
COMPUTATIONAL ASPECTS OF HIGH-THROUGHPUT CRYSTALLOGRAPHIC MACROMOLECULAR STRUCTURE DETERMINATION

Paul D. Adams, Ralf W. Grosse-Kunstleve, and Axel T. Brunger

The desire to understand biological processes at a molecular level has led to the routine application of X-ray crystallography. However, significant time and effort usually are required to solve and complete a macromolecular crystal structure. Much of this effort is in the form of manual interpretation of complex numerical data using a diverse array of software packages, 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 solution (Burley et al., 1999), and the introduction of errors due to subjective interpretation of the data (Mowbray et al., 1999). These problems present a major impediment to the success of structural genomics efforts (Burley et al., 1999; Montelione and Anderson, 1999) that require the whole process of structure solution to be as streamlined as possible. See Chapter 29 for a detailed description of structural genomics. The automation of structure solution is thus necessary as it has the opportunity to produce minimally biased models in a short time. Recent technical advances are fundamental to achieving this automation and make high-throughput structure determination an obtainable goal.

HIGH-THROUGHPUT STRUCTURE DETERMINATION

Automation in macromolecular X-ray crystallography has been a goal for many researchers. The field of small-molecule crystallography, where atomic resolution data are routinely collected, is already highly automated. As a result, the current growth rate of the Cambridge Structural Database (CCSD) (Allen, Kennard, and Taylor, 1983) is more than 15,000 new structures per year. This growth rate is approximately 10 times the growth rate of the Protein Data Bank (PDB) (Berman et al., 2000). See Chapters 9, 10, and 11 for further details of structural databases. Automation of macromolecular crystallography could significantly improve the rate at which new structures are determined. Recently, the goal of automation has moved to a position of prime importance with the development of the concept of structural genomics (Burley et al., 1999; Montelione and Anderson, 1999). In order to exploit the information present in the rapidly expanding sequence databases, 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 rate significantly faster 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 a tight integration between these algorithms. Truly automated structure solution will require the computer to make decisions about how best to proceed in the light of the available data.

The automation of macromolecular structure solution applies to all of the procedures involved. There have been many technological advances that make macromolecular X-ray crystallography easier. In particular, cryoprotection to extend crystal life (Garman, 1999), the availability of tunable synchrotron sources (Walsh et al., 1999a), high-speed charge-coupled device (CCD) data collection devices (Walsh et al., 1999b), and the ability to incorporate anomalously scattering selenium atoms into proteins have all made structure solution much more efficient (Walsh et al., 1999b). The desire to make structure solution more efficient has led to investigations into the optimal data collection strategies for multiwavelength anomalous diffraction (Gonzalez et al., 1999) and phasing using single anomalous diffraction with sulfur or ions (Dauter et al., 1999; Dauter and Dauter, 1999). Gonzalez and her colleagues have shown that multiwavelength anomalous diffraction (MAD) phasing using only two wavelengths can be successful (Gonzalez et al., 1999). 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). However, Rice and his colleagues have also shown that, in general, a single wavelength collected at the anomalous peak is sufficient to solve a macromolecular structure (Rice, Earnest, and Brunger, 2000). Such an approach minimizes the amount of data 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.

DATA ANALYSIS

The first step of structure solution, 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. However, if redundancy is low, then probabilistic methods can be used (Read, 1999). The prior expectation, given either by a Wilson distribution of intensities or 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 multiple 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, provided that a large enough anomalous signal is observed (Grosse-Kunstleve and Brunger, 1999). However, in less routine cases it can be very important to extract the maximum information from the data. One approach used in MAD phasing is to analyze the data sets to calculate F_A structure factors, which correspond to the anomalously scattering substructure (Terwilliger, 1994). Several programs are available to estimate the F_A structure factors: XPREP (Bruker, 2001), MADSYS (Hendrickson, 1991) and SOLVE (Terwilliger and Berendzen, 1999a). 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 (Blessing and Smith, 1999).

