PERGAMON

Progress in Biophysics & Molecular Biology 72 (1999) 135–155

Progress in Biophysics & Molecular Biology

www.elsevier.com/locate/pbiomolbio

Annealing in crystallography: a powerful optimization tool

Axel T. Brunger[a, b,*], Paul D. Adams[a, b], Luke M. Rice[a, b]

[a]*The Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, CT 06511, USA*
[b]*Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06511, USA*

Received 1 February 1999

1. Introduction

Over the last decade, developments in molecular biology, X-ray diffraction and nuclear magnetic resonance (NMR) instrumentation, and computational methods have allowed nearly exponential growth of macromolecular structural studies. The analysis of data from these studies generally requires sophisticated computational procedures which culminate in refinement and structure validation. These procedures can be formulated as the chemically constrained or restrained nonlinear optimization of a target function, which usually measures the agreement between observed data and data computed from an atomic model. The ultimate goal is to optimize the simultaneous agreement of an atomic model with observed data and with a priori chemical information.

The target function used for this optimization normally depends on several atomic parameters, but most importantly on atomic coordinates. The large number of adjustable parameters (typically at least three times the number of atoms in the model) gives rise to a very complicated target function. This in turn produces what is known as the multiple minima problem: the target function contains many local minima in addition to the global minimum. These local minima tend to defeat gradient-descent optimization techniques such as conjugate gradient or least-squares methods (Press et al., 1986). These methods are simply not capable of sampling molecular conformations thoroughly enough to find the most optimal model if the starting one is far from the correct structure.

Simulated annealing (Kirkpatrick et al., 1983) is an optimization technique particularly well suited to overcoming the multiple minima problem. Unlike gradient-descent methods,

* Corresponding author. Tel.: +1-203-432-6143; fax: +1-203-432-6946.
E-mail address: brunger@laplace.csb.yale.edu (A.T. Brunger)

PII: S0079-6107(99)00004-8

simulated annealing can cross barriers between minima and thus can explore a greater volume of the parameter space to find better models in deeper minima. Following its introduction to crystallographic refinement (Brunger et al., 1987), and NMR structure calculation (Brunger et al., 1986) and refinement (Kaptein et al., 1985), there have been major improvements of the original method in four principal areas: the measure of model quality, the search of the parameter space, the target function, and the modeling of conformational variability.

For crystallographic refinement, the introduction of cross-validation (the 'free' *R*-value) (Brunger, 1992) has significantly reduced the danger of overfitting the diffraction data. The complexity of the conformational space has been reduced by the introduction of torsion-angle molecular dynamics (Rice and Brunger, 1994), which decreases the number of adjustable parameters that describe a model approximately tenfold. The target function has been improved by incorporating the concept of maximum-likelihood which takes into account model error, model incompleteness, and errors in the experimental data (Read, 1986; Adams et al., 1997). Advances have also been obtained for NMR structure determination. Cross-validation has been shown to be valid for NMR structure calculation (Brunger et al., 1993), and the radius of convergence has been increased by the use of torsion-angle molecular dynamics (Stein et al., 1997) and the introduction of variable target functions (Braun and Go, 1985; Braun, 1987; Nilges et al., 1988a, 1991). Finally, the sampling power of simulated annealing can be used for exploring the molecule's conformational space in cases where the molecule undergoes dynamic motion or static disorder through multi-conformer models for both X-ray crystallography and solution NMR (Kuriyan et al., 1991; Burling and Brunger, 1994; Bonvin and Brunger, 1995, 1996; Burling et al., 1996).

2. The target function

In essence, macromolecular structure calculation and refinement is a search for the global minimum of a target function,

$$E = E_{\text{chem}} + w_{\text{data}} E_{\text{data}}, \tag{1}$$

as a function of the parameters of an atomic model, in particular atomic coordinates. E_{chem} comprises empirical information about chemical interactions; it is a function of all atomic positions, describing covalent (bond lengths, bond angles, torsion angles, chiral centers and planarity of aromatic rings) and nonbonded (intra-molecular as well as inter-molecular and symmetry-related) interactions. E_{data} describes the difference between observed and calculated data, and w_{data} is a weight appropriately chosen to balance the gradients (with respect to atomic parameters) arising from the two terms.

3. A priori chemical information

The geometric energy function E_{chem} (Hendrickson, 1985) consists of terms for covalent bonds, bond angles, chirality, planarity and nonbonded repulsion. The parameters for the covalent terms can be derived from average geometry and root-mean-square (rms) deviations

observed in a small-molecule data base. Extensive statistical analyses were undertaken for the chemical moieties of proteins (Engh and Huber, 1991) and of polynucleotides (Parkinson et al., 1996) using the Cambridge crystallographic database (Allen et al., 1983). Analysis of the ever increasing number of atomic resolution macromolecular crystal structures will no doubt cause some modifications of these parameters in the future (Dauter et al., 1995; Stec et al., 1995; Sevcik et al., 1996; Vlassi et al., 1998).

In both NMR and X-ray crystallography, it is common to use a purely repulsive quartic function ($E_{\text{repulsive}}$) for the nonbonded interactions which are included in E_{chem} (Hendrickson, 1985),

$$E = \sum_{ij} [(cR_{ij}^{\min})^n - R_{ij}^n]^m, \tag{2}$$

where R_{ij} is the distance between two atoms i and j, $R_{ij}^{\min}$ is the van der Waals radius for a particular atom pair ij, $c \leqslant 1$ is a constant that is sometimes used to reduce the radii, and $n=2$, $m=2$ or $n=1$, $m=4$. Van der Waals attraction and electrostatic interactions are usually not included in structure calculation and refinement. These simplifications are valid since the experimental data contains information that is able to produce atomic conformations consistent with actual nonbonded interactions. In fact, atomic resolution crystal structures can be used to derive parameters for electrostatic energies (Pearlman and Kim, 1990). If the experimental information is insufficient to fully determine the macromolecular structure, use of electrostatic and simulated solvent interactions can bias the structure towards the theoretical nonbonded model.

Geometric energy functions are related to empirical energy functions that were developed for energy-minimization and molecular-dynamics studies of macromolecules (see Karplus and Petsko, 1990 for a review). These empirical energy functions were not designed for structure determination, and therefore required some modification for use in macromolecular structure refinement (Brunger et al., 1986; Nilges et al., 1988b; Brunger et al., 1989, 1990; Fujinaga et al., 1989; Weis et al., 1989). Recently, crystallographic simulated annealing refinement was implemented with a purely geometric energy function (Adams et al., 1997) which provides uniformity among different crystallographic refinement programs and simplifies the generation of parameters for new chemical compounds. Similar developments for solution NMR are in progress (M. Nilges, personal communication).

4. X-ray diffraction data

The conventional form of $E_{\text{X-ray}}$ consists of the crystallographic residual E^{LSQ}, defined as the sum over the squared differences between the observed $|F_{\text{o}}|$ and calculated $|F_{\text{c}}|$ structure factor amplitudes for a particular atomic model,

$$E_{\text{X-ray}} = E^{\text{LSQ}} = \sum_{hkl} (|F_{\text{o}}| - k|F_{\text{c}}|)^2, \tag{3}$$

where *hkl* are the indices of the reciprocal lattice points of the crystal, $|F_o|$ and $|F_c|$ are the observed and calculated structure-factor amplitudes, and k is a relative scale factor.

