Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Improved free *R* factors for cross-validation of macromolecular structure – importance for real-space refinement

Zhi Chen, Eric Blanc and Michael S. Chapman*

Department of Chemistry and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-3015, USA

Correspondence e-mail: chapman@sb.fsu.edu

Received 12 December 1997
Accepted 21 May 1998

Improvements in free *R* cross-validation are based on changed scaling procedures and the use, in map calculation, of estimates of the validation amplitudes which are independent of the actual observed values. The deleterious effects of the omitted test data are mitigated by reduction of the test-set size, which is made possible by constraining test and working sets to share the same scaling coefficients, thereby reducing the degrees of freedom and the dependence of free *R* on data selection. Further improvements come with use of a modified free *R* factor, $R_{T_A}^{\rm free}$. Instead of omitting the validation reflections from map calculation, their amplitudes are replaced by the average of resolution peers that is (nearly) independent of the actual cross-validation amplitudes. The improvements are relevant to model building, phase refinement by density modification and especially to real-space refinement. Although for real data at about 3 Å resolution, free *R* factors of about 0.25 are affected little, the precision of the structure is improved by about 0.1 Å. Tests with simulated data show that with good agreement between observed and calculated amplitudes (as in very high resolution studies or simulated refinement tests), free *R* factors can be improved by factors greater than two.

1. Introduction

Following the statistical concept of cross-validation, the free *R* value (Brünger, 1992*a*) has been established as an objective quality index of at least medium-resolution macromolecular structure and has been widely used in conjunction with reciprocal-space structure refinement (for a review, see Kleywegt & Brünger, 1996), especially to optimize refinement protocols. A free *R* factor (Brünger, 1992*a*) is calculated using only a 'test' subset, *T*, of the data that has been omitted from refinement. The statistics are, therefore, free of the artifactual lowering that can be caused by overfitting. Free *R* factors are, therefore, useful in selecting appropriate parameters for a reliable refinement. At the end of a very high resolution refinement, free *R* factors perhaps should not be calculated, because the deleterious effects of omitting data on the quality of the refinement (Sheldrick, 1997) outweigh limited information to be gleaned from small differences between free and conventional *R* factors at high resolution (Brünger, 1997). Furthermore, the effects of omitting a fixed proportion of data may be more severe in high-resolution refinements where stereochemical restraints typically contribute less to the atomic shifts. Deleterious effects also apply at lower resolution, but perhaps due to cynicism in the application of refinement methods, it has become a *de facto* standard to

quote free R factors, as exemplified by standard headers of Protein Data Bank files.

Free R factors are especially useful in comparing different refinement protocols. However, our research into real-space refinement (Chapman, 1995; Chapman & Blanc, 1997) suggested potential problems: tests with synthetic data revealed that the free R factor was not always well correlated with coordinate error. Our observations complemented those of others (Bacchi *et al.*, 1996) that omission of data reduces the convergence of real-space refinement, degrades the refined model and introduces spurious features in the density map.

In calculating an electron-density map, the effect of omitting the test set is the same as resetting their amplitudes to zero. In real-space refinement, the model is optimized to best agree with a map in which these amplitudes are zeroed. It is not surprising that this can introduce errors in the model. Thus, the effect of omitting data is much less benign in real space than in reciprocal space. In the automatic part of reciprocal-space refinement, omission of data is only slightly deleterious: the fullest use of experimental restraints is not made, but no artifacts are introduced. However, most refinement protocols alternate automatic optimization with interactive fitting of models into maps. For $R^{\rm free}$ to remain a cross-validated indicator, these maps should be calculated without the test set. Currently, test data are not always omitted (Kleywegt & Jones, 1997), perhaps because fitting, like real-space refinement, can be adversely affected by the map artifacts resulting from omission of data.

