High-resolution Structure Refinement

George M. Sheldrick
Institut für Anorganische Chemie der Universität Göttingen
Tammannstraße 4, D-37077 Göttingen, Germany
gsheldr@shelx.uni-ac.gwdg.de

Abstract

The 'small-molecule' program SHELXL is finding increased application in refinement of macromolecules against high resolution data. SHELXL is slow (compared with other macromolecular programs) but very general. Even with data to atomic resolution, restraints are still required for anisotropic refinement and handling disorder. Full-matrix refinement is useful in the initial rigid group refinement and can be used to estimate standard deviations for the final refined structure. NCS, anti-bumping and chirality restraints enable SHELXL to be employed at more moderate resolution too. The solvent model and factors affecting convergence are also discussed.

1 Introduction

Recent advances in cryogenic techniques, area detectors, and the use of synchrotron radiation enable macromolecular data to be collected to higher resolution than was previously possible. In practice this tends to complicate the refinement because it is possible to resolve finer details of the structure; it is often necessary to model alternative conformations, and in a few cases even anisotropic refinement is justified. Although SHELXL-96 [1] provides a number of other features not found in many macromolecular refinement programs (e.g. refinement against data from twinned crystals), it is probably the restrained anisotropic refinement, the flexible treatment of disorder and the estimation of esds that are most likely to be of interest to macromolecular crystallographers.

SHELXL is designed to be easy to use and general for all space groups, but uses a conventional structure-factor calculation rather than a FFT summation; the latter would be faster, but in practice involves some small approximations and is not very suitable for the treatment of dispersion or anisotropic thermal motion. The price to pay for the extra generality and precision is that SHELXL is much slower than programs written specifically for macromolecules, but this is to some extent compensated for by the better convergence properties, reducing the amount of manual intervention required (and also the *R*-factor).

An auxiliary program SHELXPRO is provided as an interface to other protein programs. SHELXPRO is able to generate an .ins file for input to SHELXL from a file in PDB format, including the appropriate restraints etc. SHELXPRO can also generate map files for O and other map interpretation programs, and can display the refinement results in the form of Postscript plots, as well as including the updated coordinates in the .ins file for the next refinement. SHELXL produces PDB and CIF format files that can be read by SHELXPRO and used for archiving.

Two input files are required to run SHELXL: an intensity data file name.hkl and a file name.ins that contains crystal data, atoms, restraints, instructions, etc. All the information in the .ins file is given by four-letter keywords followed by numerical data and references to atoms etc. in free format.

Macromolecular structures are conventionally divided up into residues, for example individual amino-acids. In SHELXL residues may be referenced either individually, by '_' followed by the appropriate residue number, or as all residues of a particular class, by '_' followed by the class. For example 'DFIX 2.031 SG_9 SG_31' could be used to restrain a disulfide distance between two cystine residues, whereas 'FLAT_PHE CB > CZ' would apply planarity restraints to all atoms between CB and CZ inclusive in all PHE (phenylalanine) residues. Plus and minus signs refer to the next and previous residue numbers respectively, so 'DFIX_* 1.329 C_– N' applies a bond length restraint to all peptide bonds ('_*' after the command name applies it to all residues). This way of referring to atoms and residues is not restricted to proteins; it is equally suitable for oligonucleotides, polysaccharides, or to structures containing a mixture of all three. It enables the necessary restraints and other instructions to be input in a concise and relatively self-explanatory manner. These instructions are checked by

the program for consistency and appropriate warnings are printed.

The more important instruction keywords for macromolecular refinement are summarized in the following table (* indicates significant changes from SHELXL-93):

DEFS Set global restraint esd defaults.

DFIX Restrain 1,2-distance to target (which may be a free variable).

DANG* Restrain 1,3-distance to target (may be a free variable).

SADI Restrain distances to be equal without specifying target.

SAME Generate SADI automatically for 1,2- and 1,3-distances.

CHIV* Restrain chiral volume to target (default zero; may be f.v.).

FLAT Planarity restraint.

DELU Generate rigid bond U_{ij} restraints using connectivity.

SIMU Generate similar U (or U_{ij}) restraints using distances.

ISOR 'Approximately isotropic' restraints.

BUMP* Generate anti-bumping restraints automatically.

NCSY* Generate non-crystallographic symmetry restraints.