Reduction of E^{LSQ} can result from improvement in the atomic model, but also from an accumulation of systematic errors in the model or fitting noise in the data (Silva and Rossmann, 1985). The least-squares residual is therefore poorly justified when the model is far away from the correct one or incomplete (Read, 1986). An improved target for macromolecular refinement can be obtained using a maximum-likelihood formulation (Read, 1986, 1990; Bricogne, 1991, 1993; Pannu and Reed, 1996; Murshudov et al., 1997). The goal of the maximum-likelihood method is to determine the probability of making a measurement, given the model, estimates of the model's errors and those of the measured intensities. The effects of model errors (incorrectly placed and missing atoms) on the calculated structure factors are first quantified with σ_A values, which correspond roughly to the fraction of each structure factor that is expected to be correct. However, overfitting of the diffraction data causes the model bias to be underestimated and undercorrected in the σ_A values. The effect of this overfitting can be reduced by cross-validating σ_A values, i.e. by computing them from a randomly selected test set which is excluded from the summation on the right hand side of Eq. (3) (Brunger, 1992; Kleywegt and Brunger, 1996). The expected values of $\langle F_o \rangle$ and the corresponding variance (σ^2_{ML}) are derived from σ_A, the observed F_o, and calculated F_c (Read, 1986). These quantities can be readily incorporated into a maximum-likelihood target function,

$$E_{\text{X-ray}} = E_{\text{ML}} = \sum_{hkl \in \text{working set}} (1/\sigma^2_{\text{ML}})(|F_o| - \langle |F_o| \rangle)^2. \tag{4}$$

In order to achieve an improvement over the least-squares residual (Eq. (3)), cross-validation was found to be essential (Adams et al., 1997) for the computation of σ_A and its derived quantities in Eq. (4).

For many crystal structures, some initial experimental phase information is available from either isomorphous heavy atom replacement or multiwavelength anomalous diffraction methods (Hendrickson, 1991). These phases represent additional observations which can be incorporated in the refinement target. The maximum likelihood formulation naturally extends itself to incorporation of this information (Bricogne, 1997; Pannu et al., 1998). Tests have shown that the addition of experimental phase information greatly improves the results of refinement (Pannu et al., 1998; Adams et al., 1999).

Pannu and Reed (1996) have developed an efficient Gaussian approximation for the case of structure-factor amplitudes with no prior phase information, termed the MLF target function. In the limit of a perfect model MLF reduces to the traditional least-squares residual (Eq. (3)) with $1/\sigma^2$ weighting. In the case where prior phase information is included, the integration over the phase angles is carried out numerically, and is termed the MLHL target (Pannu et al., 1998). A maximum likelihood function which expresses the probability distributions in terms of observed intensities has also been developed, and is termed MLI (Pannu and Read, 1996).

5. NMR data

A common form of E_{NMR} describes nuclear Overhauser effect (NOE)-derived distance restraints and scalar J-coupling constants derived dihedral angle restraints using flat-bottomed parabolic functions (Clore et al., 1986),

$$w_{\mathrm{NMR}}E_{\mathrm{NMR}} = w_{\mathrm{NOE}}E_{\mathrm{NOE}} + w_{\mathrm{dihe}}E_{\mathrm{dihe}}, \tag{5}$$

$$E_{\mathrm{NOE}} = \sum_{\mathrm{NOEs}} \begin{cases} (d - d_{\mathrm{upper}})^2 & d_{\mathrm{upper}} < d \\ 0 & d_{\mathrm{lower}} < d < d_{\mathrm{upper}} \\ (d_{\mathrm{lower}} - d)^2 & d < d_{\mathrm{lower}} \end{cases}, \tag{6}$$

$$E_{\mathrm{dihe}} = \sum_{\mathrm{dihedrals}} \begin{cases} (\phi - \phi_{\mathrm{upper}})^2 & \phi_{\mathrm{upper}} < \phi \\ 0 & \phi_{\mathrm{lower}} < \phi < \phi_{\mathrm{upper}} \\ (\phi_{\mathrm{lower}} - \phi)^2 & \phi < \phi_{\mathrm{lower}} \end{cases}. \tag{7}$$

Here d denotes the distance between a particular pair of spins in the model, d_{lower} and d_{upper} are, respectively, lower and upper bounds for the distance restraint derived from the isolated spin-pair approximation or from NOE-backcalculation (see Nilges, 1996 for a review), ϕ denotes the dihedral angle formed by four atoms in the model, ϕ_{lower} and ϕ_{upper} are, respectively, lower and upper bounds for the dihedral angle restraint derived from scalar J-coupling constant measurements and empirical Karplus (Karplus, 1963) relationships. In the case of ambiguous NOE assignments, overlapping NOEs, or motion of methyl groups and aromatic rings, appropriate averaging schemes must be used (Wüthrich, 1986; Nilges, 1995, 1996) when computing d, d_{upper}, and d_{lower}. Direct refinement against scalar J-coupling constant data (Kim and Prestegard, 1990; Garrett et al., 1994; Mierke et al., 1994) or against empirical ^{1}H and ^{13}C chemical shift databases which correlate molecular conformation and chemical shifts (Kuszewski et al., 1995a,b; Oldfield, 1995) is also possible. Because early atomic models can contain very significant violations of experimental distance restraints, E_{NOE} is often modified so that it becomes linear for large violations: the 'soft-square' potential (Nilges et al., 1988c). This modification is important for convergence – without it the large violations give rise to enormous forces which cause numerical instabilities during simulated annealing.

6. Additional information

Additional constraints or restraints may be used to improve the ratio of observables to parameters. For example, atoms can be grouped so that they move as rigid bodies during refinement, or bond lengths and bond angles can be kept fixed (Diamond, 1971; Sussman et al., 1977; Rice and Brunger, 1994). In the crystallographic case, the existence of noncrystallographic symmetry can be used to average over equivalent molecules and thereby to reduce noise in the diffraction data (Weis et al., 1989). In the NMR case, deuterium exchange protection can often be used to infer the presence of amide–carbonyl hydrogen bonds. These inferred hydrogen bonds are then modeled as distance restraints, thereby increasing the

observable to parameter ratio. Empirical dihedral-angle conformational databases of proteins have been recently added to the list of possible additional restraints (Kuszewski et al., 1996).

7. Weighting

The weight w_{data} (Eq. (1)) balances the forces arising from E_{data} and E_{chem}. The choice of w_{data} can be critical: if w_{data} is too large, the refined structure will show unphysical deviations from ideal geometry; if w_{data} is too small, the refined structure will not satisfy the observed data. Automated protocols to provide initial estimates for optimal weighting have been developed (Brunger et al., 1989; Adams et al., 1997). However, independent information must be used (e.g. cross-validation) to objectively obtain the best possible weight for X-ray diffraction data (Brunger, 1992) and NMR data (Brunger et al., 1993).

8. Searching conformational space

Annealing denotes a physical process wherein a solid is heated until all particles randomly arrange themselves in a liquid phase, and then is cooled slowly so that all particles arrange themselves in the lowest energy state. By formally defining the target E (Eq. (1)) to be the equivalent of the potential energy of the system, one can simulate the annealing process (Kirkpatrick et al., 1983). There is no guarantee that simulated annealing will find the global minimum (except in the case of an infinitely long search) (Laarhoven and Aarts, 1987). Compared to conjugate-gradient minimization where search directions must follow the gradient, simulated annealing achieves more optimal solutions by allowing motion against the gradient (Kirkpatrick et al., 1983). The likelihood of uphill motion is determined by a control parameter referred to as temperature. The higher the temperature, the more likely it is that simulated annealing will overcome barriers. It should be noted that the simulated annealing temperature normally has no physical meaning and merely determines the likelihood of overcoming barriers of the target function.

The simulated annealing algorithm requires a generation mechanism to create a Boltzmann distribution at a given temperature T. Simulated annealing also requires an annealing schedule, that is, a sequence of temperatures $T_1 > T_2 > \ldots > T_n$ at which the Boltzmann distribution is computed. Implementations of the generation mechanism differ in the way they generate a transition from one set of parameters to another which is consistent with the Boltzmann distribution at given temperature. The two most widely used generation mechanisms are Metropolis Monte Carlo (Metropolis et al., 1953) and molecular dynamics (Verlet, 1967) simulations. For NMR structure calculation, both molecular dynamics and Monte Carlo have been successfully used (Curro, 1974; Kaptein et al., 1985; Brunger et al., 1986; Ulyanov et al., 1993; Xu and Krishna, 1995). For X-ray crystallographic refinement, molecular dynamics proved extremely successful (Brunger et al., 1987) whereas Monte Carlo methods have yet to be shown to be effective.

9. Metropolis Monte Carlo

The Metropolis Monte Carlo algorithm (Metropolis et al., 1953) simulates the thermal equilibrium of a system for a fixed value of the temperature T. In the limiting case of $T = 0$, Monte Carlo is equivalent to a gradient descent method; the only moves allowed are the ones that lower the target function until a local minimum is reached. At a finite temperature, however, Monte Carlo allows uphill moves and hence allows barrier-crossings.

