Recent developments in software for the automation of crystallographic macromolecular structure determination

Paul D Adams* and Ralf W Grosse-Kunstleve†

The automation of macromolecular structure determination by X-ray crystallography has long been a goal for many researchers. Recently, there have been improvements in the underlying algorithms, some of which have been implemented in software packages that deal with multiple stages of the structure determination process. These first steps towards complete automation have made X-ray crystallography more efficient.

Addresses
Lawrence Berkeley National Laboratory, Mailstop 4-230, 1 Cyclotron Road, Berkeley, CA 94720, USA
*e-mail: PDAdams@lbl.gov
†e-mail: RWGrosse-Kunstleve@lbl.gov

Current Opinion in Structural Biology 2000, **10**:564–568

0959-440X/00/$ – see front matter

Abbreviations

CCD	charge-coupled device
MAD	multiwavelength anomalous diffraction
NCS	noncrystallographic symmetry
PDB	Protein Data Bank
rms	root mean square

Introduction

The desire to understand biological processes at a molecular level has led to the routine application of X-ray crystallography. However, many people outside of the field of structural biology remain unaware of the time and effort usually required to solve a structure. Much of this effort is in the form of the manual interpretation of complex numerical data and the repeated use of interactive three-dimensional graphics. The need for extensive manual intervention leads to two major problems: significant bottlenecks that impede rapid structure determination [1•] and the introduction of errors due to subjective interpretation of the data [2••]. The automation of structure determination is thus desirable as it provides the opportunity to produce minimally biased models in a shorter time. Here, we present some of the recent technical advances that help automate the structure determination process in macromolecular X-ray crystallography.

High-throughput structure determination

The field of small-molecule crystallography, in which atomic-resolution data are routinely collected, is highly automated. As a result, the current growth rate of the Cambridge Structural Database (CSD) is more than 15,000 new structures per year. This is approximately 10 times the growth rate of the Protein Data Bank (PDB). Thus, automation of macromolecular X-ray crystallography has long been a goal and, with the recent development of the concept of structural genomics [1•,3], has moved to a position of prime importance. In order to exploit the information present in the rapidly expanding sequence databases, it has been proposed that the structural database must also grow. Increased knowledge about the relationship between sequence, structure and function will allow sequence information to be used to its full extent. For structural genomics to be successful, macromolecular structures will need to be solved at a significantly faster rate than at present. This high-throughput structure determination will require automation to reduce the bottlenecks related to human intervention. Automation will rely on the development of algorithms that minimize or eliminate subjective input, the development of algorithms that automate procedures that were traditionally performed by hand and, finally, the development of software packages that allow tight integration of these algorithms. Truly automated structure determination will require the computer to make decisions about how best to proceed in the light of the available data.

The automation of macromolecular structure determination applies to all of the procedures involved. There have been many technological advances that make macromolecular X-ray crystallography easier. In particular, the use of cryoprotection to extend crystal life [4], the availability of tunable synchrotron sources [5] and high-speed CCD data collection devices [6•], and the ability to incorporate anomalously scattering selenium atoms into proteins have all made structure determination much more efficient [6•]. The desire to make structure determination more efficient has led to investigations into the optimal data collection strategies for multiwavelength anomalous diffraction (MAD) [7•] and phasing using single anomalous diffraction with sulfur or ions [8••,9]. It has been shown that, in general, a single wavelength collected at the anomalous peak is sufficient to solve a macromolecular structure [10]. Such an approach minimizes the amount of data that need to be collected and increases the efficiency of synchrotron beamlines, and is therefore likely to become an important and widely used technique in the future. It is clear that automation will need to be applied to all aspects of structure determination, including crystal growth, crystal mounting, data collection and data processing. In this review, however, we will focus on the methods required to solve the structure once the data have been collected and processed.

Data analysis

The first step of structure determination, once the raw images have been processed, is assessment of data quality. The intrinsic quality of the data must be quantified and the appropriate signal extracted. Observations that are in error must be rejected as outliers. Some observations will be rejected at the data-processing stage, where multiple observations are available. If redundancy is low, however,

then probabilistic methods can be used [11•]. The prior expectation of the set of observations, given either by a Wilson distribution of intensities or by model-based structure factor probability distributions, is used to detect outliers. This method is able to reject strong observations that are in error, which tend to dominate the features of electron density and Patterson maps. This method could also be extended to the rejection of outliers during the model refinement process.

