J. Appl. Cryst. (1999). **32**, 1084–1089

Combined Rietveld and stereochemical restraint refinement of a protein crystal structure

R. B. Von Dreele

Manuel Lujan Jr Neutron Scattering Center, MS H805, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. E-mail: vondreele@lanl.gov

(*Received 11 April 1999; accepted 5 August 1999*)

Abstract

By combining high-resolution X-ray powder diffraction data and stereochemical restraints, Rietveld refinement of protein crystal structures has been shown to be feasible. A refinement of the 1261-atom protein metmyoglobin was achieved by combining 5338 stereochemical restraints with a 4648-step ($d_{\min}$ = 3.3 Å) powder diffraction pattern to give the residuals R_{wp} = 2.32%, R_p = 1.66%, $R(F^2)$ = 3.10%. The resulting tertiary structure of the protein is essentially identical to that obtained from previous single-crystal studies.

1. Introduction

Until recently, powder diffraction was considered to be useful for phase identification and quantitative phase analysis only (Alexander, 1976). The development of the Rietveld method for refining crystal structures, first from neutron powder diffraction data (Rietveld, 1969) and then later from X-ray powder diffraction data (Young *et al.*, 1977), has led to wide application of the method to complex oxides, zeolites and most recently small organic molecules (Cheetham & Taylor, 1977; Poojary & Clearfield, 1997; Harris & Tremayne, 1996). Consequently, it has had a significant impact on various materials sciences. Construction of high-resolution X-ray powder diffractometers (Cox *et al.*, 1986) and accurate descriptions of the diffraction line shape (Finger *et al.*, 1994) have allowed Rietveld structure refinement to rival the results commonly obtained from single-crystal diffraction data (Dinnebier *et al.*, 1998). These developments now provide the opportunity to extend the application of Rietveld refinement techniques to protein structures. Here the first refinement of a protein structure from a combination of powder diffraction data and stereochemical restraints is described and it is shown that the quality of the result is sufficient to give the tertiary structure of the protein. Metmyoglobin was selected for the first test of this refinement method because it is of moderate size (*ca* 1260 non-hydrogen atoms excluding solvent and counterions) and has a well known structure (Kendrew *et al.*, 1960; Takano, 1977, 1984; Yang & Phillips, 1996). The results suggest that Rietveld refinement of protein structures from high-resolution powder diffraction data is feasible and that application of this method may have a significant impact on future exploration of protein structure and function.

2. Stereochemical restraints

The primary and secondary structures of proteins are very well known from high-resolution single-crystal protein structural studies (Morris *et al.*, 1992) as well as studies of small peptides (Engh & Huber, 1991). Consequently, the practice of augmenting the frequently low-resolution ($d_{\min}$ > 2.5 Å) single-crystal diffraction data obtained from most proteins with a set of restraints on bond lengths, bond angles, group planarities, *etc*., is widely applied for the refinement of protein crystal structures (Konnert & Hendrickson, 1980; Driessen *et al.*, 1989). These restraints are required because the number of reflections obtained at less than 2.5 Å resolution gives less than one observation for each positional parameter, making the refinement problem underdetermined when based on the diffraction data alone (Sussman, 1984). In a powder diffraction experiment, the three-dimensional diffraction pattern one would obtain from a single crystal has been collapsed into a single dimension, resulting in a great loss of information. To overcome this loss, the restraint schemes used for single-crystal protein refinements have been adopted and applied to powder diffraction data in a Rietveld refinement procedure. The function minimized takes the form

$$\begin{aligned} M = {} & \sum w_{Yi}(Y_{oi} - Y_{ci})^2 + f_a \sum w_{ai}(a_{oi} - a_{ci})^2 \\ & + f_d \sum w_{di}(d_{oi} - d_{ci})^2 + f_t \sum w_{ti}(t_{oi} - t_{ci})^2 \\ & + f_p \sum w_{pi}(-p_{ci})^2 + f_v \sum w_{vi}(v_{oi} - v_{ci})^4 \\ & + f_h \sum w_{hi}(h_{oi} - h_{ci})^2 + f_x \sum w_{xi}(x_{oi} - x_{ci})^2 \\ & + f_R \sum w_{Ri}(-R_{ci})^2, \end{aligned}$$