FVAR Starting values for overall scale factor and free variables.

SUMP Restrain linear combination of free variables.

PART Atoms in same disorder component have same PART number.

AFIX Riding H, rigid groups and other geometric constraints.

HFIX Generate AFIX instructions to generate and refine H-atoms.

MERG Merge equivalent reflections (usually MERG 4).

SHEL Maximum and min. resolution (data ignored outside range).

SWAT* Refine diffuse solvent parameter (Babinet's principle).

HOPE* Refine anisotropic scaling.

WGHT Weighting scheme, probably best left at default 'WGHT 0.1' .

CGLS No. of conjugate gradient refinement cycles, select R_{free} data.

BLOC, L.S., DAMP Blocked-matrix least-squares (for esds).

RTAB, MPLA, HTAB* Tables of bonds, angles, planes, H-bonds etc.

WPDB, ACTA, LIST* Output PDB and CIF files for archiving.

2 Constrained and Restrained Refinement

A *constraint* is an exact mathematical condition that leads to the elimination of one or more least-squares variables; a *restraint* is an additional piece of information that is not exact but is associated with an esd. Restraints are normally applied as addition observational equations, i.e. they improve the data to parameter ratio by increasing the number of data but leaving the number of parameters unchanged.

The lower data to parameter ratio for macromolecules makes the use of constraints and especially restraints essential. Rigid group constraints enable a structure to be refined with very few parameters, especially when the (thermal) displacement parameters are held fixed (using the BLOC instruction in SHELXL). After a structure has been solved by molecular replacement using a rather approximate model for the whole protein or oligonucleotide, it may well be advisable to divide the structure up into relatively rigid domains (using a few AFIX 6 and AFIX 0 instructions) and to refine these as rigid groups, initially for a limited resolution shell (e.g. SHEL 8 3), then stepwise extending the resolution. Restraints may still be required to define flexible hinges and prevent the units from flying apart. In view of the small number of parameters and the high correlations introduced by rigid group refinement, L.S. (full-matrix refinement) should be used for this stage (but CGLS will be necessary for the subsequent refinement). After this initial step, which exploits the large convergence radius of rigid group refinement, in general the more flexible restraints will be used in preference to constraints for the rest of the refinement.

SHELXL provides distance, planarity and chiral volume restraints, but not torsion angle restraints or specific hydrogen bond restraints. For oligonucleotides, distance restraints [2] may be used, but for reasonably high resolution data it is probably better to assume that for the sugars and phosphates the chemically equivalent 1,2- and 1,3-distances are equal (using the SAME and SADI restraints) without the need to specify target values. In this way the effect of the pH on the protonation state of the phosphates and hence the P-O distances does not need to be predicted, but it is assumed the whole crystal is at the same pH. For proteins, since some amino-acid residues occur only a small number of times in a given protein, it is probably better to use 1,2- and 1,3-target distances based on the study of Engh and Huber [3]; these are employed in the restraints added by SHELXPRO to the .ins file.

The three bonds to a carbonyl carbon atom may be restrained to lie in the same plane by means of a *chiral volume restraint* [4] with a target volume of zero (e.g. 'CHIV_GLU 0 C CD'). Chiral volume restraints are also useful to prevent the inversion of α-carbon atoms and the β-carbons of Ile and Thr, e.g. 'CHIV_ILE 2.5 CA CB'. Chiral volume restraints may even be applied to non-chiral atoms such as CB of valine and CG of leucine in order to ensure conformity with conventional atom-labeling schemes (from the point of view of the atom names, these atoms could be considered to be chiral!). The FLAT instruction restrains planar groups by restraining the

(chiral) volumes of a sufficient number of atomic tetrahedra to be zero; this works well but is somewhat unconventional, and so may be changed in the future.

Although the unique features of SHELXL are primarily useful for refinement against very high resolution data, tests on SHELXL-93 indicated that only small changes would be required to extend its range of applicability to medium resolution data (say 2.8Å or better). The most important of these changes, implemented in SHELXL-96, were improved diagnostics and more sophisticated anti-bumping restraints, and the addition of non-crystallographic symmetry (NCS) restraints.