Merohedral twinning of the diffraction data can make structure solution difficult and in some cases impossible. The twinning occurs when a crystal contains multiple diffracting domains that are related by a simple transformation such as a twofold rotation about a crystallographic axis, a phenomenon that can only occur in certain space groups. As a result the observed diffraction intensities are the sum of the intensities from the two distinctly oriented domains. Fortunately, the presence of twinning can be detected at an early stage by the statistical analysis of structure factor distributions (Yeates, 1997). If the twinning is only partial, it is possible to detwin the data. Perfect twinning typically makes structure solution using experimental phasing methods difficult, but the molecular replacement method (see below) still can be successfully used.

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. However, in recent years labeling techniques such as seleno-methionyl incorporation have become widely used. Such labeling techniques lead to an increase in the number of atoms to be located, rendering manual interpretation of Patterson maps extremely difficult. As a result, automated heavy atom location methods have proliferated. The programs SOLVE (Terwilliger and Berendzen, 1999a) and CNS (Brunger et al., 1998; Grosse-Kunstleve and Brunger, 1999) use Patterson-based techniques to find a starting heavy atom configuration that is then completed using difference Fourier analyses. Both Shake-and-Bake (SnB) (Weeks and Miller, 1999) and SHELX-D (Sheldrick and Gould, 1995) use the direct methods reciprocal-space phase refinement combined with

modifications in real-space. SnB refines phases derived from randomly positioned atoms, while SHELX-D derives starting phases by automatic inspection of the Patterson map. All methods have been used with great success to solve substructures with more than 60 selenium sites. SHELX-D and SnB have been used to find up to 150 and 160 selenium sites, respectively.

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 (Otwinowski, 1991), CNS (Brunger et al., 1998), SHARP (La Fortelle and Bricogne, 1997), SOLVE (Terwilliger and Berendzen, 1999a). The SOLVE program has the advantage of fully integrating and automating heavy atom location, refinement, and phasing, and therefore is very easy to use. The SHARP program implements a more complex algorithm for phasing, making use of two-dimensional integration over both phases and amplitudes. This method is computationally expensive, rendering SHARP typically an order of magnitude slower than other phasing programs, but in the case of significant nonisomorphism between heavy atom derivative data sets the improvement in the phases can be worth the additional computing time.

DENSITY MODIFICATION

Often the raw phases obtained from the experiment are not of sufficient quality to proceed with structure determination. However, 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 prior to map interpretation and model building. However, due to the cyclic nature of the density modification process, where 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 (Abrahams, 1997). The γ correction has been generalized to the γ perturbation method in the DM program, part of the CCP4 suite (Collaborative Computational Project 4, 1994), and can be applied to any arbitrary density modification procedure, including noncrystallographic symmetry averaging and histogram matching (Cowtan, 1999). After bias removal, histogram matching is significantly more powerful than solvent flattening for comparable volumes of protein and solvent (Cowtan, 1999). More recently a reciprocal-space, maximum-likelihood formulation of the density modification process has been devised and implemented in the program RESOLVE (Terwilliger, 2000). 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 through a weighted combination of starting parameters with those derived from flattened maps. In this way the problem of choices of weights for phase combination is avoided. The SOLVE and RESOLVE programs together provide a relatively automated way to go from experimental data to a map suitable for model building.

MOLECULAR REPLACEMENT

The method of molecular replacement is commonly used to solve structures for which a homologous structure is already known. As the database of known structures expands

as a result of structural genomics efforts, this technique will become more and more important. The method attempts to locate a molecule or fragments of a molecule, whose structure is known, in the unit cell of an unknown structure for which experimental data are available. In order to make the problem tractable, it has traditionally been broken down into two consecutive three-dimensional search problems: a search to determine the rotational orientation of the model followed by a search to determine the translational orientation for the rotated model (Rossmann and Blow, 1962). The method of Patterson Correlation (PC) refinement is often used to optimize the rotational orientation prior to the translation search, thus increasing the likelihood of finding the correct solution (Brunger, 1997). With currently available programs structure solution by molecular replacement usually involves significant manual input. Recently, however, methods have been developed to automate molecular replacement. One approach has used the exhaustive application of traditional rotation and translation methods to perform a complete six-dimensional search (Sheriff, Klei, and Davis, 1999). More recently, less time-consuming methods have been developed. The EPMR program implements an evolutionary algorithm to perform a very efficient six-dimensional search (Kissinger, Gehlhaar, and Fogel, 1999). A Monte-Carlo simulated annealing scheme is used in the program Queen of Spades to locate the positions of molecules in the asymmetric unit (Glykos and Kokkinidis, 2000).