where the first term is the standard Rietveld refinement minimization function for powder diffraction data. The others are stereochemical minimization functions for bond angles, bond distances, torsion angles, planar groups, van der Waals repulsions, hydrogen bonds, chiral volumes and φ/ψ surface pseudopotentials, respectively. Each restraint term has a weight, $w_i = 1/\sigma_i^2$; the indivi-

Table 1. *φ/ψ pseudopotential surface coefficients*

Values in parentheses are standard uncertainties obtained from the least-squares fit of the φ/ψ pseudopotential function to a scored Ramachandran plot.

Term	A	φ_0	ψ_0	B ($\times 10^3$)	C ($\times 10^3$)	D ($\times 10^3$)
1	−3.23 (5)	−93.5 (12)	−25.7 (13)	0.142 (5)	0.209 (16)	0.057 (6)
2	−3.53 (6)	−115.0 (10)	144.4 (12)	0.134 (4)	0.154 (11)	0.019 (4)
3	−2.59 (10)	56.8 (10)	48.6 (16)	0.86 (7)	0.309 (24)	0.132 (29)
4	−1.85 (18)	54.2 (15)	−151.8 (19)	2.0 (4)	1.28 (24)	0.08 (21)

dual σ_i values are estimated from, for example, the distribution of bond lengths in small peptides (Engh & Huber, 1991). The weight factors (f_a, f_d, *etc.*) are normally set to unity during refinement; however, their values can be adjusted to increase or reduce the effect of each restraint component on the minimization function.

The φ/ψ torsion-angle pseudopotential term was developed for this work by assigning an energy score to each region of the Ramachandran plot (Ramachandran *et al.*, 1963) (core = −3, allowed = −2, generous = −1 and disallowed = 0) and by fitting a sum of two-dimensional Gaussians,

$$R_c = A_0 + \sum_{i=1}^{n} A_i \exp[-B_i(\phi_{oi} - \phi_c)^2 - C_i(\varphi_{oi} - \varphi_c)^2 - D_i(\phi_{oi} - \phi_c)(\varphi_{oi} - \varphi_c)],$$

to the resulting φ,ψ surface. A constant bias ($A_0 \simeq 3$) is used to place the minimum in the pseudopotential at zero. Four terms are used to represent the surface; each term has six coefficients, A_i, B_i, C_i, D_i, φ_{oi} and ψ_{oi}; these are listed in Table 1. A contour plot of the φ/ψ surface is shown in Fig. 1. The values of φ_c, ψ_c are obtained from each amino acid in the protein (excepting glycine) to calculate this contribution to the minimization function.

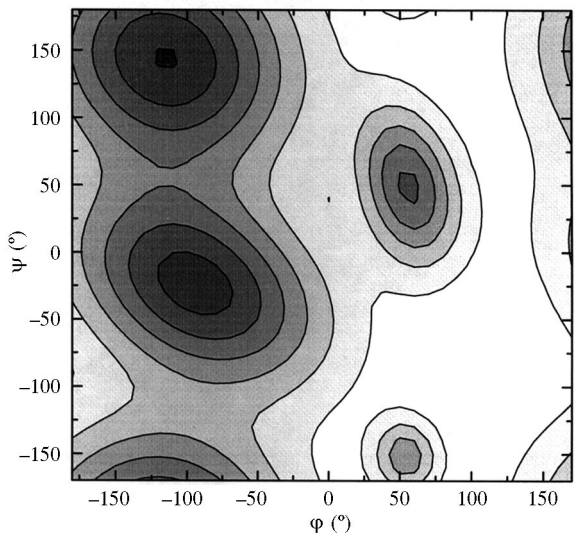


Fig. 1. φ/ψ torsion-angle pseudopotential surface obtained by fitting a four-term two-dimensional Gaussian function to the Ramachandran plot. The darkest areas are the fully allowed core regions and the lightest areas are the disallowed regions.