The advantage of the Metropolis Monte Carlo algorithm is its simplicity. A disadvantage concerns the efficient choice of the parameter shifts that define the Monte Carlo move. Ideally, this choice should in some way reflect the topology of the search space. In the case of a covalently connected macromolecule, random shifts of atomic coordinates have a high rejection rate: they immediately violate geometric restrictions such as bond lengths and bond angles. This problem can be alleviated in principle by carrying out the Monte Carlo simulation in a suitably chosen set of internal coordinates such as torsions about bonds, or normal modes of vibration, or by relaxing the strained coordinates through minimization (Li and Scheraga, 1987; Saunders, 1987; Abagyan and Argos, 1992).

10. Molecular dynamics

A suitably chosen set of atomic parameters can be viewed as generalized coordinates that are propagated in time by the classical (Hamilton) equations of motion (Goldstein, 1980). If the generalized coordinates represent the x, y, z positions of the atoms of a molecule, the Hamilton equations of motion reduce to the more familiar Newton's second law,

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = -\nabla_i E. \tag{8}$$

The quantities m_i and r_i are respectively the mass and coordinates of atom i, and E is given by Eq. (1). The solution of the partial differential equations (Eq. (8)) is achieved numerically using finite-difference methods (Verlet, 1967). This approach is referred to as molecular dynamics.

Initial velocities for the integration of Eq. (8) are usually assigned randomly from a Maxwell distribution at the appropriate temperature. Assignment of different initial velocities will produce a somewhat different structure after simulated annealing. By performing several refinements with different initial velocities one can therefore improve the chances of success of simulated annealing refinement. Furthermore, this improved sampling can be used to determine discrete disorder and conformational variability (see below).

Although Cartesian (i.e. flexible bond lengths and bond angles) molecular dynamics places restraints on bond lengths and bond angles (through E_chem, Eq. (1)), one might want to implement these restrictions as constraints, i.e. fixed bond lengths and bond angles (Diamond, 1971). This is supported by the observation that the deviations from ideal bond lengths and bond angles are usually small in X-ray crystal structures, and even smaller in NMR structures. Indeed, fixed-length constraints have been applied to structure calculation by least-squares or conjugate-gradient minimization (Diamond, 1971; Braun and Go, 1985), and by Monte Carlo minimization (Curro, 1974; Ulyanov et al., 1993; Xu and Krishna, 1995). It is only recently,

however, that efficient and robust algorithms have become available for molecular dynamics in torsion-angle space (Bae and Haug, 1987, 1988; Jain et al., 1983; Mathiowetz et al., 1994; Rice and Brunger, 1994).

Using an approach that retains the Cartesian-coordinate formulation of the target function and its derivatives with respect to atomic coordinates makes calculations remains relatively straightforward and topology independent (Rice and Brunger, 1994). In this formulation, however, the expression for the acceleration becomes a function of positions and velocities. Iterative equations of motion for constrained dynamics in this formulation can be derived and solved by finite difference methods (Abramowitz and Stegun, 1968). This method is numerically very robust and has a significantly increased radius of convergence in crystallographic refinement and NMR structure calculation compared to Cartesian molecular dynamics (Rice and Brunger, 1994; Stein et al., 1997).

11. Temperature control

Simulated annealing requires the control of the temperature during molecular dynamics. The current temperature of the simulation (T_{curr}) is computed from the kinetic energy,

$$E_{\mathrm{kin}} = \sum_{i}^{n\ \mathrm{atoms}} \frac{1}{2} m_i (\partial r_i / \partial t)^2, \tag{9}$$

of the molecular dynamics simulation,

$$T_{\mathrm{curr}} = \frac{2E_{\mathrm{kin}}}{3nk_{\mathrm{B}}}. \tag{10}$$

Here n is the number of degrees of freedom and k_{B} is Boltzmann's constant. One commonly used approach to control the temperature of the simulation consists of coupling the equations of motion to a heat bath. A friction term (γ_i) (Berendsen et al., 1984) to control the temperature,

$$-m_i \gamma_i v_i [1 - (T/T_{\mathrm{curr}})], \tag{11}$$

where v_i are the velocities of the atoms, can be added to right hand side of Eq. (8). This method generalizes the concept of friction by allowing a negative friction coefficient and by determining the friction coefficient and its sign by the ratio of the current simulation temperature to the target temperature T_{curr}.

12. Why does simulated annealing work?

The goal of any optimization problem is to find the global minimum of a target function. In the case of macromolecular structure calculation and refinement, one searches for the conformation or conformations of the molecule that best fit the experimental data and that simultaneously maintain reasonable covalent and noncovalent interactions. Simulated

annealing refinement has a much larger radius of convergence than conjugate-gradient minimization (see below). It must therefore be able to find a lower minimum of the target E (Eq. (1)) than the local minimum found by simply moving along the negative gradient of E. Paradoxically, the very reasons that make simulated annealing such a powerful refinement technique (the ability to overcome barriers in the target energy function) would seem to prevent it from working at all. If it crosses barriers so easily, what allows it to stay in the vicinity of the global minimum?

It is most easy to visualize this property of simulated annealing in the case of molecular dynamics. By specifying a fixed temperature, the system essentially gains a certain inertia which allows it to cross energy barriers of the corresponding target function (Eq. (10)). The target temperature must be large enough to overcome smaller barriers (e.g. Fig. 1a) but low enough to ensure that the system will not 'climb out' out of the global minimum if it manages to arrive there. While temperature itself is a global parameter of the system, temperature fluctuations arise principally from local conformational transitions – for example from an amino acid sidechain falling into the correct orientation. These local changes tend to lower the value of the target E, thus increasing the kinetic energy, and hence the temperature, of the system. Once the

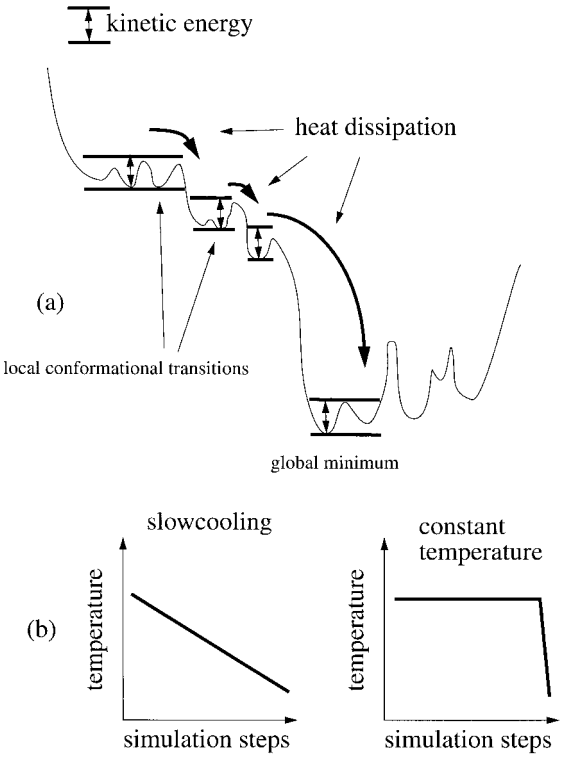

Fig. 1. (a) Schematic explanation of molecular-dynamics based simulated annealing. The kinetic energy of the system allows local conformational transitions with barriers small than the kinetic energy. If a larger drop in energy is encountered the excess kinetic energy is dissipated through the friction term (Eq. (11)). It is thus unlikely that the system can climb out of the global minimum once it has reached it. (b) Possible annealing schedules: slow-cooling and constant temperature followed by quenching.

temperature coupling (Eq. (11)) has removed this excess kinetic energy through heat dissipation, the reverse transition is very unlikely, since it would require a localized increase in kinetic energy where the conformational change occurred in the first place. Temperature coupling maintains a sufficient amount of kinetic energy to allow local conformational corrections, but does not supply enough to allow escape from the global minimum. This explains the observation that on average the agreement with the experimental data will improve rather than worsen with simulated annealing.