To improve the sensitivity of any molecular replacement search algorithm, maximum likelihood methods have been developed (Read, 2001). The traditional scoring function of the search is replaced by a function that takes into account the errors in the model and the uncertainties at each stage. This approach is seen to greatly improve the chances of finding a correct solution using the traditional approach of rotation and translation searches. In addition, the method performs a statistically correct treatment of simultaneous information from multiple search models using multivariate statistical analysis (Read, 2001). This method will allow information from different structures to be used in highly automated procedures while minimizing the risk of introducing bias. In the future molecular replacement algorithms may permit experimental data to be exhaustively tested against all known structures to determine whether a homologous structure is already present in a database, which could then be used as an aid in structure determination.

MAP INTERPRETATION

The interpretation of the initial electron density map, calculated using either experimental phasing or molecular replacement methods, is often performed in multiple stages (described below) with the final goal being the construction of an atomic model. If the interpretation cannot proceed to an atomic model, that is often an indication that the data collection must be repeated with improved crystals. Alternatively, repeating previous computational steps in data analysis or phasing may generate revised hypotheses about the crystal, such as a different space group symmetry or estimate of unit cell contents. Clearly, completely automating the process of structure solution will require that these eventualities are taken into consideration and dealt with in a rigorous manner.

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 root-mean-square 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 (Terwilliger and Berendzen, 1999b; Terwilliger, 1999). Terwilliger and Berendzen also have shown that the correlation of the local root-mean-square 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 (Terwilliger and Berendzen, 1999c).

Currently the process of analyzing an experimental electron density map to build the atomic model is a time-consuming, subjective process and almost entirely graphics based. Sophisticated programs such as O (Jones et al., 1991), XtalView (McRee, 1999), QUANTA (Oldfield, 2000), TurboFrodo (Jones, 1978), and MAIN (Turk, 2000) are commonly used for manual rebuilding. These greatly reduce the effort required to rebuild models by providing: libraries of side chain rotamers and peptide fragments (Kleywegt and Jones, 1998) and map interpretation tools and real space refinement of rebuilt fragments (Jones et al., 1991). However, Mowbray and her colleagues have shown that there are substantial differences in the models built manually by different people when presented with the same experimental data (Mowbray et al., 1999). The majority of time spent in 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.

Although interactive graphics systems for manual model building have made the process dramatically simpler, there have also 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. Location of these features can often be achieved even in electron density maps of low quality using exhaustive searches in either real space (Kleywegt and Jones, 1997) or reciprocal space (Cowtan, 1998; Cowtan, 2001), the latter having a significant advantage in speed because the translation search for each orientation can be calculated using a Fast Fourier Transform. The automatic location of secondary structure elements from skeletonized electron density maps 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 (Oldfield, 2000). This method has been seen to work even at relatively low resolution ($d_{\min} \sim 3.0\text{\AA}$). However, the implementation is still graphics based and requires user input. A related approach in the program MAID also uses a skeleton generated from the electron density map as the start point for locating secondary structure elements (Levitt, 2001). Trial points are extended in space by searching for connected electron density at C_{α} distance (approximately 3.7\AA) with standard α -helical or β -strand geometry. Real-space refinement of the fragments generated is used to improve the model. Both of these methods suffer from the limitation that they do not combine the model-building process with the generation of improved electron density maps derived from the starting phases and the partial models.