When using isomorphous substitution or anomalous diffraction methods for experimental phasing, the relevant information lies in the differences between the different sets of observations. In the case of anomalous diffraction, these differences are often very small, being of the same order as the noise in the data. In general, the anomalous differences at the peak wavelength are sufficient to locate the heavy atoms [12••]. In less routine cases, however, it can be very important to extract maximum information from the data. One approach used in MAD phasing is to analyze the datasets to calculate F_A structure factors, which correspond to the anomalously scattering substructure [13]. In another approach, a specialized procedure for the normalization of structure factor differences arising from either isomorphous or anomalous differences has been developed in order to facilitate the use of direct methods for heavy-atom location [14•].

Heavy-atom location and computation of experimental phases

The location of heavy atoms in isomorphous replacement or the location of anomalous scatterers was traditionally performed by manual inspection of Patterson maps. In recent years, however, labeling techniques, such as seleno-methionyl incorporation, have become widely used. This leads to an increase in the number of atoms that need to be located, rendering manual interpretation of Patterson maps extremely difficult. As a result, automated heavy-atom location methods have proliferated. The programs SOLVE [15••] and CNS [12••,16] use Patterson-based techniques to find a starting heavy-atom configuration that is then completed using difference Fourier analyses. Both Shake-and-Bake (SnB) [17•] and SHELX [18] use direct methods for reciprocal-space phase refinement, combined with modifications in real space. Shake-and-Bake refines phases derived from randomly positioned atoms, whereas SHELX derives starting phases by automatic inspection of the Patterson map. All these methods have been used with great success to find more than 30 selenium sites. Shake-and-Bake has been used to find up to 70 selenium sites [19].

After the heavy atom or anomalously scattering substructure has been located, experimental phases can be calculated and the parameters of the substructure refined. A number of modern maximum-likelihood-based methods for heavy-atom refinement and phasing are readily available (MLPHARE [20], CNS [16], SHARP [21], SOLVE [15••]). The SOLVE program has the advantage of fully integrating and automating heavy-atom location, refinement and phasing, and is therefore very easy to use.

Density modification

There are many real-space constraints, such as solvent flatness, that can be applied to electron density maps in an iterative fashion to improve initial phase estimates. This process of density modification is now routinely used to improve experimental phases before map interpretation and model building. However, as a result of the cyclic nature of the density modification process, whereby the original phases are combined with new phase estimates, introduction of bias is a serious problem. The γ correction was developed to reduce the bias inherent in the process and has been applied successfully in the method of solvent flipping [22]. The γ correction has been generalized to the γ perturbation method, which can be applied to any arbitrary density modification procedure, including noncrystallographic symmetry (NCS) averaging and histogram matching [23••]. More recently, a reciprocal-space maximum-likelihood formulation of the density modification process has been devised [24••]. This method has the advantage that a likelihood function can be directly optimized with respect to the available parameters (phases and amplitudes), rather than indirectly optimized through a weighted combination of starting parameters with those derived from flattened maps. In this way, the problem of choosing weights for phase combination is avoided.

Molecular replacement

The method of molecular replacement is commonly used to solve structures for which a homologous structure is already known. In order to make the problem tractable, it has traditionally been broken down into two three-dimensional search problems: a search to determine the rotation parameters of the model, followed by a search to determine the translation parameters of the rotated model. The method of Patterson correlation (PC) refinement is often used to optimize the rotational parameters before the translation search, thus increasing the likelihood of finding the correct solution [25]. With currently available programs, structure determination by molecular replacement usually involves significant manual input. Recently, however, methods have been developed to automate molecular replacement. One approach uses the exhaustive application of traditional rotation and translation methods to perform a complete six-dimensional search [26]. Less time-consuming methods have been developed using new algorithms based on evolutionary search [27•] or stochastic procedures [28•], both of which are able to perform directed six-dimensional searches in a relatively short time. In the future, similar methods may permit experimental data to be exhaustively tested against all known structures to determine whether a homologous structure, which could then be used to aid structure determination, is already present in a database.

Map interpretation

The first stage of electron density map interpretation is an overall assessment of the information contained in a given map. The standard deviation of the local rms electron density can be calculated from the map. This variation is high when the electron density map has well-defined protein and solvent regions, and is low for maps calculated with random phases [29,30]. It has also been shown that the correlation of the local rms density in adjacent regions in the unit cell can be used as a measure of the presence of distinct, contiguous solvent and macromolecular regions in an electron density map [31•].