3. Experimental

3.1. *Sample preparation and data collection*

A polycrystalline sample of sperm whale (Physeter catodon) metmyoglobin was prepared by gentle grinding of small single crystals grown from a saturated solution of $(NH_4)_2SO_4$ buffered 1:1 to pH 6.37 with 4 *M* NaH_2PO_4/K_2HPO_4; the pH was adjusted with NaOH and H_2SO_4. A slurry of the powder in mother liquor was placed in a thin-walled silica capillary of 1.5 mm diameter; the capillary was then sealed with modeling clay to prevent evaporation of the solvent. The sample was ∼4 mm long. X-ray powder diffraction data were collected on X3b1 at the National Synchrotron Light Source, Brookhaven National Laboratory, equipped with a double Si(111) monochromator and a Ge(111) analyzer, with λ = 1.14991 (1) Å, from 1.5 to 20° 2θ using 0.004° steps at 5 s step^{-1}; the sample was spun during data collection to ensure powder averaging. The wavelength was established by use of the NIST SRM1976 flat-plate alumina standard. Two scans were collected; these were identical, indicating that no sample degradation occurred from radiation exposure. The two scans were combined for subsequent data analysis (Fig. 2).

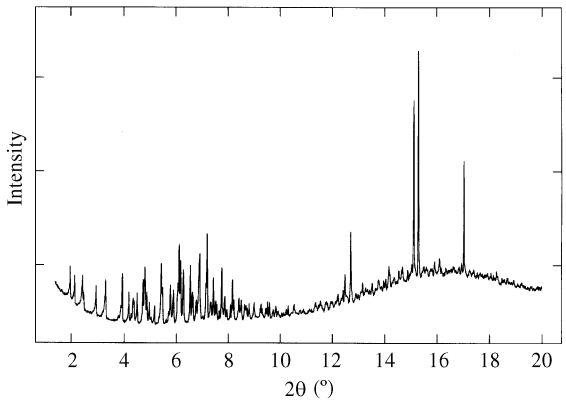


Fig. 2. The high-resolution X-ray powder diffraction pattern of whale metmyoglobin. The peaks at low 2θ are strongly asymmetric due to axial divergence in the instrument collimation. The large peaks at 12–18° 2θ are due to a second crystalline phase, $(NH_4)_2SO_4$. The background arises from diffuse scattering from the silica capillary and the mother liquor in the sample slurry.

3.2. *Structure refinement*

A modified version of the *General Structure Analysis System* (*GSAS*; Larson & Von Dreele, 1994) was used for Rietveld refinement of the metmyoglobin structure. The starting model was obtained from entry 4MBN (Takano, 1977) in the Brookhaven Protein Data Bank; all water molecules and the SO_4^{2-} groups were stripped from the structure before refinement. The powder diffraction profiles were modeled with the function described by Finger *et al.* (1994) and the background, which is a combination of amorphous scattering from the silica capillary and the liquid part of the slurry sample, was modeled with a 20-term constant-interval linear interpolation function. Solvent scattering was modeled following the procedures used by the protein structure refinement program *RESTRAIN* (Driessen *et al.*, 1989) which modify the atomic scattering factors. The eight diffraction peaks for the second phase, $(NH_4)_2SO_4$, were modeled using the LeBail method (LeBail *et al.*, 1988); only lattice parameters and profile coefficients were refined for this phase. After an initial refinement of non-structural parameters (background, profile coefficients, lattice parameters, solvent correction and diffractometer zero point), an initial fit with R_{wp} = 5.60%, R_p = 3.52% and $R(F^2)$ = 16.28% was obtained. There were clear misfits between the observed and calculated powder diffraction profiles throughout the entire diffraction pattern. These were particularly evident in the low-angle portion of the pattern. Structure refinement was initiated after developing a set of restraints as described above. There were 5338 restraints used in the final refinement cycles. Least-squares refinement of the structure was performed by use of six large (788–823 parameters) overlapping blocks. Each block was subjected to a single cycle of refinement; six successive cycles spanned the six blocks in sequence. Several adjustments of the stereochemical restraint weight factors were made during the refinement to achieve a more global minimum. After ~900 refinement cycles, convergence was obtained with R_{wp} = 2.32%, R_p = 1.66%, χ^2 = 3.97 and $R(F^2)$ = 3.10%. Periodically during the refinement, the protein stereochemistry was examined with the *PROCHECK* (Laskowski *et al.*, 1993) suite of programs.