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 (Perrakis et al., 1997) with model refinement (Murshudov, Vagin, and Dodson, 1997). An iterative procedure is used that describes the electron density map as a set of unconnected atoms from which proteinlike patterns, primarily the main-chain trace from peptide units, are extracted. From this information and knowledge of the

protein sequence, a model can be automatically constructed (Perrakis, Morris, and Lamzin, 1999). This powerful procedure, known as warpNtrace in ARP/wARP, 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 ($d_{\min} < 2.0\text{\AA}$). Data that extend to this resolution are available for less than 50% of the $\sim 16,500$ 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 (Holton et al., 2000). 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\AA . Data of this quality are available for approximately 95% of the structures in the PDB. We anticipate that the combination of secondary structure fragment location, the pattern matching methods of the TEXTAL program, and iteration with structure refinement for map improvement will in the future provide a general solution to the problem of model building at resolutions better than 3.5\AA .

REFINEMENT

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 prior 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. However, the refinement of macromolecular structures is often difficult for several reasons. First, the data-to-parameter ratio is low, creating the danger of overfitting the diffraction data. This method results in a good agreement of the model to the experimental data even when it contains significant errors. Therefore, the apparent ratio of data to parameters is often increased by incorporation of chemical information, that is, bond length and bond angle restraints obtained from ideal values seen in high-resolution structures (Hendrickson, 1985). Second, the initial model often has significant errors, often due to the limited quality of the experimental data, or a low level of homology between the search model and the true structure in molecular replacement. Third, local (false) minima exist in the target function. The more local minima and the deeper they are, the more likely refinement will fail. Fourth, model bias in the electron density maps complicates the process of manual rebuilding between cycles of automated refinement.

Methods have been devised to address these difficulties. Cross validation, in the form of the free R -value, can be used to detect overfitting (Brunger, 1992). The radius of convergence of refinement can be increased by the use of stochastic optimization methods such as molecular dynamics-based simulated annealing (Brunger, Kuriyan, and Karplus, 1987). Most recently, improved targets for refinement of incomplete, error-containing models have been obtained using the more general maximum likelihood formulation (Murshudov, Vagin, and Dodson, 1997; Pannu et al., 1998). 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 (Adams et al., 1999). 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 (Pannu et al., 1998;

Adams et al., 1999). We anticipate that the maximum likelihood refinement method will be extended further to incorporate multivariate statistical analysis, thus, allowing multiple models to be refined simultaneously against the experimental data without introducing bias (Read, 2001).

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 noncrystallographic symmetry averaging. Tests have shown that the method can be successfully combined with the technique of simulated annealing (Chen, Blanc, and Chapman, 1999).

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 to improve the radius of convergence of refinement (Rice and Brunger, 1994). 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 to refine anisotropic parameters in the program REFMAC has greatly improved the speed with which such models can be generated and tested (Murshudov et al., 1999). The method has been shown to improve the crystallographic R-value and free R-value as well as the fit to geometric targets for data with resolution higher than 2Å.

VALIDATION

Validation of macromolecular models and their experimental data (Vaguine, Richelle, and Wodak, 1999) is an essential part of structure determination (Kleywegt, 2000). Validation is important both during the structure solution process and at the time of coordinate and data deposition at the Protein Data Bank, where extensive validation criteria are also applied (Berman et al., 2000). See Chapters 14 and 15 for descriptions of validation methods based on stereochemistry and atomic packing. In the future, the repeated application of validation criteria in automated structure solution will help avoid errors that currently occur as a result of subjective manual interpretation of data and models.

CHALLENGES TO AUTOMATION

Noncrystallographic Symmetry

It is not uncommon for macromolecules to crystallize with more than one copy in the asymmetric unit. This result leads to relationships between atoms in real space and diffraction intensities in reciprocal space. These relationships can be exploited in the structure solution process. However, the identification of noncrystallographic symmetry (NCS) is generally a manual process. A method for automatic location of proper NCS (i.e., a rotation axis) has been shown to be successful even at low resolution (Vonrhein and Schulz, 1999). A more general approach to finding NCS relationships uses skeletonization of electron density maps (Spraggon, 1999). A monomer envelope is calculated from the solvent mask generated by solvent flattening. The

NCS relationships between monomer envelopes can then be determined using standard molecular replacement methods.