Here, methods are presented through which the deleterious effects of omitting the test set are mitigated. While their value is most obvious with regard to real-space refinement, they are also relevant to phase refinement by density modification and in the calculation of accurate electron-density maps. Therefore, the new procedures for cross-validation will facilitate all types of structure refinement and may extend the application of cross-validation to the high resolutions where the structure factors become correlated by Sayre's equation (Sayre, 1952). They also suggest improvements to cross-validated phase refinement by density averaging and modification (Cowtan & Main, 1996; Roberts & Brünger, 1995).

The opportunity is also taken to facilitate the incorporation of improved data sets without violating the principles of cross-validation. Scaling methods that account for solvent *etc*. are used in a way that does not introduce additional degrees of freedom that could artificially lower the free R factor for small validation sets.

2. Methods

2.1. Basic protocol

Synthetic data was partitioned into test and working sets in preparation for free R-factor calculation. Coordinates for the test structure were randomly displaced, then refined against the working-set data in real space and (for comparison) in reciprocal space. Free R factors were calculated in various ways and compared with the root-mean-square (r.m.s.) deviations between refined and ideal structures.

2.2. Free *R* calculation

Fundamentally, the procedures were similar to those of Brünger (1992*a*). A random fraction of the data was designated as 'test' (T) and not used for the refinement. The size of this portion was varied (see below). Differences to the Brünger (1992*a*) procedure implemented in *X-PLOR* (Brünger, 1992*b*) include the following. Scaling coefficients between observed amplitudes F_o and model F_c were calculated using the entire data, combining test (T) and working (W) reflections. The free R factor was then calculated using only T-set reflections and without further scaling. This has two advantages. Firstly, the scaling coefficients are independent of data selection and their use eliminates some of the variation previously observed between different test-data sets (Brünger, 1992*a*). Secondly, combination of the T and W sets has allowed us (in other work) to use more sophisticated scaling methods (Moews & Kretsinger, 1975; Tronrud, 1994) that account for the solvent continuum. With the larger $T + W$ combined data set, the additional degrees of freedom do not artificially reduce free R. The synthetic tests used here are without solvent, so the scaling routines from *X-PLOR* (Brünger, 1992*b*) were used.

The most important modification concerns the treatment of the T set. As will be detailed in the following section, instead of completely omitting them from map calculation, the observed amplitudes were replaced by estimates calculated from the average of observed amplitudes at similar resolution, or from the atomic model. How were we to judge whether a reduction in $R^{\rm free}$ reflected a real improvement in the model, or was an artifact perhaps due to biasing the $R^{\rm free}$ statistic? An additional set, T_o, was also omitted and used for standard, unmodified cross-validation. Comparisons could then be made between a standard $R^{\rm free}$, now denoted $R^{\rm free}_{T,\hat{T}_o}$, and some modified statistic, $R^{\rm free}_{T_A,\hat{T}_o}$. Complicated notation is needed to keep track of the different statistics. The subscripts to R list the (two) sets of test observations omitted from map calculation and refinement. A subscript to T other than o signifies observations that were replaced by estimates. R factors can be calculated from either T set, so a particular choice is designated with a circumflex: $\hat{T}$. A distinction is made between the standard $R^{\rm free}_T$ and $R^{\rm free}_{T_o}$, because omission of a second T set can impair refinement and elevate subsequent R factors. A comparison of $R^{\rm free}_{T,\hat{T}_o}$ and $R^{\rm free}_{T_A,\hat{T}_o}$ offers a reliable indicator of the effect of A-type T-set substitution upon the model, because the evaluation uses only the T_o reflections treated in a completely standard way. T_o is required only for the validation of indices reported in this paper and would not be required for model cross-validation in the future.

A procedure has been written to select test reflections on the basis of Miller indices rather than contents of a data file, achieving an effect similar to the *CCP*4 script *Uniqueify* (Collaborative Computational Project, Number 4, 1994), but without needing to save flags for each possible observation. Reflections can be added while preserving a prior test-set selection and maintaining a constant test-set fraction. Use of the Miller indices and an operating-system-independent

pseudo-random number generator (Press *et al.*, 1992) ensure that the same reflections are designated as test set when the procedure is run at different times or with different reflection files.