13. Practical considerations

As Fig. 1a illustrates, the simulation temperature needs to be high enough to allow conformational transitions but not too high to avoid moving too far away from the initial structure. The optimum temperature for a given starting structure is a matter of trial and error. We empirically determined starting temperatures for a variety of simulated annealing protocols (Brunger, 1988; Adams et al., 1997) which should work for the average case. However, it might be worth trying a different temperature if a particularly difficult refinement problem is encountered. In particular, significantly higher temperatures are attainable using torsion-angle molecular dynamics. Note that each simulated annealing refinement run is subject to chance by using a random number generator to generate the initial velocities. Thus, multiple refinements must be run if systematic trends resulting from changes of certain parameters of the annealing schedule are to be studied. The best structure(s) among a set of refinements using different initial velocities and/or temperatures should be taken for further refinement or averaging (see below).

The annealing schedule employed can in principle be any function of the simulation step (or time domain). The two most commonly used protocols are linear slow-cooling or constant-temperature followed by quenching (Fig. 1b). A slight advantage is obtained with slow-cooling (Brunger et al., 1990). The duration of the annealing schedule is another parameter. Too short a protocol does not allow sufficient sampling of conformational space. Too long a protocol may waste computer time since it is more efficient to run multiple trials as opposed to one long refinement protocol (unpublished results).

14. Crystallographic refinement

In the crystallographic case, the limited radius of convergence of refinement arises not only from the high dimensionality of the parameter space, but also from the crystallographic phase problem. For new crystal structures, initial electron density maps must be computed from a combination of observed diffraction amplitudes and experimental phases where the latter are typically of poorer quality and lower resolution than the former. A different problem arises when structures are solved by molecular replacement (Hoppe, 1957; Rossmann and Blow, 1962) which uses a similar structure as a search model. In this case the resulting electron density maps can be severely 'model-biased', that is, they seem to confirm the existence of the

search model without providing clear evidence of actual differences between it and the true crystal structure. In either case, initial atomic models usually require extensive refinement.

Many examples have shown that simulated annealing refinement starting from initial models (obtained by standard crystallographic techniques) produces significantly better final models compared to those produced by least-squares or conjugate-gradient minimization. In a realistic test case (Adams et al., 1999), a series of models for the aspartic proteinase penicillopepsin was generated from homologous structures present in the Protein Data Bank. The sequence identity among these structures ranged from 100 to 25%, thus providing a set of models with increasing coordinate error compared to the refined structure of penicillopepsin. These models, after truncation of all residues to alanine, were all used as search models in molecular replacement against the native penicillopepsin diffraction data. In all cases the correct placement of the model in the penicillopepsin unit cell was found.

Both conjugate gradient minimization and simulated annealing were carried out in order to compare the performance of LSQ (the least-squares residual, MLF (the maximum likelihood target using amplitudes) and MLHL (the maximum likelihood target using amplitudes and experimental phase information). In the latter case, phases from single-isomorphous replacement were used. A very large number of conjugate gradient cycles were carried out in order to make the computational requirements equivalent for both minimization and simulated annealing. The conjugate gradient minimizations were converged, i.e. there was no change when further cycles were carried out.

For a given target function, simulated annealing always outperformed minimization (Fig. 2).

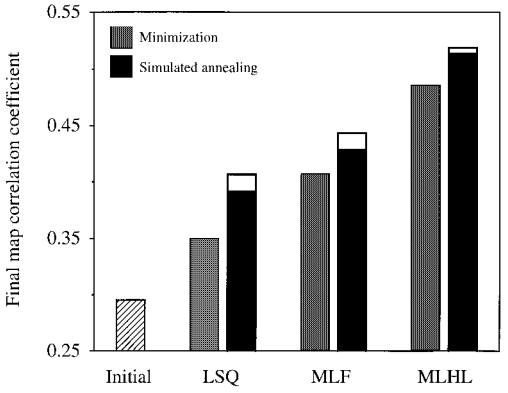

Fig. 2. Simulated annealing produces better models than extensive conjugate gradient minimization. Map correlation coefficients were computed before and after refinement against the native penicillopepsin diffraction data (Hsu et al., 1977) for the polyalanine model derived from *Mucor pusillus* pepsin (Newman et al., 1993). Correlation coefficients are between σ_A-weighted maps calculated from each model and from the published penicillopepsin structure. The observed penicillopepsin diffraction data was in space-group C2 with cell dimensions $a = 97.37$ Å, $b = 46.64$ Å, $c = 65.47$ Å and $\beta = 115.4°$. All refinements were carried out using diffraction data from the lowest resolution limit of 22.0 up to 2.0 Å. The MLHL refinements used single-isomorphous phases from a $K_3UO_2F_5$ derivative of the penicillopepsin crystal structure, which covered a resolution range from 22.0 to 2.8 Å. Simulated annealing refinements were repeated five times with different initial velocities. The numerical averages of the map correlation coefficients for the five refinements are shown as the hashed bars. The best map correlation coefficients from simulated annealing are shown as the white bars.

For a given starting model, the maximum likelihood targets outperformed the least-squares residual target for both minimization and simulated annealing, producing models with lower phase errors and higher map correlation coefficients when compared to the published penicillopepsin crystal structure (Fig. 2). This improvement is illustrated in σ_A-weighted electron density maps obtained from the resulting models (Fig. 3). The incorporation of experimental phase information further improved the refinement significantly despite the ambiguity in the SIR phase probability distributions. Thus, the most efficient refinement will make use of torsion angle dynamics simulated annealing, and prior phase information in the MLHL maximum likelihood target function.

Cross-validation is essential in the calculation of the maximum likelihood target (Read, 1986; Kleywegt and Brunger, 1996; Adams et al., 1997). Maximum-likelihood refinement without cross-validation gives much poorer results, as indicated by higher free R-values, higher $R_{free}-R$ differences, and larger phase errors. It should be noted that the normal R-value usually increases upon using the cross-validated maximum likelihood formulation. This is a consequence of the reduction of overfitting by this method.

Simulated annealing refinement is most useful when the initial model is relatively crude. Given a well-refined model, it offers little advantage over conventional methods, with the exception of providing information about the accuracy and conformational variability of the refined structure (see below).

15. NMR structure calculation

NMR experiments provide very specific, local information about macromolecules. In contrast to X-ray crystallography, where each data point contains information about the entire molecule, NMR spectroscopy provides information about inter-atomic distance pairs and specific dihedral angles (see above). These fundamental differences in the experimental data manifest themselves in fundamental differences in refinement requirements. The global nature of crystallographic data means that the initial model must not deviate too significantly from the final, refined one. The local nature of NMR data, on the other hand, means that random initial models can be productively refined (Nilges et al., 1988a). Furthermore, molecular-dynamics based simulated annealing can be used to automate the NOE-assignment process (Nilges, 1995). It is a testament to the remarkable power of molecular dynamics-based simulated annealing that it has enjoyed such great success in widely divergent applications.

Recent tests comparing torsion angle molecular dynamics refinement to Cartesian molecular dynamics refinement illustrate the advantages of the reduced variable approach (Stein et al., 1997). Torsion angle dynamics was compared to commonly used strategies which rely on Cartesian molecular dynamics, distance geometry, or both. The method has a higher success rate and efficiency than conventional simulated annealing algorithms which use Cartesian molecular dynamics or distance geometry combined with Cartesian molecular dynamics.