Currently, the process of analyzing an experimental electron density map to build the atomic model is a time-consuming, subjective process and is almost entirely graphics based. It has been shown that there are substantial differences in the models built by different people presented with the same experimental data [2••]. The majority of time spent completing a crystal structure is in the use of interactive graphics to manually modify the model. This manual modification is required either to correct parts of the model that are incorrectly placed or to add parts of the model that are currently missing. This process is prone to human error because of the large number of degrees of freedom of the model and the possible poor quality of regions of the electron density map.

Fortunately, much of the subjectivity of manual rebuilding has been removed by incorporating information from databases of known structures [32]. However, there have been significant advances in making the process of map interpretation and model building truly automated. One route to automated analysis of the electron density map is the recognition of larger structural elements, such as α helices and β strands. The location of these features can often be achieved, even in electron density maps of low quality, using exhaustive searches in either real space [33] or reciprocal space [34], the latter having a significant advantage in speed. The automatic location of secondary structure elements can be combined with sequence information and databases of known structures to build an initial atomic model with little or no manual intervention from the user [35]. This method has been seen to work even at relatively low resolution (d_{min} ~3.0 Å). However, its implementation is still graphics based and requires user input. In order to completely automate the model building process, a method has been developed that combines automated identification of potential atomic sites in the map [36] with model refinement [37]. The atomic sites are then analyzed to determine both a mainchain protein trace and the identity of amino acid residues. From this information and knowledge of the protein sequence, a model can be automatically constructed [38••]. This powerful procedure, known as warpNtrace, can gradually build a more complete model from the initial electron density map and, in many cases, is capable of building the majority of the protein structure in a completely automated way. Unfortunately, this method currently has the limitation of a need for relatively high resolution data (d_{min} < 2.0 Å). Data that extend to this resolution are available for only about 50% of the approximately 10,000 X-ray structures in the PDB. To extend the applicability of automated map interpretation to lower resolution data, work has started using pattern recognition methods [39••]. The resulting program is called TEXTAL and shows great promise for the interpretation of maps even at a data resolution as low as 3.0 Å. Data of this quality are available for approximately 95% of the structures in the PDB.

Refinement and validation

In general, the atomic model obtained by automatic or manual methods contains some errors and must be optimized to best fit the experimental data and previously known chemical information. In addition, the initial model is often incomplete and refinement is carried out to generate improved phases that can then be used to compute a more accurate electron density map. Most recently, improved targets for the refinement of incomplete, error-containing models have been obtained using the more general maximum-likelihood formulation [37,40]. The resulting maximum-likelihood refinement targets have been successfully combined with the powerful optimization method of simulated annealing to provide a very robust and efficient refinement scheme [41]. For many structures, some initial experimental phase information is available from either isomorphous heavy-atom replacement or anomalous diffraction methods. These phases represent additional observations that can be incorporated in the refinement target. Tests have shown that the addition of experimental phase information greatly improves the results of refinement [40,41].

The refinement methods used in macromolecular structure determination work almost exclusively in reciprocal space. However, there has been renewed interest in the use of real-space refinement algorithms that can take advantage of high-quality experimental phases from anomalous diffraction experiments or NCS averaging. Tests have shown that the method can be successfully combined with the technique of simulated annealing [42].

The parameterization of the atomic model in refinement is of great importance. When the resolution of the experimental data is limited, then it is appropriate to use chemical constraints on bond lengths and angles. This torsion angle representation is seen to decrease overfitting and improve the radius of convergence of refinement [43]. If data are available to high enough resolution, additional atomic displacement parameters can be used. Macromolecular structures often show anisotropic motion, which can be resolved at a broad spectrum of levels ranging from whole domains down to individual atoms. The use of the fast Fourier transform has greatly improved the speed with which such models can be generated and tested [44•].

Validation of macromolecular models and their experimental data [45] is an essential part of structure determination [46]. This is important both during the structure determination process and at the time of coordinate and data deposition at the PDB, when extensive validation criteria are also applied [47]. In the future, the repeated application of validation criteria in automated structure determination will help avoid errors that currently occur as a result of the subjective manual interpretation of data and models.

