

New applications of simulated annealing in X-ray crystallography and solution NMR

Axel T Brünger*, Paul D Adams and Luke M Rice

Address: The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

*Corresponding author.
E-mail: brunger@laplace.csb.yale.edu

Electronic identifier: 0969-2126-005-00325

Structure 15 March 1997, 5:325–336

© Current Biology Ltd ISSN 0969-2126

Introduction

Over the past 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 culminating in refinement and structure validation. These procedures can be formulated as the chemically-constrained or restrained non-linear 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 [1]. 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 [2,3] is an optimization technique particularly well suited to overcoming the multiple minima problem. Unlike gradient-descent methods, 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 [4], and NMR structure calculation [5] and refinement [6], 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) [7] 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 [8], 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 [9,10]. Advances have also been made for NMR structure determination. Cross-validation has been shown to be appropriate for NMR structure calculation [11], and the radius of convergence has been increased by the use of torsion-angle molecular dynamics [12] and the introduction of variable target functions [13–16]. Finally, the sampling power of simulated annealing can be combined with multiconformer models for exploring the molecule's conformational space in cases where the molecule undergoes dynamic motion or static disorder for both X-ray crystallography and solution NMR [17–21].

Setting the stage: the target function

In essence, macromolecular structure calculation and refinement is a search for the global minimum of a target function (E) which is a function of the parameters of an atomic model, in particular atomic coordinates,

$$E = E_{\text{chem}} + w_{\text{data}} E_{\text{data}} \quad (1)$$

E_{chem} comprises empirical information about chemical interactions; it is a function of all atomic positions, describing covalent and nonbonded (intramolecular as well as intermolecular 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.

A priori chemical information

E_{chem} consists of terms for covalent bonds, bond angles, chirality, planarity and nonbonded repulsion [22]. The parameters for the covalent terms can be derived from average geometry and root mean square (rms) deviations observed in a small molecule database. Extensive statistical analyses were undertaken for the chemical moieties of proteins [23] and of polynucleotides [24] using the

Cambridge Crystallographic Database [25]. Analysis of the ever increasing number of atomic resolution macromolecular crystal structures will no doubt lead to some modifications of these parameters in the future [26–28].

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} [22],

$$E_{\text{repulsive}} = \sum_{ij} ((cR_{ij}^{\text{min}})^n - R_{ij}^n)^m \quad (2)$$

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 , $c \leq 1$ is a constant that is sometimes used to reduce the radii, and $n=2$, $m=2$ or $n=1$, $m=4$. Electrostatic interactions and van der Waals attraction are usually not included in structure calculation and refinement. These simplifications are valid as 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 [29]. 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 the empirical energy functions that were developed for energy minimization and molecular dynamics studies of macromolecules (see [30] for an introduction). These empirical energy functions were not designed for structure determination, and therefore required some modification for use in macromolecular structure refinement [5,31–35]. Recently, crystallographic simulated-annealing refinement was implemented with a purely geometric energy function [10], 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).

X-ray diffraction data

The conventional form of E_{xray} consists of the crystallographic residual E^{LSQ} , defined as the sum over the squared differences between the observed (F_o) and calculated (F_c) structure-factor amplitudes for a particular atomic model:

$$E_{\text{xray}} = E^{\text{LSQ}} = \sum_{hkl} (|F_o| - k|F_c|)^2 \quad (3)$$

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 improvements in the atomic model, but also from an accumulation of systematic errors in the model or fitting noise in the data [36]. The

least-squares residual is, therefore, poorly justified when the model is incomplete or far away from the correct one [9]. An improved target for macromolecular refinement can be obtained using a maximum-likelihood formulation [37–39,9,10]. 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 estimates of the errors 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 Equation 3 [7,40]). The expected values of $\langle F_o \rangle$ and the corresponding variance (σ_{ML}^2) are derived from σ_A , the observed (F_o), and calculated (F_c) structure-factor amplitudes [9]. These quantities can be readily incorporated into a maximum-likelihood target function:

$$E_{\text{xray}} = E^{ML} = \sum_{hkl \in \text{working set}} (1/\sigma_{ML}^2) (|F_o| - \langle F_o \rangle)^2 \quad (4)$$

In order to achieve an improvement over the least-squares residual (Equation 3), cross-validation was found to be essential [10,40] for the computation of σ_A and its derived quantities in Equation 4.

