

Finding the position for H atoms is extremely important for the catalytic reaction because it determines the protonation state and therefore the activity of the catalytic residues [5]. Figure 3 shows the protonation state of the nicotinamide cycle in the active site of aldose reductase. This cycle is a part of NADPH coenzyme which is transformed to NADP⁺ during the enzymatic reaction. The hydrogen atom H4 clearly shows up which demonstrates that the coenzyme is in the NADP⁺ state.

Charge density refinement: the multipolar model

The IAM model is too primitive to take into account all the information existing at subatomic resolution and a new model derived from small molecules crystallography has been developed which is called charge density refinement [6, 7].

In contrast to the IAM model where all atoms of a molecule or protein are supposed to be neutral with a spherical valence electron distribution (promolecule), the valence charge density is modelled by a sum of multipolar pseudo atoms lying at atomic positions. The valence electron density of such a pseudo atom is projected on the basis of real spherical harmonics functions centred on each pseudoatom.

$$\rho_{static}(\mathbf{r}) = \rho_{coeur}(r) + P_v\kappa^3\rho_v(\kappa r) + \sum_{l=0}^{lmax}P_{lmp}R_l(r)y_{lmp}(\theta,\varphi) \quad (1)$$

The radial functions used are of Slater type. First applications of this formalism to mono or dipeptides were calibration of ab initio HF calculations: it was clearly shown that triple zeta basis sets with polarisation functions were necessary to quantitatively reproduce the X-ray diffraction experiment [8]. The P_v, P_{lmp}, κ, κ' (dilation contraction of the density) are directly obtained from least squares refinement against the structure factors amplitudes [2, 7].

This analytical representation of the charge density is used to calculate crystal and molecular properties such as electrostatic potential, electric field, net charges, higher moments [9] and topology of the electron density [10]. As an example figure 4 gives the experimental electrostatic potential around NADP⁺, which is the cofactor of most enzymatic oxydo-reduction reactions. This potential was derived from X-ray multipolar analysis [11].

High resolution X-ray diffraction studies have been performed in Nancy on all amino-acids or monopeptides existing in Nature in order to precise their electron density distribution [12, 13].

These studies allowed building a data base of atomic charge density parameters (P_v, P_{lm}, κ, κ' see formula 1). These parameters were shown to be transferable to peptide functions in proteins [13]. Figure 5 gives the static deformation electron density calculated from the data base parameters for the CONH peptide group.

$$\delta\rho(r) = \sum_{j=1}^{Nat}\left[\rho_{base}(\mathbf{r} - \mathbf{r}_j) - \rho_{promolecule}(\mathbf{r} - \mathbf{r}_j)\right]$$

This deformation density reveals the valence electron redistribution due to covalent interactions. The electrons built up in the

features
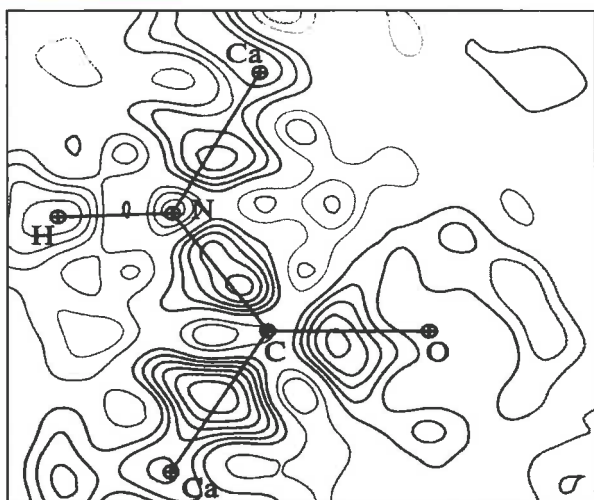
▲ **Fig. 6:** Residual electron density averaged over the 34 non disordered peptide groups of crambin after an IAM modelling. Contours as figure 5.

bonding region and the oxygen lone pairs are clearly evidenced. Then, the next step has been testing this data base on a protein. X-ray diffraction data on crambin, which is a small 46 residues protein, have been measured on BW7A line of DORIS (Hamburg) synchrotron to a resolution of d = 0.54 Å, which is still the world record for a protein [14]. This protein possesses all necessary criteria like low Debye Waller factors (B ~3 Å^2) for ordered parts of the protein. Taking advantage of the repetition of the same CONH chemical motif along the polypeptide main chair, the average dynamic deformation map over the 34 non disordered peptide residues was calculated according to:

$$\delta\rho(r) = \sum_{H'} (F_o - F_c) \exp(i\varphi_c) \exp(-2i\pi \, H.r)$$

F_c/φ_c are respectively the structure factor amplitude and the phase calculated from the IAM model (neutral, spherical atoms). F_o is the structure factor amplitude derived from the synchrotron experiment.

This average deformation density map displays significant residual density in the bonds between non hydrogen atoms and on oxygen lone pairs. These features clearly demonstrate that the IAM model does not provide an adequate fit to the experimental diffraction data.