These methods could be used in the future to automate the location of NCS operators and the determination of molecular masks. In the case of experimental phasing using heavy atoms or anomalous scatterers, it is possible to locate the NCS from the sites (Lu, 1999). The RESOLVE program automates this process such that NCS averaging can be automatically performed as part of the phase improvement procedure.

Disorder

Except in the rare case of very well-ordered crystals of extremely rigid molecules, disorder of one form or another is a component of macromolecular structures. This disorder may take the form of discrete conformational substates for side chains (Wilson and Brunger, 2000) or surface loops, or small changes in the orientation of entire molecules throughout the crystal. The degree to which this disorder can be identified and interpreted typically depends on the quality of the diffraction data. With low-to medium-resolution data, dual side chain conformations are occasionally observed. With high-resolution data (1.5Å or better) multiple side chain and main chain conformations are often seen. The challenge for automated structure solution is the identification of the disorder and its incorporation into the atomic model without the introduction of errors as a result of misinterpreting the data. Disorder of whole molecules within the crystal, as a result of small differences in packing between neighboring unit cells, cannot be visualized in electron density maps. However, the effect on refinement statistics such as the R and free-R value can be significant because no single atomic model can fit the observed diffraction data well. One approach to the problem is to simultaneously refine multiple models against the data (Burling and Brunger, 1994). An alternative approach is the refinement of Translation-Libration-Screw (TLS) parameters for whole molecules or subdomains of molecules (Winn, Isupov, and Murshudov, 2001). This introduces only a few additional parameters to be refined while still accounting for the majority of the disorder. However, it still remains a challenge to automatically identify subdomains.

CONCLUSION

Over the last decade of the twentieth century there have been many significant advances toward automated structure determination. Programs such as SOLVE (Terwilliger and Berendzen, 1999a), RESOLVE (Terwilliger, 2000), and the warpNtrace suite (Perrakis et al., 1999) combine large functional blocks in an automated fashion. The program CNS (Brunger et al., 1998) provides a framework in which different algorithms can be combined and tested using a powerful scripting language. Progress toward full automation will be made in the short term by linking existing programs together using scripting languages or the World Wide Web. However, a long-term solution will require the construction of a fully integrated system that makes use of the latest advances in crystallographic algorithms and computer science. The software that truly automates the crystallographic process will need to be intimately associated with data collection and processing. We anticipate that the next generation of automated software will permit the heavy atom location and phasing steps of structure solution to be performed in a few minutes. This speed will enable real time assessment of diffraction data as

it is collected at synchrotron beamlines. Map interpretation will be significantly faster than at present, with initial analysis of the electron density taking minutes rather than the hours or days required currently.

REFERENCES

- Abrahams JP (1997): Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. *Acta Crystallogr D* 53:371–76.
- Adams PD, Pannu NS, Read RJ, Brunger AT (1999): Extending the limits of molecular replacement through combined simulated annealing and maximum likelihood refinement. *Acta Crystallogr D* 55:181–90.
- Allen FH, Kennard O, Taylor R (1983): Systematic analysis of structural data as a research technique in organic chemistry. *Acc Chem Res* 16:146–53.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The Protein Data Bank. *Nucleic Acids Res* 28:235–42.
- Blessing RH, Smith GD (1999): Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. *J Appl Crystallogr* 32:664–70.
- Bruker AXS, Inc. (2001): *Analytical X-ray Solutions*. Madison, WI.
- Brunger AT (1992): The Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472–4.
- Brunger AT (1997): Patterson correlation searches and refinement. *Methods Enzymol* 276:558–80.
- Brunger AT, Kuriyan J, Karplus M (1987): Crystallographic R factor refinement by molecular dynamics. *Science* 235:458–60.
- Brunger AT, Adams PD, Clore GM, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998): Crystallography & NMR system (CNS) A new software system for macromolecular structure determination. *Acta Crystallogr D* 54:905–21. [The design and implementation of the widely used CNS program is described. The use of a scripting language to develop, test, and implement new features is a powerful feature of the software.]
- Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Sali A, Studier FW, Swaminathan S (1999): Structural genomics: beyond the human genome project. *Nat Genet* 23:151–7.
- Burling FT, Brunger AT (1994): Thermal motion and conformational disorder in protein crystal structures: Comparison of multi-conformer and time-averaging models. *Israel J Chem* 34:165–75.
- Chen Z, Blanc E, Chapman MS (1999): Real-space molecular-dynamics structure refinement. *Acta Crystallogr D* 55:464–8.
- Collaborative Computational Project, Number 4 (1994): The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 50:760–3.
- Cowtan K (1998): Modified phased translation functions and their application to molecular-fragment location. *Acta Crystallogr D* 54:750–6.
- Cowtan K (1999): Error estimation and bias correction in phase-improvement calculations. *Acta Crystallogr D* 55:1555–67.
- Cowtan K (2001): Fast Fourier feature recognition. *Acta Crystallogr D* 57:1435–44.
- Dauter Z, Dauter M (1999): Anomalous signal of solvent bromides used for phasing of lysozyme. *J Mol Biol* 289:93–101.