4. Discussion

The behavior of this first refinement of a protein structure from high-resolution synchrotron X-ray powder data is summarized in Fig. 3. In the early stages of the refinement, fitting of the diffraction data dominated the change in the minimization function, even at the expense of the stereochemical restraints. After about 50 refinement cycles, the fitting of the diffraction data was essentially complete and then further changes in the minimization function reflected changes in the stereochemical restraint contributions. The weight factors were manipulated in subsequent refinement cycles to allow the structure to find a more global minimum; these were reset to unity for the final refinement cycles. The values of the contributors to the minimization function at the completion of the refinement, along with the final values of the weight factors, are given in Table 2, and the final fit to the powder diffraction data is shown in Fig. 4.

A comparison of the starting 4MBN metmyoglobin structure and the results from this refinement is shown in Fig. 5. The tertiary structure of this protein obtained from the Rietveld refinement is essentially identical to that of the single-crystal structure. The overall arrangement and extent of the helix and coil regions of the polypeptide chain are the same for the two models and the position and orientation of the heme group are also unchanged. Detailed differences are evident, however, in the conformation of both the peptide backbone and the side chains. Fig. 6 shows the magnitude of the shift in the atom positions from those of 4MBN to the results of this refinement. The mean shift for the peptide backbone atoms was 0.94 Å; the largest shift was 4.38 Å. For all atoms the mean shift was 1.27 Å and the largest shift was 5.63 Å. However, the centroid of the metmyoglobin molecule shifted by only 0.17 Å; thus these shifts reflect changes in the molecular conformation rather than a simple shift in position. A second refinement, which followed a different scheme for manipulation of the restraint weight factors, gave a similar result. The largest atom shifts in the peptide backbone are at the ends of the polypeptide chain and for residues Phe123–Ala125 that are in a coil region (GH corner) of the protein. The largest atom shifts for the entire structure are in the larger and more extended side chains (Leu, Phe, Tyr, *etc.*); probably their orientations are not well determined from this low-resolution (d_{min} = 3.3 Å) powder diffraction data.

An evaluation (Table 3) of the whale metmyoglobin structure obtained from Rietveld refinement by the *PROCHECK* (Laskowski *et al.*, 1993) suite of programs shows the strong impact of the stereochemical restraints

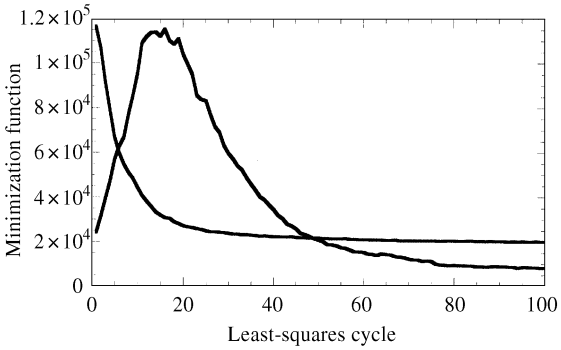


Fig. 3. Contributions to the minimization function for the first 100 cycles of least squares from the powder data and stereochemical restraints for the Rietveld refinement of whale metmyoglobin (see §4).