Anti-bumping restraints are distance restraints that are only applied if the two atoms are closer to each other than the target distance. They can be generated automatically by SHELXL, taking symmetry equivalent atoms and disorder into account. Since this step is relatively time-consuming, in the 1993 release it was performed only before the first refinement cycle, and the anti-bumping restraints were generated automatically only for the solvent (water) atoms (however they could be inserted by hand for any pairs of atoms). In practice this proved to be too limited, so in SHELXL-96 the automatic generation of anti-bumping restraints was extended to all C, N, O and S atoms (with an option to include H...H interactions) and was performed each refinement cycle.

Anti-bumping restraints are not generated automatically for (a) atoms connected by a chain of three bonds or less in the connectivity array (unless separated by more than a specified number of residues), (b) atoms with different non-zero PART numbers, and (c) pairs of atoms for which the sum of occupancies is less than 1.1. The target distances for the O...O and N...O distances are less than for the other atom pairs to allow for possible hydrogen bonds. The H...H anti-bumping restraints are applied to all pairs of hydrogen atoms not bonded to the same atom; they help to ensure chemically reasonable conformations for flexible side-chains.

The use of *NCS restraints* considerably improves the effective data to parameter ratio, and the resulting Fourier maps often look as though they were calculated with higher resolution data than were actually used (because the phases are more accurate). Two types of NCS restraint may be generated automatically with the help of the NCSY instruction. The first type uses the connectivity table to define equivalent 1,4-distances, which are then restrained to be equal. The second restrains the isotropic *U*-values of equivalent atoms to be equal. It is not normally necessary to restrain equivalent 1,2- and 1,3-distances to be equal because the DFIX and DANG restraints will have this effect anyway; but SAME may be used to add such restraints in the absence of DFIX and DANG. The use of *restraints* rather than applying NCS as an exact constraint (e.g. in the structure factor calculation) is more flexible (but slower) and does not require the specification of transformation matrices and real-space masks. Experience indicates that NCS restraints should be used wherever possible; it is not difficult to relax them later (e.g. for specific side-chains involved in interactions with other non-NCS related molecules) should this prove to be necessary. Note that restraining equivalent 1,4-distances to be equal is not quite as restrictive as restraining equivalent torsion angles: the 1,4-distances are equal for gauche$^+$ and gauche$^-$ conformations. On the other hand such conformational differences are chemically plausible for exposed side-chains.

Constraints and restraints greatly increase the radius and rate of convergence of crystallographic refinements, so they should be employed in the early stages of refinement wherever feasible. The difference electron density syntheses calculated after such restrained refinements are often more revealing than those from free refinements. In large small-molecule structures with poor data to parameter ratios, the last few atoms can often not be located in a difference map until an anisotropic refinement has been performed with geometrical and ADP restraints. Atoms with low displacement parameters that are well determined by the X-ray data will be relatively little affected by the restraints, but the latter may well be essential for the successful refinement of poorly defined regions of the structure. Premature removal or softening the restraints (to improve the *R*-value !) often impedes further progress.

3 Restrained Anisotropic Refinement

There is no doubt that macromolecules are better described in terms of anisotropic displacements, but the data to parameter ratio is very rarely adequate for a free anisotropic refinement. Such a refinement often results in 'non-positive definite' (NPD) displacement tensors, and at the best will give probability ellipsoids that do not conform to the expected dynamical behavior of the molecule. Clearly constraints or restraints must be applied to obtain a chemically sensible model. It is possible to divide a macromolecule up into relatively rigid domains, and to refine the 20 TLS parameters of rigid body motion for each domain [5]. This may be a good model for the bases in oligonucleotides and for the four aromatic side-chains in proteins, but otherwise macromolecules are probably not sufficiently rigid for the application of TLS constraints, or they would have to be divided up into such small units that too many parameters would be required.

As with the refinement of atomic positions, restraints offer a more flexible approach.

The *rigid-bond restraint* (DELU) [6] assumes that the components of the anisotropic displacement parameters (ADPs) along bonded (1,2-) or 1,3-directions are zero within a given esd. This restraint should be applied with a low esd, i.e. as a 'hard' restraint. For many non-planar groups of atoms, rigid-bond restraints effectively impose TLS conditions of rigid body motion [7]. Although rigid-bond restraints involving 1,2- and 1,3-distances reduce the effective number of free ADPs per atom from 6 to less than 4 for typical organic structures, further restraints are often required for the successful anisotropic refinement of macromolecules.