- Dauter Z, Dauter M, de La Fortelle E, Bricogne G, Sheldrick GM (1999): Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J Mol Biol* 289:83–92. [If accurate diffraction data are collected at a wavelength close to that of Cu-K α radiation the anomalous signal present from sulfur atoms can be used to generate useful phase information. This technique will probably be used extensively in the future to solve protein structures.]
- Garman E (1999): Cool data: quantity AND quality. *Acta Crystallogr* D55:1641–53.
- Glykos NM, Kokkinidis M (2000): A stochastic approach to molecular replacement. *Acta Crystallogr* D56:169–74.
- Gonzalez A, Pedelacq J-D, Sola M, Gomis-Rueth FX, Coll M, Samama J-P, Benini S (1999): Two-wavelength MAD phasing: in search of the optimal choice of wavelengths. *Acta Crystallogr* D55:1449–58.
- Grosse-Kunstleve RW, Brunger AT (1999): A highly automated heavy-atom search procedure for macromolecular structures. *Acta Crystallogr* D55:1568–77.
- Hendrickson WA (1985): Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol* 115:252–70.
- Hendrickson WA (1991): Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254:51–8.
- Holton T, Ioerger TR, Christopher JA, Sacchettini JC (2000): Determining protein structure from electron-density maps using pattern matching. *Acta Crystallogr* D56:722–34.
- Jones TA (1978): A graphics model building and refinement system for macromolecules. *J Applied Crystallogr* 11:268–72.
- Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M (1991): Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr* A47:110–9.
- Kissinger CR, Gehlhaar DK, Fogel DB (1999): Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr* D55:484–91.
- Kleywegt GJ (2000): Validation of protein crystal structures. *Acta Crystallogr* D56:249–65.
- Kleywegt GJ, Jones TA (1997): Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallogr* D53:179–85.
- Kleywegt GJ, Jones TA (1998): Databases in Protein Crystallography. *Acta Crystallogr* D54:1119–31.
- La Fortelle E de, Bricogne G (1997): Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol* 276:472–94.
- Levitt DG (2001): A new software routine that automates the fitting of protein X-ray crystallographic electron-density maps. *Acta Crystallogr* D57:1013–9.
- Lu G (1999): FINDNCS: A program to detect non-crystallographic symmetries in protein crystals from heavy atoms sites. *J Applied Crystallogr* 32:365–8.
- McRee DE (1999): XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* 125:156–65.
- Montelione GT, Anderson S (1999): Structural genomics: keystone for a human proteome project. *Nat Struct Biol* 6:11–2.
- Mowbray SL, Helgstrand C, Sigrell JA, Cameron AD, Jones TA (1999): Errors and reproducibility in electron-density map interpretation. *Acta Crystallogr* D55:1309–19.
- Murshudov GN, Vagin AA, Dodson EJ (1997): Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr* D53:240–55.
- Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999): Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr* D55:247–55.