Noncrystallographic symmetry

It is not uncommon for macromolecules to crystallize with more than one copy in the asymmetric unit. This leads to relationships between atoms in real space and reflections in reciprocal space. These relationships can be exploited in the structure determination process. However, the identification of NCS is generally a manual process. A method for the automatic location of proper NCS (i.e. a rotation axis) has been shown to be successful even at low resolution [48]. A more general approach to finding NCS relationships uses skeletonization of electron density maps [49•]. It is possible that these methods might be used in the future to automate the location of NCS operators and the determination of molecular masks.

Conclusions

Although there are many details that still need to be resolved in order to generate a truly automated procedure for structure determination, there have been many advances. Programs such as SOLVE [15••] and the warpNtrace suite [38••] combine large functional blocks in an automated fashion. The program CNS [16] provides a framework in which different algorithms can be combined and tested using a powerful scripting language. However, building a fully automated system covering all aspects of structure determination and refinement will depend on the continued development of algorithms that are more efficient, more robust and more objective. Adequate support for method development is therefore a prerequisite to realizing the goal of automation, which will be of crucial importance in extending our understanding of biology through structural genomics.

Acknowledgements

We thank A Brunger for critical reading of the manuscript. This work was funded by a grant from the Department of Energy and Lawrence Berkeley National Laboratory (DE-AC03-76SF00098) to PDA.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Sali A, Studier FW, Swaminathan S: **Structural genomics: beyond the human genome project.** *Nat Genet* 1999, **23**:151-157.
The idea of structural genomics is clearly presented. The complexity of the problem and the methodological advances required for success are described.

2. •• Mowbray SL, Helgstrand C, Sigrell JA, Cameron AD, Jones TA: **Errors and reproducibility in electron-density map interpretation.** *Acta Crystallogr D* 1999, **55**:1309-1319.
The same data were given to three researchers, who proceeded to build and refine an atomic model. Significant differences are seen in the resulting models. Some of the criteria and tools that were useful in detecting errors are discussed.

3. Montelione GT, Anderson S: **Structural genomics: keystone for a Human Proteome Project.** *Nat Struct Biol* 1999, **6**:11-12.

4. Garman E: **Cool data: quantity AND quality.** *Acta Crystallogr D* 1999, **55**:1641-1653.

5. Walsh MA, Evans G, Sanishvili R, Dementieva I, Joachimiak A: **MAD data collection – current trends.** *Acta Crystallogr D* 1999, **55**:1726-1732.

6. • Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A: **Taking MAD to the extreme: ultrafast protein structure determination.** *Acta Crystallogr D* 1999, **55**:1168-1173.
This paper demonstrates that it is possible to solve structures using the multiwavelength anomalous diffraction (MAD) method in a very short time with a third-generation synchrotron source and semiautomated software.

7. • Gonzalez A, Pedelacq J-D, Sola M, Gomis-Ruth FX, Coll M, Samama J-P, Benini S: **Two-wavelength MAD phasing: in search of the optimal choice of wavelengths.** *Acta Crystallogr D* 1999, **55**:1449-1458.
The authors show that multiwavelength anomalous diffraction (MAD) phasing using only two wavelengths can be successful. The optimum wavelengths for such an experiment are those that give a large contrast in the real part of the anomalous scattering factor (e.g. the inflection point and high-energy remote).

8. •• Dauter Z, Dauter M, de La Fortelle E, Bricogne G, Sheldrick GM: **Can anomalous signal of sulfur become a tool for solving protein crystal structures?** *J Mol Biol* 1999, **289**:83-92.
If accurate diffraction data are collected at a wavelength close to that of CuKα radiation, the anomalous signal present from sulfur atoms can be used to generate useful phase information. In this example, the location of the sulfur atoms was already known, however, at longer wavelengths (approaching 2 Å) there should be sufficient signal for determination of unknown sulfur positions.

9. Dauter Z, Dauter M: **Anomalous signal of solvent bromides used for phasing of lysozyme.** *J Mol Biol* 1999, **289**:93-101.

10. Rice LM, Earnest TN, Brunger AT: **Single wavelength anomalous diffraction phasing revisited: a general phasing method?** *Acta Crystallogr* 2000, in press.

11. • Read RJ: **Detecting outliers in non-redundant diffraction data.** *Acta Crystallogr D* 1999, **55**:1759-1764.
This paper describes the use of structure factor probability distributions to identify outliers in experimental data. This outlier removal has a significant impact on Patterson and electron density maps.