2.3. Map calculation

To mitigate the deleterious effects upon maps, rather than omitting the *T*-set reflections, they are replaced by a resolution-shell average of the *W*-set amplitudes. There are several precedents for the 'filling' of missing or weak observations: (*a*) the use of back-transformed amplitudes in the application of high-order non-crystallographic symmetry phase refinement (Chapman *et al.*, 1998; Rossmann, 1995); (*b*) Bayesian re-estimation of 'true' intensities in the data-normalization routines of Blessing *et al.* (1998) and (*c*) the use of model-derived amplitudes in σ_A difference maps (Read, 1986) used to check the map–model correlation following maximum-likelihood refinement (Murshudov *et al.*, 1997).

For our free *R* calculation, it is assumed that (at sub-atomic resolutions) the amplitudes of *T* reflections are only infinitesimally correlated to the average of their resolution peers. Although not strictly true in the presence of non-crystallographic symmetry, it remains a valid approximation because of the limited extent of the interference function (Rossmann, 1995) that correlates reflections of symmetry-equivalent regions of reciprocal space. This means that pairs of correlated reflections will be a tiny minority. Assuming insignificant correlation, assessment against the actual *T*-set amplitudes remains cross-validation, even after refinement against *T*-set substituted resolution averages.

$R_{T_A}^{\rm free}$ is calculated in the same way as $R_T^{\rm free}$, but the subscript *A* denotes (real-space) refinement against a map calculated with average-substituted *T* reflections. An alternative, $R_{T_C}^{\rm free}$ was also calculated analogously, except that the observed amplitudes of the *T* set were replaced by those calculated from the current model, instead of their resolution averages. With the simulated data used here, all reflections were used with unit weight.

2.4. Real-space refinement

The least-squares stereochemically restrained protocol was similar to that previously described (Chapman, 1995; Chapman & Blanc, 1997), and calculation of the model electron density accounted for the experimental resolution limits. Instead of using our previously described implementation that is an add-on module for *TNT* refinement (Tronrud *et al.*, 1987), a new equivalent was used in which real-space refinement is implemented through an alternative target for *X-PLOR* refinement (Brünger, 1992*b*; Chen *et al.*, 1999). This allowed the combination of real-space refinement with the *X-PLOR* stereochemical potentials and methods of optimization (Brünger *et al.*, 1997).

2.5. Test system

Synthetic amplitudes (4315 between 20 and 2 Å) were calculated from protein G (Derrick & Wigley, 1994; PDB accession number 1IGD; 576 protein non-H atoms). Prior to refinement, the 'ideal' model was degraded by random displacement of atoms to generate the initial model, which had an r.m.s. error of 0.53 Å and a conventional *R* factor of 0.357.

2.6. Test calculations

The dependence of standard $R_T^{\rm free}$ and modified $R_{T_A}^{\rm free}$ or $R_{T_C}^{\rm free}$ upon the size of the *T*, T_A and T_C sets was tested over the range 0–16% with or without the omission of an additional 3% T_o set. For each *R*-factor calculation, the degraded model was refined either in real space (Chapman, 1995; Chen *et al.*, 1999) or in reciprocal space (Brünger, 1992*b*), using conjugate-gradient optimization in both cases. Real- and reciprocal-space refinements used identical stereochemical potentials given by *X-PLOR* (Brünger, 1992*b*) and used equivalent weights. For averaging, the overall 20–2 Å resolution range was split into 20 d^* bins with approximately 200 reflections in each. Phases were calculated from the 'ideal' model.