Table 2. *Terms in least-squares minimization function from Rietveld refinement of whale metmyoglobin*

Term	No.	Value
Powder data	4648	16295.00
Bond angles	1046	297.09
Bond distances	3040	850.66
Chiral volumes	156	62.55
Van der Waals and hydrogen interactions	314	1391.00
Planar groups	632	1738.70
φ/ψ surface	151	3876.30
Total	9987	24511.30

on the quality of the result. In particular, the bond lengths and angles show little deviation from the Engh & Huber (1991) small-molecule values, while the side-chain torsion angles (χ_1 and χ_2) that were not subject to any stereochemical restraints show a wider variation from expected values (Morris *et al.*, 1992). The effective resolution for this powder diffraction data obtained by comparing these χ values with those from single-crystal analyses (Laskowski *et al.*, 1993) is about 4.0 Å; this corresponds to a scattering angle of 16.5° 2θ in the powder pattern.

The lattice parameters obtained from this powder diffraction study are remarkably different from those obtained in single-crystal studies of whale metmyoglobin (Takano, 1977; Yang & Phillips, 1996). While the a and c axes and β angle are similar, the b axis is 0.77 Å (2.5%) shorter than that of 4MBN. This corresponds to a unit-cell volume that is 1610 Å^3 smaller. However, the powder diffraction pattern shows only slight evidence of any sample broadening that would arise from either particle size effects or microscopic strains commonly encountered after grinding. An analysis (Larson & Von Dreele, 1994) of the profile coefficients obtained from the Rietveld refinement gives an average metmyoglobin crystallite size of approximately 1 μm and a uniaxial 0.6% microstrain along the b axis. There was no microstrain effect for the ac plane. A possible explanation for this result is that the grinding or sample handling resulted in some loss of hydration of the metmyoglobin crystals that was not subsequently reabsorbed from the mother liquor. Consequently, some of the structural differences noted above may have their origin in this change in hydration. In particular, the GH corner noted above is located at an extreme edge of the molecule and is probably distorted from interaction with

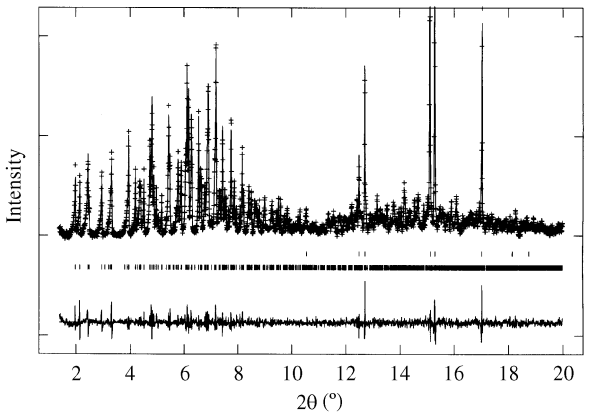
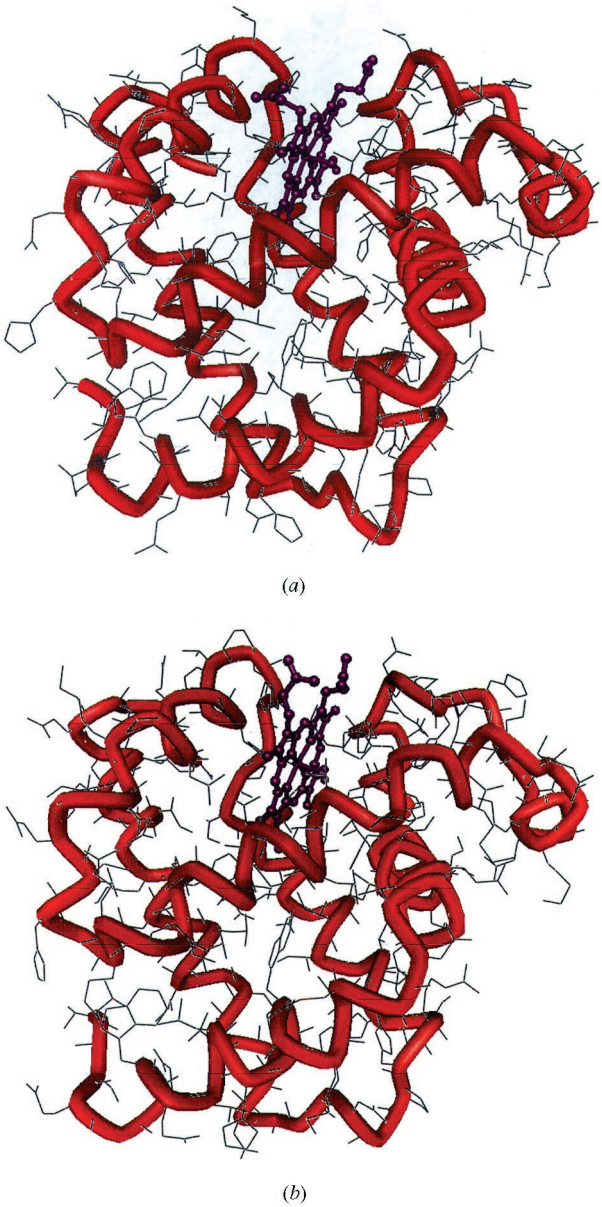