The *similar ADP restraint* (SIMU) restrains the corresponding U_{ij}-components to be approximately equal for atoms which are spatially close (but not necessarily bonded because they may be in different components of a disordered group). The isotropic version of this restraint has been employed frequently in protein refinements. This restraint is consistent with the characteristic patterns of thermal ellipsoids in many organic molecules; on moving out along side-chains, the ellipsoids become more extended and also change direction gradually.

Neither of these restraints are suitable for isolated solvent (water) molecules. A linear restraint (ISOR) restrains the ADP's to be *approximately isotropic*, but without specifying the magnitude of the corresponding equivalent isotropic displacement parameter. Both SIMU and ISOR restraints are clearly only approximations to the truth, and so should be applied as 'soft' restraints with high esds. When all three restraints are applied, structures may be refined anisotropically with a much smaller data to parameter ratio, and still produce chemically sensible ADP's. Even when more data are available, these restraints are invaluable for handling disordered regions of the structure.

An ensemble distribution created by molecular dynamics is an alternative to the harmonic description of anisotropic motion [8,9], and may be more appropriate for structures with severe conformational disorder that do not diffract to high resolution.

4 The Free *R*-factor

The question of whether the restraints can be removed in the final refinement, or what the best values are for the corresponding esds, can be resolved elegantly by the use of R_{free} [10]. To apply this test, the data are divided into a working set (90 to 95% of the reflections) and a reference set (the remaining 5 to 10%). The reference set is only used for the purpose of calculating a conventional *R*-factor that is called R_{free}. It is very important that the structural model is not in any way based on the reference set of reflections, so these are left out of *all* refinement and Fourier map calculations. If the original model was in any way derived from the same data, then many refinement cycles are required to eliminate memory effects. This ensures that the *R*-factor for the reference set provides an objective guide as to whether the introduction of additional parameters or the weakening of restraints has actually improved the model, and not just reduced the *R*-factor for the data employed in the refinement ('*R*-factor cosmetics'). The second parameter on the CGLS instruction specifies the ratio of the number of working set reflections to the number of reference set reflections, but for comparisons involving other programs it is better to use two separate .hkl files (SHELXPRO can be used to generate these).

R_{free} is invaluable in deciding whether a restrained anisotropic refinement is significantly better than an isotropic refinement. Experience indicates that both the resolution and the quality of the data are important factors, but that restrained anisotropic refinement is unlikely to be justified for crystals that do not diffract to better than 1.5 Å

Despite the overwhelming arguments for using R_{free} to monitor macromolecular refinements, it is only a single number, and is itself subject to statistical uncertainty because it is based on a limited number of reflections. Thus R_{free} may be insensitive to small structural changes, and small differences in R_{free} should not be taken as the last word; one should always consider whether the resulting geometrical and displacement parameters are *chemically reasonable*. The final refinement and maps should always be calculated with the *full* data, but without introducing additional parameters or changing the weights of the restraints. R_{free} is most useful for establishing refinement protocols; for a series of closely similar refinements (e.g. for mutants to similar resolution) the R_{free} tests only need to be applied to the first.

5 Disorder Made Simple

To obtain a chemically sensible refinement of a disordered group, we will probably need to constrain or restrain a sum of occupation factors to be unity, to restrain equivalent interatomic distances to be equal to each other or to standard values (or alternatively apply rigid group constraints), and to restrain the displacement parameters of overlapping atoms. In the case of a tight unimodal distribution of conformations, restrained anisotropic refinement may provide as good a description as a detailed

manual interpretation of the disorder in terms of two or more components, and is much simpler to perform. With high-resolution data it is advisable to make the atoms anisotropic *before* attempting to interpret borderline cases of side-chain disorder; it may well be found that no further interpretation is needed, and in any case the improved phases from the anisotropic refinement will enable higher quality difference maps to be examined.

Typical warning signs for disorder are large (and pronounced anisotropic) apparent thermal motion (in such cases the program may suggest that an atom should be split and estimate the coordinates for the two new atoms), residual features in the difference electron density and violations of the restraints. This information in summarized by the program on a residue by residue basis, separately for main-chain, side-chain and solvent atoms. In the case of two or more discrete conformations, it is usually necessary to model the disorder at least one atom further back than the maps indicate, in order that the restraints on the interatomic distances are fulfilled. The different conformations should be assigned different PART numbers so that the connectivity array is set up correctly by the program; this enables the correct rigid bond restraints on the anisotropic displacement parameters and idealized hydrogen atoms to be generated automatically even for disordered regions (it is advisable to model the disorder before adding the hydrogens).