3. Results and discussion

Fig. 1 shows that the consequences of omitting data are dire for real-space refinement. It is the omission of data, and not the refinement method, which is critical. It has previously been shown with synthesized data that real-space refinement gives a structure with tiny error (Chapman, 1995). The common *y* intercept of Fig. 1 confirms that when all of the (synthetic) data are used, the accuracies of the refinement methods are indistinguishable. However, it is striking that an r.m.s. error of 0.14 Å is introduced during real-space refinement by following the original free-*R* protocol (Brünger, 1992*a*) and omitting

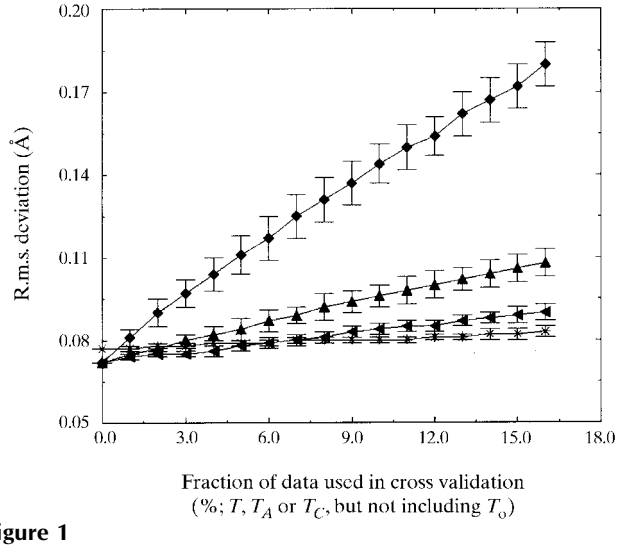

Figure 1
Accuracy of the structure following refinement while monitoring quality with several cross-validation indices. Standard error bars were generated through the use of 25 different random selections for the cross-validation data. $R_{T_A}^{\rm free}$ was calculated after refinement against maps with the resolution average substituted for the test reflections. $R_{T_C}^{\rm free}$ was similarly calculated with substitution of amplitudes calculated from the current model. ◆, Real-space refinement monitored by $R_T^{\rm free}$; ▲, real-space refinement monitored by $R_{T_A}^{\rm free}$; ◀, real-space refinement monitored by $R_{T_C}^{\rm free}$; ×, reciprocal-space refinement monitored by $R_T^{\rm free}$.

Table 1
Comparison of free *R*-factor statistics for real-space refined models.

Coordinates of the mannose-binding protein A (Burling *et al.*, 1996) were displaced by an r.m.s. of 0.33 Å by application of torsion-angle dynamics (Brünger & Rice, 1997) while ignoring the experimental data. The experimental data was partitioned into test (*T*) and working (*W*) sets. The model was refined against maps calculated using experimental phases (Burling & Brünger, 1994), either completely omitting ($R_T^{\rm free}$) or substituting independent approximations ($R_{T_A}^{\rm free}$, $R_{T_C}^{\rm free}$) for the *T*-set reflections. Refinement used a real-space implementation (Chen *et al.*, 1999) of torsion-angle dynamics (Brünger & Rice, 1997). Maps and *R* factors were calculated at 3 Å resolution. The r.m.s. error is calculated with respect to the published refined structure at 1.8 Å resolution (Burling & Brünger, 1994).

Statistic	Validation set size (%)	R.m.s. model error (Å)	Free *R*
$R_T^{\rm free}$ (Brünger, 1992*a*)	10	0.33	0.25
	2	0.22	0.26
$R_{T_A}^{\rm free}$ (this paper)	10	0.21	0.26
	2	0.22	0.26
$R_{T_C}^{\rm free}$ (this paper)	10	0.21	0.26
	2	0.24	0.26

10% of the data. By contrast, omission of this data has little consequence with reciprocal-space refinement, at least in this favorable error-free test case with a high data-to-parameter ratio. Through substitution of resolution-shell average amplitudes, most of the deleterious effects of data omission associated with real-space refinement can be overcome. The effects are nearly obliterated through use of model amplitudes (Fig. 1).