Fig. 4. X-ray powder diffraction profile from the final Rietveld refinement of whale metmyoglobin. Observed intensities are shown as crosses (+), calculated and difference curves as lines, and reflection positions for $(NH_4)_2SO_4$ and protein are shown as tick marks (|). The background intensity found in the refinement has been subtracted from the observed and calculated intensities for clarity.



(*a*)

(*b*)

Fig. 5. Molecular structure of metmyoglobin. The starting single-crystal model from 4MBN is shown in part (*a*) and the final Rietveld refinement result is shown in (*b*).

a neighboring molecule upon collapse of the *b* axis from the loss of hydration.

5. Conclusions

In this work, it has been shown that by the addition of stereochemical restraints, high-resolution powder diffraction data can be employed in a Rietveld refinement of a protein structure. Although this particular result is of relatively low structural resolution, further development of the method should provide significant improvement with a potentially very significant impact on protein crystallography. In particular, improved data collection strategies and adoption of cryocooling techniques should allow collection of useful powder data to $d_{\min} = 2.0$ Å. Further integration of Rietveld refinement with protein visualization and Fourier map display tools should allow full utilization of powder diffraction data in protein structure determination. It is anticipated that powder diffraction and Rietveld refinement of protein structures can readily be applied to structural studies of protein derivatives where a starting structural model for one member is known from a single-crystal study. Although a 'pure' *ab initio* protein structure determination from powder data is unlikely, the coupling of NMR (Wüthrich, 1995) and model-building techniques with powder diffraction may be a feasible route to solving a protein crystal structure from powder diffraction data.

The author thanks B. Schoenborn for the sample of whale metmyoglobin, and P. Stephens and R. Dinnebier for beam time on line X3b1 at NSLS and assistance in data collection. This work was supported by Basic Energy Sciences/US Department of Energy under contract W-7405-ENG-36 (LANSCE/LANL); the SUNY X3 beamline at NSLS is supported by Grant No. DE-FG02-86ER45231.

Table 3. *Summary of crystallographic and stereochemical parameters from program PROCHECK (Laskowski et al., 1993) for results of whale metmyoglobin Rietveld refinement*