Figure 1 illustrates the improvement of the crystallographic target function by maximum likelihood for a test case: the false minimum is less pronounced compared to the least-squares residual. Despite this major improvement, local minima still exist (Fig. 1) which cause a limited radius of convergence when gradient-descent minimization is used.

NMR data

A common form of E_{NMR} describes nuclear Overhauser effect (NOE) derived distance restraints and dihedral angle restraints derived from J-coupling constants using flat-bottomed parabolic functions [41]:

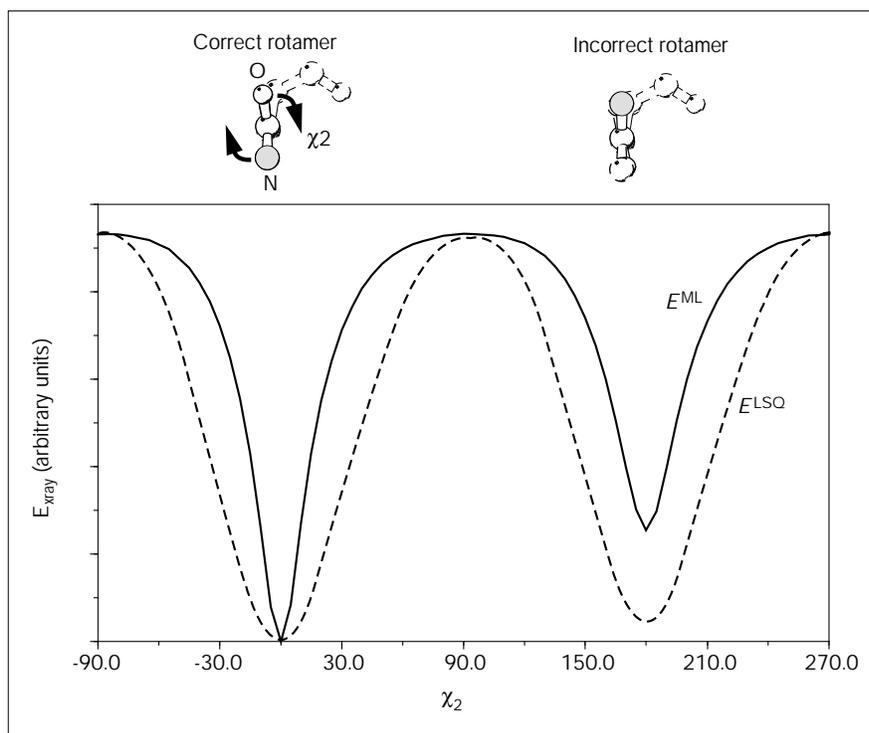
$$\omega_{\text{nmr}} E_{\text{nmr}} = \omega_{\text{NOE}} E_{\text{NOE}} + \omega_{\text{dihedral}} E_{\text{dihedral}} \quad (5)$$

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

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}} \begin{cases} (\phi - \phi_{\text{upper}})^2 & \phi_{\text{upper}} < \phi \\ 0 & \phi_{\text{lower}} < \phi < \phi_{\text{upper}} \\ (\phi_{\text{lower}} - \phi)^2 & \phi < \phi_{\text{lower}} \end{cases} \quad (7)$$

Figure 1

A comparison of the energy landscapes of the least-squares residual (Equation 3, dotted line) and the maximum-likelihood target (Equation 4, solid line). An asparagine residue was placed in a P1 unit cell of size $a = 40 \text{ \AA}$, $b = 40 \text{ \AA}$, $c = 40 \text{ \AA}$. Diffraction data from 20 to 2 \AA resolution were calculated from the structure in its initial conformation; this data was modified by addition of Gaussian error ($\pm 10\%$ of $|F|$). A single value of σ_F , 10% of the average amplitude, was used for all reflections. The χ_2 dihedral was then rotated in 5° steps and the value of the X-ray target (Equations 3 and 4) compared to the model diffraction data calculated at each position. The two minima correspond to the correct and incorrect rotamers with the positions of OD1 and ND2 inverted. The maximum-likelihood target clearly shows sharper minima, broader, flatter maxima, and a less deep false minimum relative to the global minimum.