Fig. 2 shows that substitution of average amplitudes markedly reduces the free *R* factor. The conventional free *R* for real-space refined models is high for two reasons. The increasing *R* factor is likely to result from model degradation due to real-space refinement against incomplete data (see also Fig. 1). The high *y* intercept (0.13) for the standard free *R* is likely to result from implicitly restraining model *T*-set amplitudes to zero immediately before *R*-factor calculation. Use of average amplitudes (instead of zero) leads to partial correction (*y* intercept of 0.07). Use of model amplitudes roughly halves the remaining discrepancy with $R_T^{\rm free}$ for reciprocal-space refinement (Fig. 2). $R_{T_A}^{\rm free}$ for real-space refinement still exceeds $R_T^{\rm free}$ for reciprocal-space refinement, even with small *T*, for which the model qualities are similar (Fig. 1). This indicates that $R_{T_A}^{\rm free}$ is a more conservative estimate of model quality.

Comparison of Figs. 1 and 2 has interesting implications for the optimal size of the cross-validation set. Similar plots, following reciprocal-space refinement (Fig. 6 in Brünger, 1997), suggested to Brünger that a suitable standard test set balanced precision of $R^{\rm free}$ with model error and was the greater of 10% or 500 of the reflections. They also led to the postulate that the error of $R^{\rm free}$ approximates $R^{\rm free}/n^{1/2}$, where *n* is the number of test-set reflections. Fig. 2 is consistent with the prior data and shows that Brünger's postulate holds approximately true for smaller test sets than previously evaluated, including a 43-reflection (1%) test set.

More importantly, Figs. 1 and 2 emphasize the different dependence of $R^{\rm free}$ on test-set size following real- and reciprocal-space operations. The more severe consequences of data omission in real space lead to a different optimal balance of $R^{\rm free}$ precision and model error while maps are still being used for model building or real-space refinement. A modified protocol is suggested that uses smaller test sets at such (lower resolution) stages of structure determination. Often, refinement starts at the common ∼3 Å resolution limit of experimental phases, but later extends to higher resolution – for example, 2 Å. A test set is selected to satisfy the Brünger standard at high resolution (ready for the final reciprocal-space refinement), even though there may be as few as 150 test-set reflections to 3 Å resolution. The minimal test set reduces map degradation at the starting resolution where real-space operations (averaging, building and refinement) are most critical. Early $R^{\rm free}$ are expected to have errors of ±0.08 $R^{\rm free}$, but, at a given resolution, evaluations will use the same test set, and sampling errors will therefore be systematic. Thus, even though the absolute value of $R^{\rm free}$ is imprecise, the relative values of $R^{\rm free}$ remain a reliable indicator of the progress of refinement. At the final resolution, the test set is now large enough to give $R^{\rm free}$ to the desired absolute precision.

Table 1 shows the effects of applying the modified cross-validation methods to real experimental data. The modified methods result in structures improved in precision by about 0.1 Å. Subtle differences between the various modifications are not apparent owing to the large discrepancies between model and experimental data and the small impact on typically large *R* factors ($R_{T_A}^{\rm free} \simeq 25\%$). Thus, the effect of the new methods on the statistics of real medium-resolution models is less than with very accurate ($R \simeq 0.1$) structures or with the