The application of torsion-angle molecular dynamics to refinement of a DNA dodecamer against NMR data provided even more striking results (Fig. 4). The starting atomic model consisted of two extended nucleotide strands. The 'YASAP' Cartesian molecular dynamics protocol (Nilges et al., 1991) failed to produce the correct structure whereas with torsion-angle

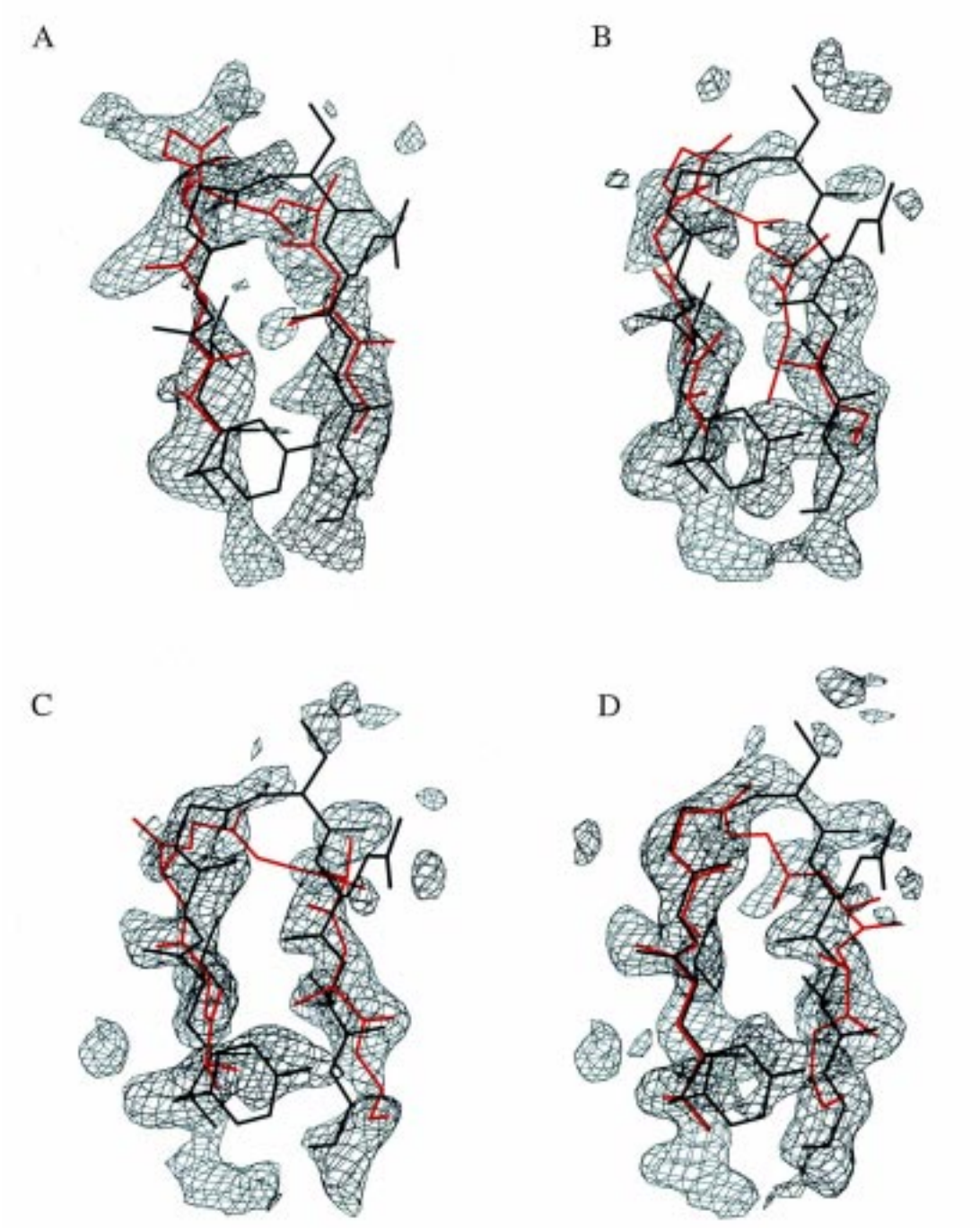

Fig. 3. Maximum likelihood targets significantly decrease model bias in simulated annealing refinement. σ_A-weighted electron density maps contoured at 1.25' for models from simulated annealing refinement with different targets. Residues 198–205 are shown with the published penicillopepsin crystal structure (Hsu et al., 1977) in black, and the model with the lowest free *R*-value from five independent refinements in red. (a) Initial electron-density map prior to refinement, (b) after refinement with the LSQ target, (c) after refinement with the MLF target, (d) after refinement with the MLHL target.

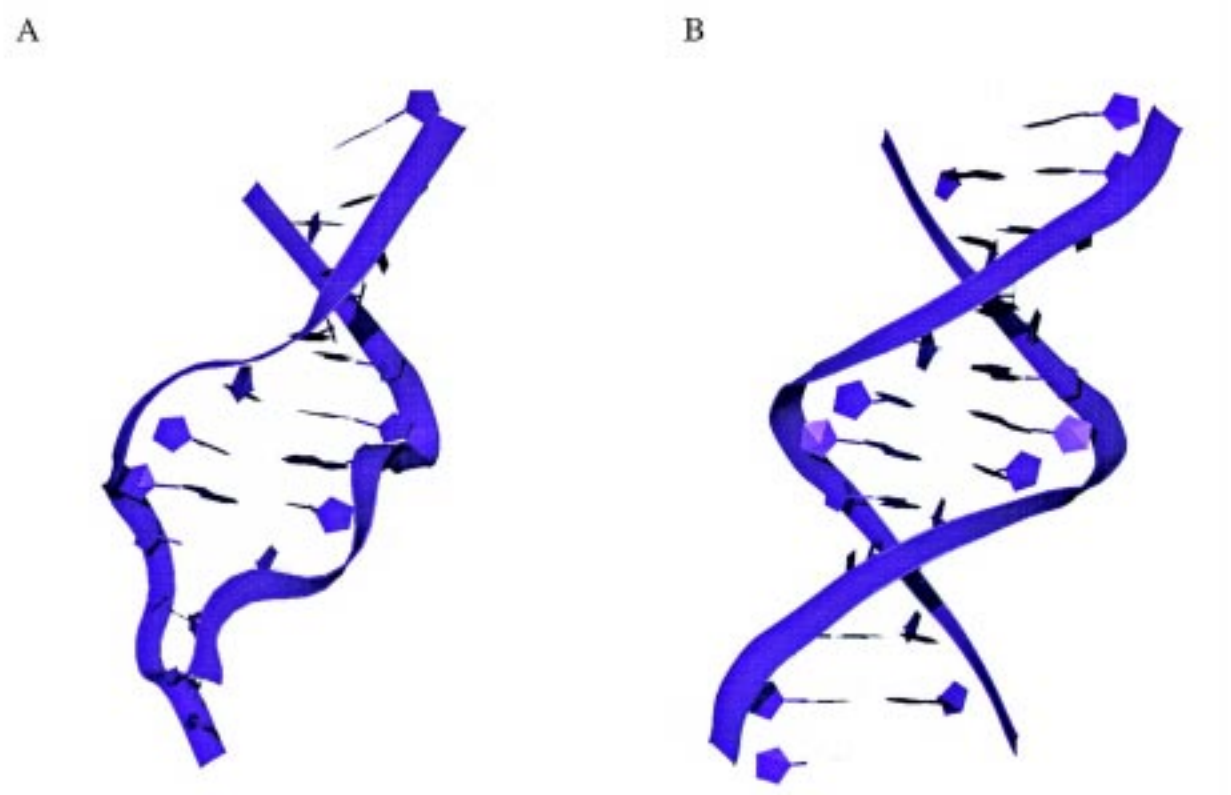

Fig. 4. Average structures from refinements of the DNA dodecamer (CGCGPATTCGCG) (Clore et al., 1988). No acceptable (i.e. without NOE violations) structures were generated by methods involving distance geometry or Cartesian molecular dynamics. In contrast, torsion angle molecular dynamics produced acceptable structures in 52% of the trials. It should be noted that the original structure was obtained by restrained molecular dynamics refinement, starting from canonical A- and B-form DNA. (a) Best structure (i.e. that with the smallest number of NOE violations) produced using 50 trials of Cartesian molecular dynamics simulated annealing starting from extended nucleotide strands. (b) Structure produced using torsion angle molecular dynamics starting from extended nucleotide strands.

molecular dynamics, convergence was achieved in about half of the trials. The method should be applicable to refinement of other nucleic acid structures which may adopt noncanonical structures.

16. Averaging of independently refined structures

As mentioned above, multiple simulated annealing refinements will generally produce somewhat different structures, some of which may be better (as assessed, for example, in terms of the free *R*-value for X-ray structures and NOE violations for NMR structures) than others. This approach offers several advantages. First, a more optimum structure can be obtained

from multiple trials as opposed to a single simulated annealing calculation. This is routinely done for NMR structure calculation where typically 20–100 trials are performed. Second, each member of the family of refined structures may be better in different regions of the molecule. Thus, by examining the ensemble during model-building, one may gain insights into possible local conformations of the molecule. Third, the structure factors of all structures of the family may be averaged in the X-ray crystallographic case. This averaging will reduce the effect of local errors (noise) that are presumably different in each member of the family.