Space group	$P2_1$
Lattice parameters	
a (Å)	64.4315 (11)
b (Å)	30.2046 (9)
c (Å)	34.9557 (5)
β (°)	105.838 (1)
V_c (Å^3)	65445.6 (22)
d-spacing range (Å)	$3.31 \leq d \leq 33.66$
Profile coefficients	
X	0.87 (6)
X_e	−0.37 (11)
Y_e	35.5 (21)
S/L	0.01218 (14)
H/L	0.00667 (17)
Main-chain parameters	
Bond error (Å)	0.003
Angle error (°)	0.88
Residues in core regions (%)	99.3
ω torsion angle standard deviation (°)	6.0
Bad contacts/100 residues	3.3
ζ angle standard deviation (°)	0.9
Hydrogen-bond energy standard deviation	1.26
Overall G factor	−0.27
Side-chain parameters	
Root-mean-square deviations from planes (Å)	0.01
χ_1 *gauche* minus standard deviation (°)	42.9
χ_1 *trans* standard deviation (°)	32.0
χ_1 *gauche* plus standard deviation (°)	30.5
χ_1 pooled standard deviation (°)	33.7
χ_2 *trans* standard deviation (°)	32.4
Morris *et al.* (1992) classification	
φ–ψ distribution	1
χ_1 standard deviation	4
Hydrogen-bond energy standard deviation	4

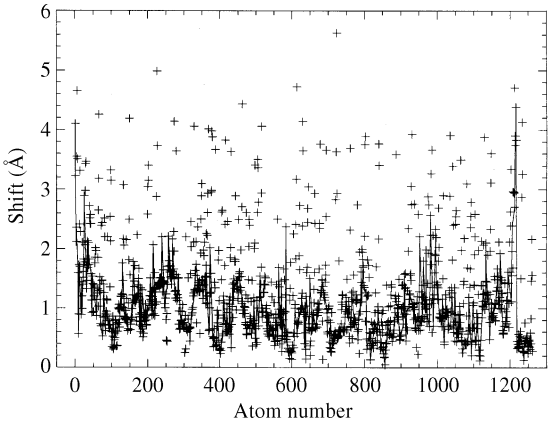


Fig. 6. Magnitude of atom shifts between single-crystal positions (4MBN) and positions from Rietveld refinement of metmyoglobin. Polypeptide main-chain atoms (N, C_α, C and O) and the heme core are shown as a line; all atoms are shown as crosses (+).

References

Alexander, L. E. (1976). *Advances in X-ray Analysis*, Vol. 20, edited by H. F. McMurdie, C. S. Barrett, J. B. Newkirk & C. O. Ruud, pp. 1–13. New York: Plenum.

Cheetham, A. K. & Taylor, J. C. (1977). *J. Solid State Chem.* **21**, 253–275.

Cox, D. E., Hastings, J. B., Cardoso, L. P. & Finger, L. W. (1986). *Mater. Sci. Forum*, **9**, 1–20.

Dinnebier, R. E., Pink, M., Sieler, J., Norby, P. & Stephens, P. W. (1998). *Inorg. Chem.* **37**, 4996–5000.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Kahn, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Finger, L. W., Cox, D. E. & Jephcoat, A. P. (1994). *J. Appl. Cryst.* **27**, 892–900.

Harris, K. D. M. & Tremayne, M. (1996). *Chem. Mater.* **8**, 2554–2570.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature (London)*, **185**, 422–427.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.
Larson, A. C. & Von Dreele, R. B. (1986). *General Structure Analysis System (GSAS)*, Los Alamos National Laboratory Report LAUR 86–748.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Le Bail, A. F., Duroy, H. & Fourquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
Poojary, D. M. & Clearfield, A. (1997). *Acc. Chem. Res.* **30**, 414–422.
Takano, T. (1977). *J. Mol. Biol.* **110**, 537–568.
Takano, T. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. R. Hall & T. Ashida, pp. 262–272. Oxford University Press.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
Sussman, J. L. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. R. Hall & T. Ashida, pp. 206–237. Oxford University Press.
Wüthrich, K. (1995). *Acta Cryst.* D**51**, 249–270.
Yang, F. & Phillips, G. N. Jr (1996). *J. Mol. Biol.* **256**, 762–774.
Young, R. A., Mackie, P. E. & Von Dreele, R. B. (1977). *J. Appl. Cryst.* **10**, 262–269.