Here d denotes the distance between a particular pair of spins in the model, d_{lower} and d_{upper} are, respectively, the lower and upper bounds for the distance restraint derived from the isolated spin-pair approximation or from NOE backcalculation (see [42] for a review), ϕ denotes the dihedral angle formed by four atoms in the model, ϕ_{lower} and ϕ_{upper} are, respectively, the lower and upper bounds for the dihedral-angle restraint derived from scalar J-coupling constant measurements and empirical Karplus [43] relationships. In the case of ambiguous NOE assignments, overlapping NOEs, or motion of methyl groups and aromatic rings, appropriate averaging schemes must be used [42,44,45] when computing d , d_{upper} and d_{lower} . Direct refinement against scalar J-coupling constant data [46–48], or against empirical ^1H and ^{13}C chemical shift databases which correlate molecular conformation and chemical shifts [49–51] 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 (‘softsquare’ potential [52]). This modification is important for convergence — without it the large violations give rise to enormous forces which cause numerical instabilities during simulated annealing.

Additional information

Additional constraints or restraints may be used to improve the ratio of observed data to refined 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 [8,53,54]. In the crystallographic case, the existence of noncrystallographic symmetry can be used to average equivalent molecules, and thereby to reduce noise in the diffraction data [33]. 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 modelled 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 [55].

Phase information can be incorporated into E_{xray} in order to improve the ratio of observables to parameters and, thus, to obtain a more accurate model especially when only low-resolution diffraction data is available [8]. As the experimental phase information for macromolecules is usually not very accurate, the errors of the phase observations have to be taken into account. Ideally, this can be accomplished by computing a phase probability distribution for each reflection and by refining the model phases against these distributions (ATB, unpublished results). Several approximations to this computationally expensive procedure have been developed. A simple representation of the phase probability distribution consists of a square-well potential around the phase centroid, where the width

of the square-well is determined by the individual figure of merit [56]. For well defined phases, for example those obtained from highly redundant virus crystal structures, structures solved by multiple wavelength anomalous dispersion (MAD) phasing [57], or good multiple isomorphous replacement (MIR) solutions, one can restrain the real and imaginary parts, A and B, of the structure factor simultaneously [58,8]. When phase information is incomplete, as is often the case, the phase-restraint target can be combined with an amplitude-based target using figure of merit weighting [59] (LMR, Y Shamoo and AJB, unpublished data).

Weighting

The weight (w_{data} ; Equation 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 [32,10]. However, independent information must be used (e.g. cross-validation) to objectively obtain the best possible weight for X-ray diffraction [7] and NMR data [16].

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 (equation 1) to be the equivalent of the potential energy of the system, one can simulate the annealing process [2]. There is no guarantee that simulated annealing will find the global minimum (except in the case of an infinitely long search) [3]. Compared to conjugate-gradient minimization, where search directions must follow the gradient, simulated annealing achieves more optimal solutions by allowing motion against the gradient [2]. 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 $T_1 \geq T_2 \geq \dots T_j$ at which the Boltzmann distribution is computed. Implementations of the generation mechanism differ in the way they generate a transition, or move, 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 [60] and molecular dynamics [61] simulations. For NMR structure calculation, both molecular dynamics and Monte Carlo have been successfully used [6,5,62–64]. For X-ray crystallographic refinement, molecular dynamics has proven to be extremely successful [4], whereas the Monte Carlo methods have yet to be shown to be effective.