The effect of averaging multiple simulated annealing trials is illustrated in Fig. 5. Each simulated annealing run employed torsion-angle molecular dynamics with the maximum likelihood target (Eq. (4)). The calculations were performed on human heterogeneous ribonucleoprotein A1, hnRNP (Shamoo et al., 1997). Averaging produced the least model-biased map (as indicated by the lowest free R-value and the lowest $R_{free}-R$ difference) with the polypeptide backbone being completely connected (Rice et al., 1998). This example is another demonstration that cross-validation of the R-value is essential for assessing model correctness (Brunger, 1992) since the normal R-value decreases with increasing model-bias of the electron density maps whereas the free R-value shows the correct behavior.

17. Ensemble models

In cases of conformational variability or discrete disorder, there is not a single correct solution to the optimization problem Eq. (1). Rather, the X-ray diffraction or NMR data represent a spatial and temporal average over all conformations that are assumed by the molecule. Ensembles of structures, which are simultaneously refined against the observed data, may thus be a more appropriate description of the data. This has been used for some time in X-ray crystallography when alternate conformations are modeled locally. Alternate conformations can be generalized to global conformations (Gros et al., 1990; Kuriyan et al., 1991; Burling and Brunger, 1994; Bonvin and Brunger, 1995), i.e. the model is duplicated n-fold, the corresponding calculated structure factors are added and refined simultaneously against the observed X-ray diffraction data or solution NMR data, and each member of the family is chemically 'invisible' to all other members. The number n can be determined by cross-validation in the crystallographic case (Burling and Brunger, 1994; Burling et al., 1996). However, this is difficult for solution NMR structures (Bonvin and Brunger, 1996) because of the unfavorable observable to parameter ratio.

An advantage of a multi-conformer model is that it directly incorporates many possible types of disorder and motion (global disorder, local sidechain disorder, local wagging and rocking motions). Furthermore, it can be used to automatically detect the most variable regions of the molecule by inspecting the atomic root-mean-square difference around the mean as a function of residue number. Thermal factors of single conformer models may sometimes be misleading by underestimating the degree of motion or disorder (Kuriyan et al., 1986) and, thus, the multiple-conformer model is a more faithful representation of the diffraction data. A disadvantage of the multi-conformer model is that it introduces many more degrees of

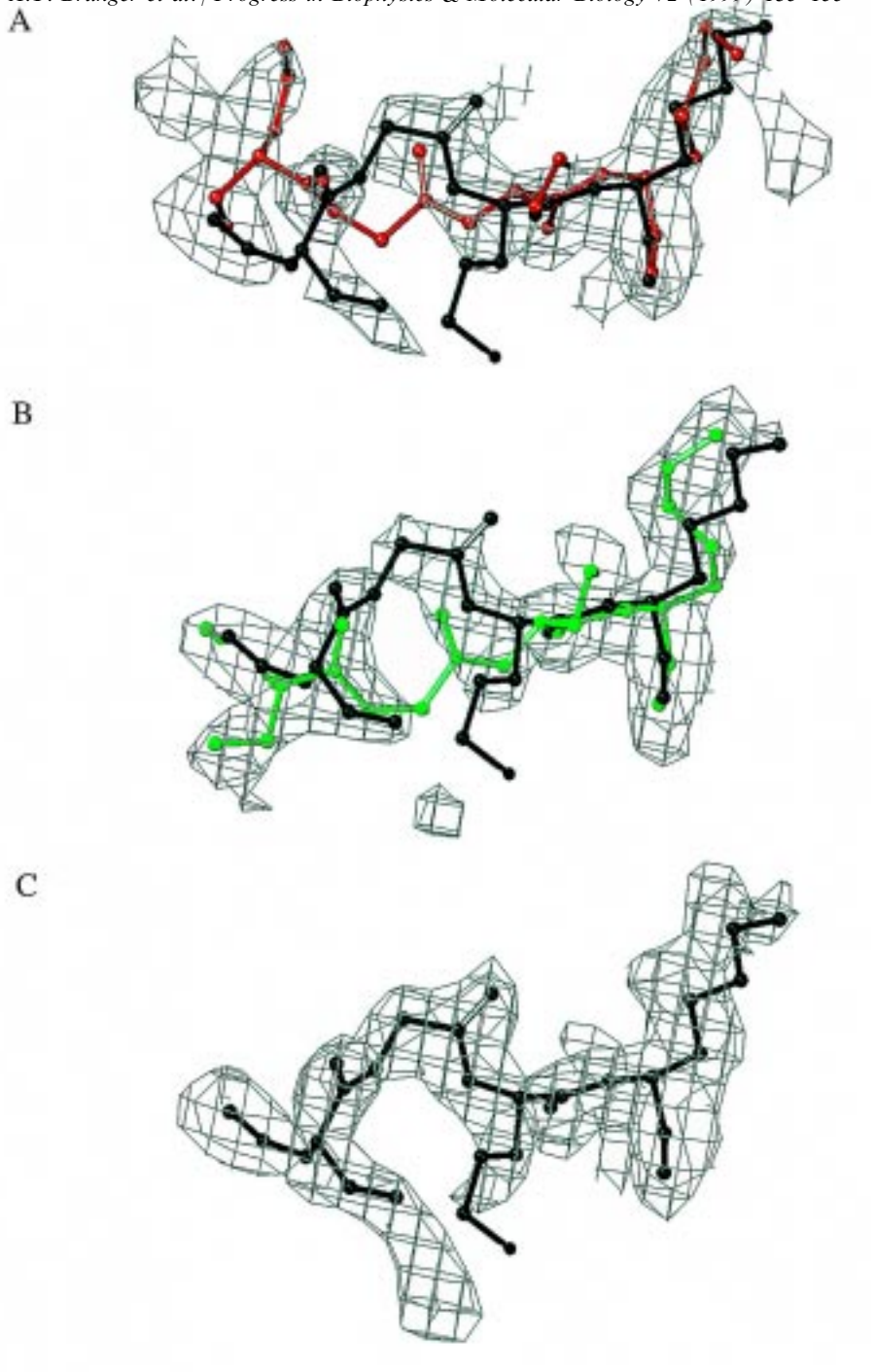

Fig. 5. Demonstration of the improvement provided by averaging several independently refined structures. Five torsion angle molecular dynamics simulated annealing refinements using the maximum-likelihood target were carried out for hnRNP (Shamoo et al., 1997; Rice et al., 1998). Cross-validated σ_A-weighted electron density maps, contoured at 1σ, are shown for (a) the typical result of refinement (red line, R = 35.7%, R_{free} = 41.3%), (b) the best result of the refinements (green line, R = 36.2%, R_{free} = 40.9%), and (c) the average map from averaging structure factors calculated from the four best models (R = 36.6%, R_{free} = 38.9% for the average F_c). In all cases the refined structure is shown in black. This figure was generated with the program O (Jones et al., 1991).

freedom. However, cross-validated maximum-likelihood refinement can address this problem. For example, the R_{free} and R-values were 0.239 and 0.237 for a single conformer refinement and 0.231 and 0.230, respectively, for a four-conformer refinement at 50–1.7 Å resolution data of a fragment of mannose-binding protein A (Burling et al., 1996) illustrating that introduction of multiple conformers did not increase the amount of overfitting compared to the single-conformer case (unpublished results).

Although there are some similarities between averaging individually refined structures and multi-conformer models, there are also fundamental differences. For example, in the case of X-ray crystallography, averaging seeks to improve the calculated electron density map by averaging out the noise present in the individual models, each of which is still a good representation of the diffraction data. This method is most useful at the early stages of refinement when the model still contains errors. In contrast, multi-conformer refinement seeks to create an ensemble of structures at the final stages of refinement which, taken together, best represent the data. It should be noted that each individual conformer of the ensemble does not necessarily remain a good description of the data since the whole ensemble is refined against the data. Clearly, this method requires high-quality data and a high observable-to-parameter ratio.