Several strategies are possible for modeling disorder with SHELXL, but for macromolecules the simplest is to include all components of the disorder in the same residues and use the same atom names, the atoms belonging to different components being distinguished only by their different PART numbers. This procedure enables the standard restraints etc. to be used unchanged, because the same atom and residue names are used. No special action is needed to add the disordered hydrogen atoms, provided that the disorder is traced back one atom further than it is visible (so that the hydrogen atoms on the PART 0 atoms bonded to the disordered components are also correct). The following extract from an .ins file illustrates the action necessary:

```
RESI 38 SER
N   3 0.7714 0.9267 0.0062  11.0 0.1093
CA  1 0.7887 0.9740 0.0744  11.0 0.1370
PART  1
CB  1 0.8386 1.0427 0.0551  41.0 0.1188
OG  4 0.8994 1.0027 0.0230  41.0 0.1820
PART  2
CB  1 0.8414 1.0366 0.0653 -41.0 0.1493
OG  4 0.8368 1.1036 0.0102 -41.0 0.1732
PART  0
C   1 0.7414 1.0167 0.1038  11.0 0.0840
O   4 0.7072 1.0231 0.0690  11.0 0.1018
```

Atoms in PART 1 may bond to other atoms in PART 1 and also to those in PART 0, but not to those in PART 2 etc. Component (PART) 1 has been assigned an occupancy equal to free variable number 4, and component 2 has been assigned an occupancy equal to one minus free variable 4 (this is specified by the codes 41 and –41), so that a single occupancy parameter is refined, and the occupancies of the disordered atoms sum to unity. All other instructions are the same as for non-disordered residues. The last column contains the (isotropic) U-values ($B = 8\pi^2U$). The program works out itself how to apply the restraints, add H-atoms etc. Note that this very simple and effective treatment of disorder was not available in SHELXL-93.

6 The Solvent Model

It is relatively common practice in the refinement of macromolecular structures to insert water molecules with partial occupancies at the positions of difference electron density map peaks in order to reduce the *R*-factor (another example of ' *R*-factor cosmetics'). Usually when two different determinations of the same protein structure are compared, only the most tightly bound waters, which usually have full occupancies and smaller displacement parameters, are the same in each structure. The refinement of partial occupancy factors for the solvent atoms (in addition to their displacement parameters) is rarely justified by R_{free}, but sometimes the best R_{free} value is obtained for a model involving some water occupancies fixed at 1.0 and some at 0.5.

Regions of diffuse solvent may be modeled using *Babinet's principle* [11]; the same formula is employed in the program TNT [12], but the implementation is somewhat different. In SHELXL it is implemented as the SWAT instruction and usually produces a significant but not dramatic improvement in the agreement of the very low angle data. Anti-bumping restraints may be input by hand or generated automatically by the program, taking symmetry equivalents into account. After each refinement job, the displacement parameters of the water molecules should be examined, and waters with very high values (say U greater than 0.8 Å^2, corresponding to a B of 63) eliminated. The F_o-F_c map is then analyzed automatically to find the highest peaks that involve no bad contacts and make at least one geometrically plausible hydrogen bond to an electronegative atom. These peaks are then included with full occupancies and oxygen scattering factors in the next refinement job. This procedure is repeated several times; in general R_{free} rapidly reaches its minimum value, although the conventional *R*-index continues to fall as further waters are added. It should be noted that the

automatic generation of anti-bumping restraints is less effective when the water occupancies are allowed to have values other than 1.0 or 0.5. This approach provides an efficient way of building up a chemically reasonable (but not necessarily unique) network of waters that are prevented from diffusing into the protein, thus facilitating remodeling of disordered side-chains etc. The occupancies of specific waters may also be tied (using free variables) to the occupancies of particular components of disordered side-chains where this makes chemical sense. This procedure may be facilitated by using SHELXPRO to convert the .res output file from one refinement job to the .ins file for the next, or fully automated using the program SHELXWAT that calls SHELXL repeatedly. A related but much more sophisticated approach (ARP) described by Lamzin and Wilson [13] may also be used in conjunction with SHELXL.