12. •• Grosse-Kunstleve RW, Brunger AT: **A highly automated heavy-atom search procedure for macromolecular structures.** *Acta Crystallogr D* 1999, **55**:1568-1577.
This paper describes the combination of direct and reciprocal space Patterson search methods for the location of heavy atoms. The fast Fourier transform translation search is used to make the procedure computationally efficient.

13. Terwilliger TC: **MAD phasing: Bayesian estimates of F_A.** *Acta Crystallogr D* 1994, **50**:11-16.

14. • Blessing RH, Smith GD: **Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations.** *J Appl Crystallogr* 1999, **32**:664-670.
The theory and practical computation of normalized difference structure factors is described in some detail. It is also shown that these normalized differences can successfully be used to locate large heavy-atom substructures in macromolecules.

15. •• Terwilliger TC, Berendzen J: **Automated MAD and MIR structure solution.** *Acta Crystallogr D* 1999, **55**:849-861.
This important paper describes in detail the SOLVE program and the underlying algorithms that are used to automate heavy-atom location.

16. Brunger AT, Adams PD, Clore GM, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ *et al.*: **Crystallography and NMR system (CNS): a new software system for macromolecular structure determination.** *Acta Crystallogr D* 1998, **54**:905-921.

17. • Howell PL, Blessing RH, Smith GD, Weeks CM: **Optimizing DREAR and SnB parameters for determining Se-atom substructures.** *Acta Crystallogr D* 2000, **56**:604-617.
The optimal parameters for difference structure factor normalization and heavy-atom location with the program Shake-and-Bake are described.

18. Sheldrick GM, Gould RO: **Structure solution by iterative peaklist optimization and tangent expansion in space group P1.** *Acta Crystallogr* 1995, **B51**:423-431.

19. Deacon AM, Ealick SE: **Selenium-based MAD phasing: setting the sites on larger structures.** *Structure* 1999, **7**:161-166.

20. Otwinowski Z: **Maximum likelihood refinement of heavy atom parameters.** In *Isomorphous Replacement and Anomalous Scattering; Proceedings of the Daresbury Study Weekend.* Edited by Wolf W, Evans PR, Leslie AGW. Warrington: SERC Daresbury Laboratory; 1991:80-85.

21. de La Fortelle E, Bricogne G: **Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods.** *Methods Enzymol* 1997, **276**:472-494.

22. Abrahams JP: **Bias reduction in phase refinement by modified interference functions: introducing the gamma correction.** *Acta Crystallogr D* 1997, **53**:371-376.

23. •• Cowtan K: **Error estimation and bias correction in phase-improvement calculations.** *Acta Crystallogr D* 1999, **55**:1555-1567.
The γ correction was extended to arbitrary density modification techniques using a perturbation method. It was shown that, after bias removal, histogram matching is significantly more powerful than solvent flattening for comparable volumes of protein and solvent.

24. •• Terwilliger TC: **Reciprocal-space solvent flattening.** *Acta Crystallogr D* 1999, **55**:1863-1871.
A procedure is described for reciprocal-space maximization of a likelihood function based on experimental phases and characteristics of the electron density map. One of the significant improvements over real-space methods is that the weighting of previously known phase information relative to the information from the modified map is automatically dealt with.

25. Brunger AT: **Patterson correlation searches and refinement.** *Methods Enzymol* 1997, **276**:558-580.

26. Sheriff S, Klei HE, Davis ME: **Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement problems.** *J Appl Crystallogr* 1999, **32**:98-101.

27. • Kissinger CR, Gehlhaar DK, Fogel DB: **Rapid automated molecular replacement by evolutionary search.** *Acta Crystallogr D* 1999, **55**:484-491.
The EPMR program implements an evolutionary algorithm to perform a very efficient six-dimensional search. Results are given for a number of difficult molecular replacement problems.

28. • Glykos NM, Kokkinidis M: **A stochastic approach to molecular replacement.** *Acta Crystallogr D* 2000, **56**:169-174.
A Monte Carlo simulated annealing scheme was used to locate the positions of molecules in the asymmetric unit, starting from random positions. Random moves in the position of a molecule are accepted with a probability related to the change in a measure-of-fit, such as the R factor or linear correlation coefficient.

29. Terwilliger TC, Berendzen J: **Discrimination of solvent from protein regions in native Fouriers as a means of evaluating heavy-atom solutions in the MIR and MAD methods.** *Acta Crystallogr D* 1999, **55**:501-505.