18. Conclusions

Simulated annealing has improved the efficiency of macromolecular structure calculation and refinement significantly in both X-ray crystallography and solution NMR spectroscopy. A case in point is the combination of torsion angle molecular dynamics with a cross-validated maximum-likelihood target for X-ray crystallography which interact synergistically to produce less model bias than any other method to date. The combined method also increases the radius of convergence allowing the refinement of poor initial models, e.g. those obtained by weak molecular replacement solutions (Rice and Brunger, 1994; Adams et al., 1999). However, simulated annealing refinement alone is still insufficient to refine a structure automatically without human intervention. For example, crystallographic refinement using simulated annealing typically cannot correct chain tracing errors such as register shifts. In the NMR case, it is sometimes necessary to correct misassignments or decide between different models that both appear to fit the experimental data equally well although ambiguous NOE restraints can help automate this process (Nilges, 1995). Fully automatic structure determination remains a distant goal that will probably require significant new algorithmic developments (Lamzin and Wilson, 1993; Perrakis et al., 1997).

Simulated annealing can also be used to provide new physical insights into molecular function which may depend on conformational variability. The sampling characteristics of simulated annealing allow the generation of multi-conformer models which can represent molecular motion and discrete disorder, especially when combined with the acquisition of high-quality data (both X-ray diffraction and solution NMR) (Burling et al., 1996). Simulated annealing is thus also a stepping stone towards development of improved models of macromolecules in solution and in the crystalline state.

Acknowledgements

Many of the recent computational developments discussed in this review are available in the program Crystallography&NMR System (Brunger, Adams, Clore, DeLano, Gros, Grosse-Kunstleve, Jiang, Kuszewski, Nilges, Pannu, Read, Rice, Simonson and Warren; URL: http//atb.csb.yale.edu). We thank Dr. Michael Nilges and Dr. Gregory Warren for critical reading of the manuscript. LMR is an HHMI predoctoral fellow. This work was funded in part by grants from the National Science Foundation to ATB (BIR 9514819 and ASC 93-181159).

References

Abagyan, R., Argos, P., 1992. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J. Mol. Biol. 225, 519–532.

Abramowitz, M., Stegun, I., 1968. Handbook of Mathematical Functions, Applied Mathematics Series, vol. 55. Dover, New York, p. 896.

Adams, P.D., Pannu, N.S., Read, R.J., Brunger, A.T., 1997. Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. Proc. Natl. Acad. Sci. USA 94, 5018–5023.

Adams, P.D., Pannu, N.S., Read, R.J., Brunger, A.T., 1999. Extending the limits of molecular replacement through combined simulated annealing and maximum likelihood refinement. Acta Cryst. D 55, 181–190.

Allen, F.H., Kennard, O., Taylor, R., 1983. Systematic analysis of structural data as a research technique in organic chemistry. Acc. Chem. Res. 16, 146–153.

Bae, D.-S., Haug, E.J., 1987. A recursive formulation for constrained mechanical system dynamics. I. Open loop systems. Mech. Struct. Mach. 15, 359–382.

Bae, D.-S., Haug, E.J., 1988. A recursive formulation for constrained mechanical system dynamics. II. Open loop systems. Mech. Struct. Mach. 15, 481–506.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684–3690.

Bonvin, A.M.J.J., Brunger, A.T., 1995. Conformational variability of solution nuclear magnetic resonance structures. J. Mol. Biol. 250, 80–93.

Bonvin, A.M.J.J., Brunger, A.T., 1996. Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? J. Biomol. NMR 7, 72–76.

Braun, W., 1987. Distance geometry and related methods for protein structure determination from NMR data. Q. Rev. Biophys. 19, 115–157.

Braun, W., Go, N., 1985. Calculation of protein conformations by proton–proton distance constraints. A new efficient algorithm. J. Mol. Biol. 186, 611–626.

Bricogne, G., 1991. A multisolution method of phase determination by combined maximization of entropy and likelihood. III. Extension to powder diffraction data . Acta Cryst. A 47, 803–829.

Bricogne, G., 1993. Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. Acta Cryst. D 49, 37–60.

Bricogne, G., 1997. Bayesian statistical viewpoint on structure determination: basic concepts and examples. Meth. Enzy. 276, 361–423.

Brunger, A.T., 1988. Crystallographic refinement by simulated annealing: application to a 2.8 Å resolution structure of aspartate aminotransferase. J. Mol. Biol. 203, 803–816.

Brunger, A.T., 1992. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355, 472–474.

Brunger, A.T., Clore, G.M., Gronenborn, A.M., Karplus, M., 1986. Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin. Proc. Natl. Acad. Sci. USA 83, 3801–3805.

Brunger, A.T., Kuriyan, J., Karplus, M., 1987. Crystallographic *R* factor refinement by molecular dynamics. Science 235, 458–460.

Brunger, A.T., Karplus, M., Petsko, G.A., 1989. Crystallographic refinement by simulated annealing: application to a 1.5 Å resolution structure of crambin. Acta Cryst. A 45, 50–61.

Brunger, A.T., Krukowski, A., Erickson, J., 1990. Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Cryst. A 46, 585–593.

Brunger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R., Nilges, M., 1993. Assessment of the quality of solution nuclear magnetic resonance structures by complete cross-validation. Science 261, 328–331.

Burling, F.T., Brunger, A.T., 1994. Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. Israel J. Chem. 34, 165–175.

Burling, F.T., Weis, W.I., Flaherty, K.M., Brunger, A.T., 1996. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. Science 271, 72–77.

Clore, G.M., Nilges, M., Sukumaran, D.K., Brunger, A.T., Karplus, M., Gronenborn, A.M., 1986. The three-dimensional structure of a-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. EMBO J. 5, 2729–2735.

Clore, G.M., Oschkinat, H., McLaughlin, L.W., Benseler, F., Scalfi-Happ, C., Happ, E., Gronenborn, A.M., 1988. Refinement of the solution structure of the DNA dodecamer 5′d(CGCGPATTCGCG)2 containing a stable purine–thymine base pair: combined use of nuclear magnetic resonance and restrained molecular dynamics. Biochemistry 27, 4185–4197.

Curro, J., 1974. Computer simulation of multiple chain systems – the effect of density on the average chain dimension. J. Chem. Phys. 61, 1203–1207.

Dauter, Z., Lamzin, V.S., Wilson, K.S., 1995. Proteins at atomic resolution. Curr. Opin. Struct. Biol. 5, 784–790.

Diamond, R., 1971. A real-space refinement procedure for proteins. Acta Cryst. A 27, 436–452.

Engh, R.A., Huber, R., 1991. Accurate bond and angle parameters for X-ray structure refinement. Acta Cryst. A 47, 392–400.

Fujinaga, M., Gros, P., van Gunsteren, W.F., 1989. Testing the method of crystallographic refinement using molecular dynamics. J. Appl. Cryst. 22, 1–8.

Garrett, D.S., Kuszewski, J., Hancock, T.J., Lodi, P.J., Vuister G, W., Gronenborn, A.M., Clore, G.M., 1994. The impact of direct refinement against three-bond HN–C alpha H coupling constants on protein structure determination by NMR. J. Magn. Res. B 104, 99–103.

Goldstein, H., 1980. Classical Mechanics, 2nd ed. Addison-Wesley, Reading, MA.

Gros, P., van Gunsteren, W.F., Hol, W.G.J., 1990. Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. Science 249, 1149–1152.

Hendrickson, W.A., 1985. Stereochemically restrained refinement of macromolecular structures. Meth. Enzymol. 115, 252–270.

Hendrickson, W.A., 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254, 51–58.

Hoppe, W., 1957. Die Faltmolekülmethode – eine neue Methode zur Bestimmung der Kristallstruktur bei Ganz oder Teilweise bekannter Molekülstruktur. Acta Cryst. 10, 750–751.

Hsu, I.N., Delbaere, L.T.J., James, M.N.G., Hoffman, T., 1977. Penicillopepsin from *Penicillium janthinellum* crystal structure at 2.8 Å and sequence homology with porcine pepsin. Nature 266, 140–145.

Jain, A., Vaidehi, N., Rodriguez, G., 1983. A fast recursive algorithm for molecular dynamics simulation. J. Comp. Phys. 106, 258–268.

Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M., 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst. A 47, 110–119.

Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., van Gunsteren, W.F., 1985. A protein structure from nuclear magnetic resonance data. lac repressor headpiece. J. Mol. Biol. 182, 179–182.

Karplus, M., 1963. Vicinal proton coupling in nuclear magnetic resonance. J. Am. Chem. Soc. 85, 2870–2871.

Karplus, M., Petsko, G.A., 1990. Molecular dynamics simulations in biology. Nature 347, 631–639.

Kim, Y., Prestegard, J.H., 1990. Refinement of the NMR structures for acyl carrier protein with scalar coupling data. Proteins Structure Function Genetics 8, 377–385.

Kirkpatrick, S., Gelatt, C.D., Vecchi Jr., M.P., 1983. Optimization by simulated annealing. Science 220, 671–680.

Kleywegt, G.J., Brunger, A.T., 1996. Cross-validation in crystallography: practice and applications. Structure 4, 897–904.

Kuriyan, J., Petsko, G.A., Levy, R.M., Karplus, M., 1986. Effect of anisotropy and anharmonicity on protein crystallographic refinement. J. Mol. Biol. 190, 227–254.

Kuriyan, J., Ösapay, K., Burley, S.K., Brunger, A.T., Hendrickson, W.A., Karplus, M., 1991. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins 10, 340–358.

Kuszewski, J., Gronenborn, A.M., Clore, G.M., 1995a. The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. J. Magn. Res. B 107, 293–297.

Kuszewski, J., Qin, J., Gronenborn, A.M., Clore, G.M., 1995b. The impact of direct refinement against 13C alpha and 13C beta chemical shifts on protein structure determination by NMR. J. Magn. Res. B 106, 92–96.

Kuszewski, J., Gronenborn, A.M., Clore, G.M., 1996. Improving the quality of NMR and crystallographic structures by means of a conformational potential derived from structure databases. Prot. Sci. 5, 1067–1080.

Laarhoven, P.J.M., Aarts, E.H.L. (Eds.), 1987. Simulated Annealing: Theory and Applications. Reidel, Dordrecht, p. 187.

Lamzin, V.S., Wilson, K.S., 1993. Automated refinement of protein models. Acta Cryst. D 49, 129–147.

Li, Z., Scheraga, H.A., 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc. Natl. Acad. Sci. USA 84, 6611–6615.

Mathiowetz, A.M., Jain, A., Karasawa, N., Goddard, W.A., 1994. Protein simulations using techniques suitable for very large systems: the cell multipole method for nonbond interactions and the Newton–Euler inverse mass operator method for internal coordinate dynamics. Proteins Structure Function Genetics 20, 227–247.

Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Mierke, D.F., Huber, T., Kessler, H.J., 1994. Coupling constants again: experimental restraints in structure refinement. Comput. Aided Mol. Des. 8, 29–40.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. D 53, 240–255.

Newman, M., Watson, F., Roychowdhury, P., Jones, H., Badasso, M., Cleasby, A., Wood, S.P., Tickle, I.J., Blundell, T.L., 1993. X-ray analyses of aspartic proteinases. V. Structure and refinement at 2.0 Å resolution of the aspartic proteinase from *Mucor pusillus*. J. Mol. Biol. 230, 260–283.

Nilges, M., 1995. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulfide connectivities. J. Mol. Biol. 245, 645–660.

Nilges, M., 1996. Structure calculation from NMR data. Curr. Opin. Struct. Biol. 6, 617–623.

Nilges, M., Clore, G.M., Gronenborn, A.M., 1988a. Determination of three dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. FEBS Lett. 239, 129–136.

Nilges, M., Clore, G.M., Gronenborn, A.M., 1988b. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. FEBS Lett. 229, 317–324.

Nilges, M., Gronenborn, A.M., Brunger, A.T., Clore, G.M., 1988c. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng. 2, 27–38.

Nilges, M., Kuszewski, J., Brunger, A.T., 1991. In: Hoch, J.C., Poulsen, F.M., Redfield, C. (Eds.), Sampling Properties of Simulated Annealing and Distance Geometry. Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy. Plenum Press, New York, pp. 451–455.

Oldfield, E., 1995. Chemical shifts and three-dimensional protein structures. J. Biomol. NMR 5, 217–225.

Pannu, N.S., Read, R.J., 1996. Improved structure refinement through maximum likelihood. Acta Cryst. A 52, 659–668.

Pannu, N.S., Murshudov, G.N., Dodson, E.J., Read, R.J., 1998. Incorporation of prior phase information strengthens maximum likelihood structural refinement. Acta Cryst. D 54, 1285–1294.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A.T., Berman, H.M., 1996. New parameters for the refinement of nucleic acid containing structures. Acta Cryst. D 52, 57–64.

Pearlman, D.A., Kim, S.-H., 1990. Atomic charges for DNA constituents derived from single-crystal X-ray diffraction data. J. Mol. Biol. 211, 171–187.

Perrakis, A., Sixma, T.K., Wilson, K.S., Lamzin, V.S., 1997. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Cryst. D 53, 448–455.

Press, W.H., Flannery, B.P., Teukolosky, S.A., Vetterling, W.T., 1986. In: Numerical Recipes. Cambridge University Press, Cambridge, pp. 498–546.

Read, R.J., 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Cryst. A 42, 140–149.

Read, R.J., 1990. Structure-factor probabilities for related structures. Acta Cryst. A 46, 900–912.

Rice, L.M., Brunger, A.T., 1994. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins Structure Function Genetics 19, 277–290.

Rice, L.M., Shamoo, Y., Brunger, A.T., 1998. Phase improvement by multi-start simulated annealing refinement and structure factor averaging. J. Appl. Cryst. 31, 798–805.

Rossmann, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst. A 15, 24–51.

Saunders, M., 1987. Stochastic exploration of molecular mechanics energy surfaces: hunting for the global minimum. J. Am. Chem. Soc. 109, 3150–3152.

Sevcik, J., Dauter, Z., Lamzin, V.S., Wilson, K.S., 1996. Ribonuclease from Streptomyces aureofaciens at atomic resolution. Acta Cryst. D 52, 327–344.

Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R., Steitz, T.A., 1997. Crystal structure of the two RNA-binding domains of human hnRNP A1 at 1.75 Å resolution. Nat. Struct. Biol. 3, 215–222.

Silva, A.M., Rossmann, M.G., 1985. The refinement of southern bean mosaic virus in reciprocal space. Acta Cryst. B 41, 147–157.

Stec, B., Zhou, R., Teeter, M.M., 1995. Full-matrix refinement of the protein crambin at 0.83 Å and 130 K. Acta Cryst. D 51, 663–681.

Stein, E.G., Rice, L.M., Brunger, A.T., 1997. Torsion angle molecular dynamics is a new, efficient tool for NMR structure calculation. J. Magn. Reson. B 124, 154–164.

Sussman, J.L., Holbrook, S.R., Church, G.M., Kim, S.-H., 1977. Structure-factor least-squares refinement procedure for macromolecular structure using constrained and restrained parameters. Acta Cryst. A 33, 800–804.

Ulyanov, N.B., Schmitz, U., James, T.L., 1993. Metropolis Monte Carlo calculations of DNA structure using internal coordinates and NMR distance restraints: an alternative method for generating a high-resolution solution structure. J. Biomol. NMR 3, 547–568.

Verlet, L., 1967. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. 159, 98–105.

Vlassi, M., Dauter, Z., Wilson, K.S., Kokkinidis, M., 1998. Structural parameters for proteins derived from the atomic resolution (1.09 Å) structure of a designed variant of the colE1 ROP protein. Acta Cryst. D 54, 1245–1260.

Weis, W.I., Brunger, A.T., Skehel, J.J., Wiley, D.C., 1989. Refinement of the influenza virus haemagglutinin by simulated annealing. J. Mol. Biol. 212, 737–761.

Wüthrich, K., 1986. NMR of Proteins and Nucleic Acids. Wiley, New York.

Xu, Y., Krishna, N.R., 1995. Structure determination from NOESY intensities using a metropolis simulated-annealing (MSA) refinement of dihedral angles. J. Mag. Res. B 108, 192–196.