7 The Radius of Convergence

A crucial aspect of any macromolecular refinement program is the radius of convergence. A larger radius of convergence reduces the amount of time-consuming manual intervention using interactive graphics. There are probably a number of contributing factors to the good convergence typically observed for SHELXL, e.g. the refinement against properly weighted F^2 values for *all* data, the inclusion of important off-diagonal terms in the least-squares algebra [4], the ability to refine all parameters at once (i.e. coordinates and displacement parameters in the same cycle), and the restriction to unimodal restraint functions; multimodal restraint functions such as torsion angles or hydrogen bonds tend to increase the number of spurious local minima. It is much better to reserve the multimodal chemical information such as torsion angles for verifying the structure with an independent program such as PROCHECK [14], and to use the unimodal information as restraints. The errors in the FFT calculation of derivatives are larger that those in the structure factors (for the same grid intervals); this would also impede convergence.

Many claims that SHELXL gives *R*-factors one or two percent lower than other programs have been tracked down either to subtle differences in the model or to not getting trapped in local minima. The differences in the model include the treatment of diffuse solvent and hydrogen atoms, and the ability to refine common occupancies for disordered groups. The inclusion of dispersion terms and the use of a conventional rather than a FFT structure factor summation are also more precise; the approximations in the FFT summation may become significant for high resolution data and atoms with small displacement parameters.

8 Estimated Standard Deviations

No small molecule crystallographer would contemplate publishing a structure without estimated standard deviations, but they are rarely quoted for macromolecules, and then usually only in the form of a Luzzati plot (which is rather inappropriate and was never intended for the purpose [15]!). Provided that there are appreciably more data than parameters, it is in fact possible to invert the full least-squares normal matrix (or at least large blocks of it) from the refinement of a macromolecule, and so derive the esds in all parameters by small-molecule methods. SHELXL uses the full covariance matrix for the estimation of the esds in all dependent parameters such as bond lengths, torsion angles etc.

The structure should be refined to convergence by conjugate gradient least-squares (CGLS) so that the matrix needs to be inverted only once, at the end of the refinement. It turns out that the inversion produces sensible esds even when the calculated shifts would lead to instability. The esds take the restraints into account (in a Bayesian sense) so all restraints should be switched off for this final full-matrix cycle, which is performed with L.S. 1 and DAMP 0 0. This DAMP instruction specifies zero damping (which would otherwise artificially reduce the esds) and zero shift multipliers. All the reflection data should of course be used.

If the full-matrix cycle would take longer than a week or require the purchase of extra memory, an adequate compromise is to use BLOC 1 N_1 > LAST (or something similar) to set up a full-matrix block consisting of all positional but no thermal displacement parameters.

SHELXPRO can be used to plot the atomic positional and bond length esds (a BOND instruction is needed for SHELXL to generate the latter) against the B or B_{eq} values. Preliminary tests suggest that a formula recently proposed by Durward Cruickshank [15] models the dependence of the esds on B_{eq}, effective atomic number, the *R*1-value and the completeness of the data rather well (much better that the Luzzati method).

9 Example of a High-Resolution Refinement

An example of a SHELXL refinement against high resolution data is summarized in Table 1. The starting coordinates were taken from the 1.5Å structure. The first batch of waters produced a big drop in *R*1 and R_{free}, but further dilution only reduced *R*1, not R_{free}. Restrained

anisotropic refinement reduced R_{free} by a total of 6.5%, which is highly significant; the addition of riding hydrogen atoms did not increase the number of parameters but reduced R_{free} by 0.9%. At this point the better phases from the anisotropic refinement made it possible to locate automatically further significant water molecules (this is invariably observed when there is a big drop in R_{free} on going anisotropic). The best value of R_{free} was obtained with a solvent model containing 79 fully occupied waters and 18 half occupied waters; finally modeling disordered conformations for six side-chains produced a further small improvement. Refinement of all water occupancies reduced $R1$ by 0.1% but increased R_{free} by 0.2%, so would not have been appropriate here.

Table 1. SHELXL refinement of rubredoxin against 0.92A synchrotron data. 24770 reflections with $F>4\sigma$ were used for $R1$, and 1350 for R_{free}.