Metropolis Monte Carlo

The Metropolis Monte Carlo algorithm [60] 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 using a suitably chosen set of internal coordinates, such as torsions about bonds, normal modes of vibration, or by relaxing the strained coordinates through minimization [65–67].

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 [68]. 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:

$$m_i(\partial^2 \vec{r}_i / \partial t^2) = -\nabla_i E \quad (8)$$

The quantities m_i and \vec{r}_i are the mass and coordinates of atom i , respectively, and E is given by Equation 1. The solution of the partial differential equations (Equation 8) is achieved numerically using finite difference methods [61]. This approach is referred to as molecular dynamics [61].

Initial velocities for the integration of Equation 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} , Equation 1), one might want to implement these restrictions as constraints (i.e. fixed bond lengths and bond angles [54]). 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 [54,11], and by Monte Carlo minimization [62–64]. It is only recently, however, that efficient and robust algorithms have become available for molecular dynamics in torsion-angle space [8,69–72].

We chose an approach that retains the Cartesian-coordinate formulation of the target function and its derivatives with respect to atomic coordinates so that calculation remains relatively straightforward and topology independent [8]. 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 [73]. 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 [8,15].

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_{kin}):

$$E_{\text{kin}} = \sum_i^{\text{atoms}} 1/2 m_i (\partial r_i / \partial t)^2 \quad (9)$$

of the molecular dynamics simulation,

$$T_{\text{curr}} = 2E_{\text{kin}} / 3nk_b \quad (10)$$

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) [74] can be added to the right-hand side of Equation 8 to lower the temperature:

$$-m_i \gamma_i v_i (1 - (T/T_{\text{curr}})) \quad (11)$$

where v_i is the velocity of atom i and m_i is the mass. This method generalizes the concept of friction by determining

the friction coefficient and its sign by the ratio of the current simulation temperature (T_{curr}) to the target temperature T .

Why does simulated annealing work?

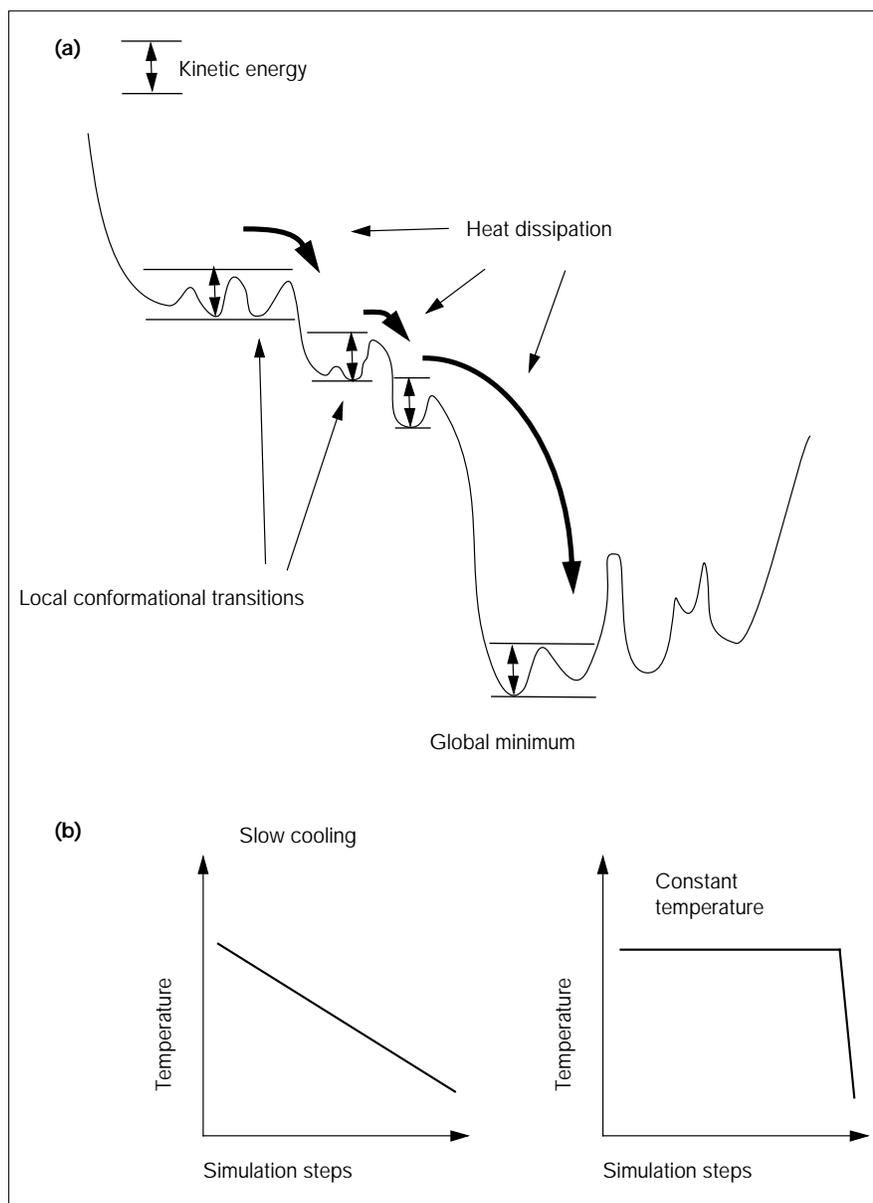
The goal of any optimization problem is to find the global minimum of a target function. In the case of macromolecular structure calculation and refinement, one searches for the conformation, or conformations, of the molecule that best fit the experimental data and that simultaneously maintain reasonable 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 (Equation 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 (Equation 10). The target temperature must be large enough to overcome smaller barriers (e.g. Fig. 2) 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 temperature coupling (Equation 11) has removed this excess kinetic energy through 'heat dissipation', the reverse transition is very unlikely, as 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.