- Oldfield T (2002): A semi-automated map fitting procedure. In: Bourne PE, Watenpaugh K, editors. *Crystallographic Computing 7: Macromolecular Crystallographic Data (Crystallographic Computing)*. Oxford: Oxford University Press. In press.
- Otwinowski Z (1991): Maximum likelihood refinement of heavy atom parameters. In: Wolf W, Evans PR and Leslie AGW, editors. *Isomorphous Replacement and Anomalous Scattering, Proc. Daresbury Study Weekend*. Warrington: SERC Daresbury Laboratory, p. 80–5.
- Pannu NS, Murshudov GM, Dodson EJ, Read RJ (1998): Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr D*54:1285–94.
- Perrakis A, Sixma TK, Wilson KS, Lamzin VS (1997): wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D*53:448–55.
- Perrakis A, Morris R, Lamzin VS (1999): Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–63. [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 proteinlike patterns are extracted. This method is currently used by crystallographers to automate model building when high resolution data are available (approximately 2.0Å or better).]
- Read RJ (1999): Detecting outliers in non-redundant diffraction data. *Acta Crystallogr D*55:1759–64.
- Read RJ (2001): Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D*57:1373–82.
- Rice LM, Brunger AT (1994): Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* 19:277–90.
- Rice LM, Earnest TN, Brunger AT (2000): Single wavelength anomalous diffraction phasing revisited: a general phasing method? *Acta Crystallogr D*56:1413–20.
- Rossmann MG, Blow DM (1962): The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr* 15:24–31.
- Sheldrick GM, Gould RO (1995): Structure solution by iterative peaklist optimization and tangent expansion in space group P1. *Acta Crystallogr B*51:423–31.
- Sheriff S, Klei HE, Davis ME (1999): Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement problems. *J Applied Crystallogr* 32:98–101.
- Spraggon G (1999): 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*55:458–63.
- Terwilliger TC (1994): MAD phasing: Bayesian estimates of F_A . *Acta Crystallogr D*50:11–6.
- Terwilliger TC (1999): σ_R^2 , a reciprocal-space measure of the quality of macromolecular electron-density maps. *Acta Crystallogr D*55:1174–1178.
- Terwilliger TC (2000): Maximum-likelihood density modification. *Acta Crystallogr D*56:965–72. [A procedure is described for reciprocal-space maximization of a likelihood function based on experimental phases and characteristics of the electron-density map. This powerful approach to phase improvement is able to generate minimally biased phase estimates and will be a valuable tool in the future for all aspects of phase improvement and phase combination.]
- Terwilliger TC, Berendzen J (1999a): Automated MAD and MIR structure solution. *Acta Crystallogr D*55:849–61. [This important paper describes in detail the SOLVE program and the underlying algorithms that are used to automate heavy atom location and phasing. It is a good example of the considerations that go into automating a complex crystallographic process.]

- Terwilliger TC, Berendzen J (1999b): Discrimination of solvent from protein regions in native Fourier as a means of evaluating heavy-atom solutions in the MIR and MAD methods. *Acta Crystallogr D* 55:501–5.
- Terwilliger TC, Berendzen J (1999c): Evaluation of macromolecular electron-density map quality using the correlation of local r.m.s. density. *Acta Crystallogr D* 55:1872–7.
- Turk D (2000): MAIN 96: An interactive software for density modifications, model building, structure refinement and analysis. In: Bourne PE, Watenpaugh K, editors. *Crystallographic Computing 7: Macromolecular Crystallographic Data (Crystallographic Computing)*. Oxford: Oxford University Press.
- Vaguine AA, Richelle J, Wodak SJ (1999): 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* 55:191–205.
- Vonrhein C, Schulz GE (1999): Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX. *Acta Crystallogr D* 55:225–9.
- Walsh MA, Evans G, Sanishvili R, Dementieva I, Joachimiak A (1999a): MAD data collection—current trends. *Acta Crystallogr D* 55:1726–32.
- Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A (1999b): Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr D* 55:1168–73.
- Weeks CM, Miller R (1999): The design and implementation of SnB v2.0. *J Applied Crystallogr* 32:120–4.
- Wilson MA, Brunger AT (2000): The 1.0Å crystal structure of Ca²⁺ bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J Mol Biol* 301:1237–56.
- Winn MD, Isupov MN, Murshudov GN (2001): Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* 57:122–33.
- Yeates TO (1997): Detecting and overcoming crystal twinning. *Methods Enzymol* 276:344–58.