After transfer of the statistically significant multipole from the data base and after multipolar refinement with MOPRO [7], the residual density does not exceed 0.06 e Å^{-3} which is about the estimated error: the progressive flattening of residual density features through the refinement stages is convincing physical evidence of real improvement in the modelling [2]. At the end of the refinement, the static deformation electron density of the average peptide residue (figure 7) is in almost quantitative agreement with that derived from an triple zeta HF calculation on a single monopeptide [8].

Conclusion

Ultra high resolution structure determination opens a new frontier in macromolecular crystallography; observing protonation states and shifts to standard geometry shows that protein structures are not rigid entities on which functionally active residues are linked, but active machineries in which most residues play an active role for the fine tuning of the mechanism. Subatomic resolution studies also allow determination of charge and electronic distribution and electrostatic potential for the atoms in the active site of enzymes and enable a better understanding of their function. New important applications are foreseen in the determination of electronic properties and oxidation states of reactive metallic centres in redox and electron transfer metalloproteins. With the enhancement of charge density modelling, the development of the data base of transferable parameters [12] and the continuing technological advances in experimental synchrotron crystallography, analysing the electronic structure of macromolecules has considerable unexplored potential.

Acknowledgements
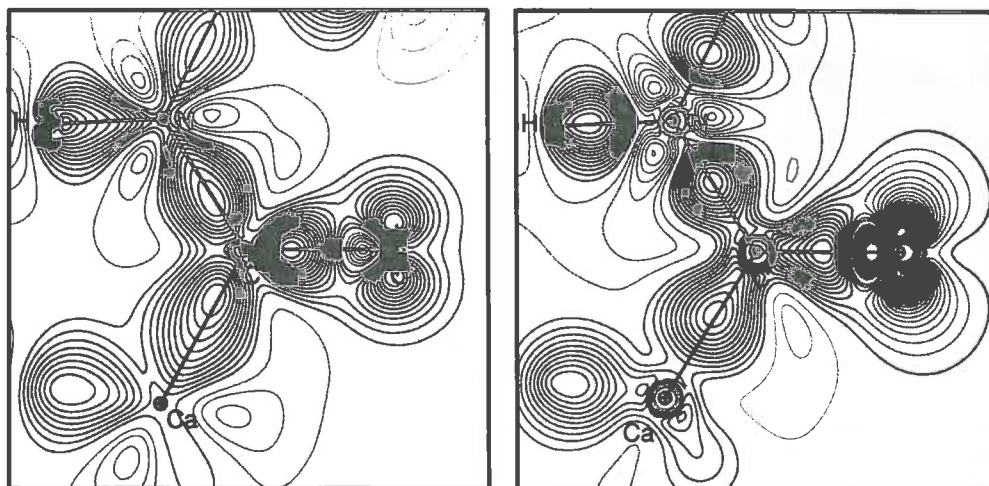
◀ **Fig. 7:** Static deformation density of a peptide plane in crambine a), compared to a HF SCF calculation on a single peptide b); contours: some as figure 5.

(INSERM), by the Hôpital Universitaire de Strasbourg (H.U.S), and by Institute for Diabetes Discovery, Inc.

References

[1] Dauter, Z., Lamzin, V.S. & Wilson, K.S. (1995) *Curr. Opin. Struct. Biol.* **6**, 784-90.

[2] Jelsch, C., Teeter, M.M., Lamzin, V., Pichon-Pesme, V., Blessing, R.H. & Lecomte, C. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3171-3176.

[3] Kuhn, P., Knapp, M., Soltis, M., Ganshaw, G., Thoene, M. & Bott, R. (1998) *Biochemistry* **37**, 13446-13452.

[4] Howard, E., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Joachimiak, A., Sanishvili, R., Van Zandt, M., Sibley, E., Moras, D. & Podjarny, A. (2002) *Sent to publication.*

[5] Cachau, R., Howard, E., Barth, P., Mitschler, A., Chevrier, B., Lamour, V., Joachimiak, A., Sanishvili, R., Van Zandt, M., Sibley, E., Moras, D. & Podjarny, A. (2000) *Journal de Physique* **10**, 3-13.

[6] Hansen, N.K. & Coppens P. (1978) *Acta Cryst.* **A34**, 909-921.

[7] Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001) *J. Appl. Cryst.* **34**, 214-223.

[8] Souhassou, M., Lecomte, C., Ghermani, N.E., Rohmer, M.M., Wiest, R., Bénard, M. & Blessing, R.H. (1992) *J. Amer. Chem. Soc.* **108**, 6942-6950.

[9] Bouhmaida, N., Ghermani, N.E., Lecomte, C. & Thalal, A. (1999) *Acta Cryst* **A55**, 729-739.

[10] Souhassou, M. & Blessing, R.H. (1999) *J. Appl. Cryst.* **32**, 210-217.

[11] Guillot, B., Jelsch, C., Muzet, N. & Lecomte, C. *Submitted for publication.*

[12] Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995) *J. Phys. Chem.* **99**, 6242-6250.

[13] Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998) *Acta Cryst.* **D54**, 1306-1318.

[14] Yamano, A., Heo, N.H. & Teeter, M. (1997) *J. Mol. Biol.* **230**, 292-311.