30. Terwilliger TC: **σ^2_R, a reciprocal-space measure of the quality of macromolecular electron-density maps.** *Acta Crystallogr D* 1999, **55**:1174-1178.

31. • Terwilliger TC, Berendzen J: **Evaluation of macromolecular electron-density map quality using the correlation of local rms density.** *Acta Crystallogr D* 1999, **55**:1872-1877.
The local correlation of rms density is shown to be a good measure of how contiguous the solvent regions of the map are. This can be used to automatically identify maps that are representative of a macromolecular structure.

32. Kleywegt GJ, Jones TA: **Databases in protein crystallography.** *Acta Crystallogr D* 1998, **54**:1119-1131.

33. Kleywegt GJ, Jones TA: **Template convolution to enhance or detect structural features in macromolecular electron-density maps.** *Acta Crystallogr D* 1997, **53**:179-185.

34. Cowtan K: **Modified phased translation functions and their application to molecular-fragment location.** *Acta Crystallogr D* 1998, **54**:750-756.

35. Oldfield T: **A semi-automated map fitting procedure.** In *Crystallographic Computing 7: Macromolecular Crystallographic Data (Crystallographic Computing).* Edited by Bourne PE, Watenpaugh K. Oxford: Oxford University Press; 2000:in press.

36. Perrakis A, Sixma TK, Wilson KS, Lamzin VS. **wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models.** *Acta Crystallogr D* 1997, **53**:448-455.

37. Murshudov GN, Vagin AA, Dodson EJ: **Refinement of macromolecular structures by the maximum-likelihood method.** *Acta Crystallogr D* 1997, **53**:240-255.

38. •• Perrakis A, Morris R, Lamzin VS: **Automated protein model building combined with iterative structure refinement.** *Nat Struct Biol* 1999, **6**:458-463.
An automated method for building and refining a protein model is described. An iterative procedure is used that describes the electron density map as a set of unconnected atoms from which protein-like patterns are extracted. The method is shown to be successful with good quality data having an upper resolution limit better than 2.3 Å.

39. •• Holton T, Ioerger TR, Christopher JA, Sacchettini JC: **Determining protein structure from electron-density maps using pattern matching.** *Acta Crystallogr D* 2000, **56**:722-734.
A new pattern matching method is described for interpreting electron density maps. Small regions of the map are matched against a database of patterns derived from known structures and their maps. This method shows great promise for the automated interpretation of maps even at data resolution as low as 3.0 Å.

40. Pannu NS, Murshudov GM, Dodson EJ, Read RJ: **Incorporation of prior phase information strengthens maximum-likelihood structure refinement.** *Acta Crystallogr D* 1998, **54**:1285-1294.

41. Adams PD, Pannu NS, Read RJ, Brunger AT: **Extending the limits of molecular replacement through combined simulated annealing and maximum likelihood refinement.** *Acta Crystallogr D* 1999, **55**:181-190.

42. Chen Z, Blanc E, Chapman MS: **Real-space molecular-dynamics structure refinement.** *Acta Crystallogr D* 1999, **55**:464-468.

43. Rice LM, Brunger AT: **Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement.** *Proteins* 1994, **19**:277-290.

44. • Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ: **Efficient anisotropic refinement of macromolecular structures using FFT.** *Acta Crystallogr D* 1999, **55**:247-255.
The implementation of individual anisotropic temperature factor refinement using the fast Fourier transform method in the program REFMAC is described. The method is shown to improve the crystallographic R-value and R_{free}, as well as the fit to geometric targets for data with resolution higher than 2 Å.

45. Vaguine AA, Richelle J, Wodak SJ: **SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model.** *Acta Crystallogr D* 1999, **55**:191-205.

46. Kleywegt GJ: **Validation of protein crystal structures.** *Acta Crystallogr D* 2000, **56**:249-265.

47. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235-242.

48. Vonrhein C, Schulz GE: **Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX.** *Acta Crystallogr D* 1999, **55**:225-229.

49. • Spraggon G: **Envelope skeletonization as a means to determine monomer masks and non-crystallographic symmetry relationships: application in the solution of the structure of fibrinogen fragment D.** *Acta Crystallogr D* 1999, **55**:458-463.
An algorithm is described that utilizes skeletonization of the solvent mask generated by the solvent-flattening technique to calculate a monomer molecular envelope. It is shown that the noncrystallographic symmetry relationships between monomer envelopes can be determined using standard molecular replacement methods.