Job	Action taken	N(parm.)	N(restr.)	$R1$%	R_{free}%
1	**PDB 7rxn, no water**	**1**	**1544**	**22.9**	**23.4**
2	**+38 waters**	**1734**	**1554**	**16.3**	**18.6**
3	**+14 more waters**	**1790**	**1563**	**15,8**	**18.7**
4	**+8 more waters**	**1822**	**1567**	**15.7**	**18.7**
5	**Fe,S anisotropic**	**1857**	**1597**	**14.8**	**17.7**
6	**+all H-atoms**	**1857**	**1600**	**14.0**	**16.8**
7	**all C,N,O anisotropic**	**4097**	**4888**	**8.8**	**11.3**
8	**+28 more waters**	**4358**	**5086**	**7.8**	**10.7**
9	**water –> 79(1)+18(½)**	**4556**	**5291**	**7.5**	**10.3**
10	**6 disordered sidechains**	**4698**	**5530**	**6.9**	**9.7**

I am very grateful to Zbigniew Dauter, Keith Wilson and Larry Sieker for providing the data for this test.

References

[1] G. M. Sheldrick and T. R. Schneider, "SHELXL: High-Resolution Refinement," Methods in Enzymology (eds. C. W. Carter and R. M. Sweet), Academic Press, 1996 (in press).

[2] G. Parkinson, J. Voitechovsky, L. Clowney, A. T. Brünger and H. M. Berman, "New Parameters for the Refinement of Nucleic Acid-Containing Structures," Acta Crystallogr., Vol. D52, pp. 57-64, 1996.

[3] R. A. Engh and R. Huber, "Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement," Acta Crystallogr., Vol. A47, pp. 392-400, 1991.

[4] W. A. Hendrickson, "Stereochemically Restrained Refinement of Macromolecular Structures," Methods in Enzymology (eds. H. W. Wyckoff, C. H. W. Hirs and S. G. Timasheff), Academic Press, Vol. 115B, pp. 252-270, 1985.

[5] H. Driessen, M. I. J. Haneef, G. W. Harris, B. Howlin, G. Khan and D. S. Moss, "RESTRAIN: Restrained Structure-Factor Least-Squares Refinement Program for Macromolecular Structures," J. Appl. Crystallogr., Vol. 22, pp. 510-516, 1989.

[6] J. S. Rollett, "Least-Squares Procedures in Crystal Structure Analysis," Computing in Crystallography (eds. F. R. Ahmed, S. R. Hall and C. P. Huber), Munksgaard, pp. 167-181, 1970.

[7] J. J. Didisheim and D. Schwarzenbach, "Rigid-Link Constraints and Rigid-Body Molecules," Acta Crystallogr., Vol. A43, pp. 226-232, 1987.

[8] P. Gros, W. F. van Gunsteren and W. G. Hol, "Inclusion of Thermal Motion in Crystallographic Structures by Restrained Molecular Dynamics," Science., Vol. 249, pp. 1149-1152, 1990.

[9] J. B. Clarage and G. N. Phillips, "Cross-Validation Tests of Time-Averaged Molecular Dynamics Refinements for Determination of Protein Structures by X-ray Crystallography," Acta Crystallogr., Vol. D50, pp. 24-36, 1994.

[10] A. T. Brünger, "Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures," Nature, Vol. 355, pp. 472-475, 1992.

[11] P. C. Moews and R. H. Kretsinger, "Refinement of the Structure of Carp Muscle Calcium-Binding Parvalbumin by Model Building and Difference Fourier Analysis," J. Mol. Biol., Vol. 91, pp. 201-228, 1975.

[12] D. E. Tronrud, L. F. Ten Eyck and B. W. Matthews, "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecules," Acta Crystallogr., Vol. A43, pp. 489-501, 1987.

[13] V. S. Lamzin and K. S. Wilson, "Automated Refinement of Protein Models," Acta Crystallogr., Vol. D49, pp. 129-147, 1993.

[14] R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, "PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures," J. Appl. Crystallogr., Vol. 26, pp. 283-291, 1993.

[15] D. W. J. Cruickshank, "Protein Precision Re-Examined: Luzzati Plots do not Estimate Final Errors," Abstract, I.U.Cr. Meeting, Seattle, 1996.