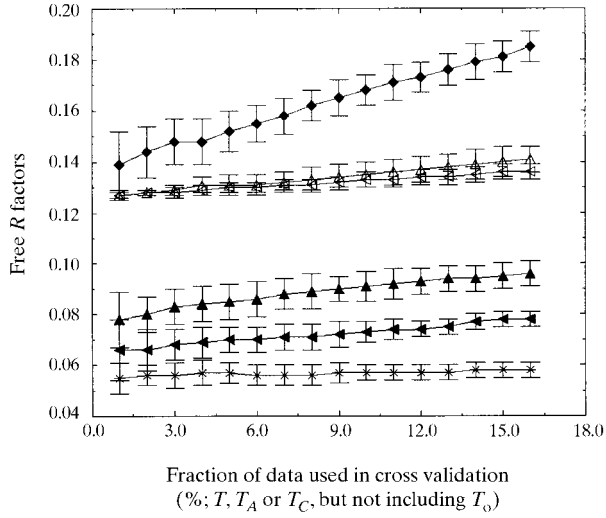

Figure 2
Comparison of different cross-validation statistics. $R_{T_A}^{\rm free}$ and $R_{T_A,\hat{T}_o}^{\rm free}$ were calculated after refinement against maps with the resolution average substituted for the test reflections. The graph of $R_{T_A}^{\rm free}$ was calculated using only the 3% T_o subset for which averages were not substituted, thereby enabling independent evaluation of the index. $R_{T_C}^{\rm free}$ and $R_{T_C,\hat{T}_o}^{\rm free}$ were similarly calculated with substitution of amplitudes calculated from the current model. ◆, Real-space refinement monitored by $R_T^{\rm free}$; △, real-space refinement monitored by $R_{T_A,\hat{T}_o}^{\rm free}$; ◁, real-space refinement monitored by $R_{T_A,\hat{T}_o}^{\rm free}$; ▲, real-space refinement monitored by $R_{T_A}^{\rm free}$; ◀, real-space refinement monitored by $R_{T_C}^{\rm free}$; ×, reciprocal-space refinement monitored by $R_T^{\rm free}$.

simulated models and data used for testing refinement methods. However, even if the modifications have an imperceptible effect on high *R*-factor values, they have a substantial impact on the precision of the atomic model.

In mitigating the deleterious effects of missing data, the potential for bias toward the current model has been introduced into the map. The bias of (F_o, φ_c) maps is well documented (Luzzati, 1953). Maps used in calculating $R_{T_A}^{\rm free}$ might carry slightly more bias, because the average-substituted *T*-set reflections have information from the current model in the phases (like the *W* set), but no offsetting information about the correct structure in the amplitudes. The bias could slow refinement and give maps with which it is difficult to diagnose model errors. Fig. 1 shows that bias is a lesser evil than data omission. In real applications, it would be wise to use well established methods for reducing and eliminating phase bias. These methods include difference and omit-map calculations (Bhat & Cohen, 1984; Hodel *et al.*, 1992; Read, 1986).

The bias remaining in omit maps calculated with average-substituted amplitudes should be no worse than in maps calculated with the real amplitudes, because all of the bias comes through the phases.

The potential for bias with $R_{T_C}^{\rm free}$ is greater owing to the presence of model information in some of the amplitudes and not just the phases. The authors of this report believe that the modest improvement in $R_{T_C}^{\rm free}$ over $R_{T_A}^{\rm free}$ as a statistic does not justify the increased risk of its use in refining a structure. As the truest indicator of the quality of real-space refinement, two limited uses of $R_{T_C}^{\rm free}$ are suggested: (i) use in a few terminal cycles of refinement if a conventional *R* factor does not suffice for a real-space refined final model and (ii) use in making subtle comparisons of refinement strategies. After finding the optimal strategy (or continuing beyond a 'terminal' refinement), changes made to the model during $R_{T_C}^{\rm free}$ monitoring should be discarded and the refinement repeated using the safer $R_{T_A}^{\rm free}$.

This work has focused on reflections that are missing from map calculation, because of their use in cross-validation. Reflections may also be missing because they have not been measured. Fig. 1 suggests that failure to measure reflections could also have deleterious effects on real-space refinement. The data sets used in our work were nearly complete, but in other cases estimation of missing scattering amplitudes by the methods discussed above may be beneficial. There is one danger. Unlike the test set, missing observations may not be a quasi-random subset. They may be preferentially weaker, overloaded or otherwise systematically different. Filling with the average amplitude of measured peers might, in this case, be an inappropriately simple remedy.

4. Conclusions

The improvements reported here lead to structures that are more accurate at times when data is being omitted for free *R*-factor calculation. Accuracy of intermediate models may not be critically important if refinement is completed using all data, as is common practice – at least at very high resolution (Sheldrick, 1997). At medium resolution, there is the possibility that refinement will be continued when better data become available. Some refrain from using the test data to avoid the possibility of mistakenly later calculating free *R* factors from data used to refine a forerunner model. Especially when refinement is in real space, the improved cross-validation impacts the quality of structures refined only against working-set data. The modified free *R* factors also give a more absolute measure of the quality of the structure, and are therefore important in comparing real- and reciprocal-space protocols of (intermediate) refinement.

The sources of the improvements reported here are changes in scaling that allow statistically reliable *R* factors to be calculated from fewer reflections, and approximation of the test-set reflections rather than complete omission. Fig. 1 demonstrates that these changes lead to more accurate maps. The changes will have greatest impact for real-space refinement. They will also be valuable in the more common case of alternated reciprocal-space refinement and interactive remodeling. Firstly, with smaller cross-validation sets, more data can be used for refinement. Secondly, the quality of the maps for interactive rebuilding can be improved through reduction of the cross-validation set and through approximation of the test reflections rather than omission.

Similar techniques should also lead to improvements in phase refinement by density modification or electron-density averaging. Use of cross-validation to monitor the progress of solvent flattening and density modification has recently been advocated (Cowtan & Main, 1996; Roberts & Brünger, 1995), but application is complicated by the need for complete cross-validation using distinct validation sets, and the different phase sets that these generate. One of the problems is common to that of electron-density averaging (for review, see Chapman, 1998), in that the electron density can not be expected to have exactly flat solvent, or exact equivalency between non-crystallographically related molecules (*etc.*) when maps are calculated omitting some of the data. The benefits of reducing the cross-validation set and approximating for test-set amplitudes in map calculation should simplify the application of cross-validation to phase refinement.

This work was generously supported by the National Science Foundation (MSC; BIR9418741). Real-space refinement programs are distributed from http://www.sb.fsu.edu/~rsref.

References

Bacchi, A., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 641–646.
Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
Blessing, R. H., Guo, D. Y. & Langs, D. A. (1998). In *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier. Dortrecht: Kluwer Academic Publishers.
Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1992*b*). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.
Brünger, A. T. (1997). *Methods Enzymol.* **277**, 366–396.

Brünger, A. T., Adams, P. D. & Rice, L. M. (1997). *NATO Advanced Study Institute on Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 149–163. Dortrecht: Kluwer.
Brünger, A. T. & Rice, L. M. (1997). *Methods Enzymol.* **277**, 243–269.
Burling, F. T. & Brünger, A. T. (1994). *Isr. J. Chem.* **34**, 165–174.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. (1996). *Science*, **271**, 72–77.
Chapman, M. S. (1995). *Acta Cryst.* A**51**, 69–80.
Chapman, M. S. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 99–108. Dordrecht: Kluwer.
Chapman, M. S. & Blanc, E. (1997). *Acta Cryst.* D**53**, 203–206.
Chapman, M. S., Blanc, E., Johnson, J. E., McKenna, R., Munshi, S., Rossmann, M. G. & Tsao, J. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 433–442. Dortrecht: Kluwer Academic Publishers.
Chen, Z., Blanc, E. & Chapman, M. S. (1999). *Acta Cryst.* D**55**. In the press.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Derrick, J. P. & Wigley, D. B. (1994). *J. Mol. Biol.* **243**, 906.
Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.
Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 208–230.
Luzzati, V. (1953). *Acta Cryst.* **6**, 142–152.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.
Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* D**53**, 240–255.
Press, W. H., Teukolsky, S. A., Vettering, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. New York: Cambridge University Press.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Roberts, A. L. U. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 990–1002.
Rossmann, M. G. (1995). *Curr. Opin. Struct. Biol.* **5**, 650–655.
Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
Sheldrick, G. M. (1997). *NATO Advanced Study Institute on Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 125–136. Dortrecht: Kluwer Academic Publishers.
Tronrud, D. E. (1994). *The TNT Refinement Package 5-E*. University of Oregon, Eugene, Oregon, USA.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.