Practical considerations

As Figure 2a illustrates, the simulation temperature needs to be high enough to allow conformational transitions but not too high so as to avoid moving too far away from the initial structure. The optimal temperature for a given starting structure is a matter of trial and error. We empirically

Figure 2



Simulated annealing. (a) Schematic explanation of molecular dynamics-based simulated annealing. The kinetic energy of the system allows local conformational transitions with barriers smaller than the kinetic energy. If a larger drop in energy is encountered the excess kinetic energy is dissipated through the friction term (Equation 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.

determined starting temperatures for a variety of simulated-annealing protocols [10,56], 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 calculation 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. 2b). A slight advantage in terms of final R values is obtained with slow cooling [34]. 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 as it is more efficient to run multiple trials as opposed to one

long refinement protocol (PDA, LMR and ATB, unpublished results).

Crystallographic refinement

In the case of crystallographic refinement, the limited radius of convergence 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 [75,76] 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 recent tests [8,10], arbitrarily 'scrambled' models were generated from an initial model of α -amylase inhibitor built using experimental phase information from MIR. Scrambling of this initial model was obtained by using increasingly long molecular dynamics simulations at 600K, computed without reference to the X-ray diffraction data. These scrambled structures were energy minimized in order to provide chemically reasonable starting models. It was shown that these models represented realistic test cases for phasing by molecular replacement: the scrambled structures contained sufficient information to solve the crystal structure through an extensive molecular replacement search [10].

Both conjugate-gradient minimization and simulated annealing were carried out in order to compare the performance of the maximum-likelihood target against the least-squares residual. Figure 3 shows that the maximum-likelihood target produces models with lower phase error. For conjugate-gradient minimization the average phase improvement, compared to the least-squares residual, is approximately 5° . It can also be seen that simulated annealing with the least-squares target has a larger radius of convergence than minimization with maximum likelihood, and that this convergence is further improved by the use of the maximum-likelihood target (Fig. 3). For the most scrambled model, an average phase improvement of more than 15° is obtained, compared to the least-squares residual. The resulting structures are very close to the published crystal structure (Fig. 3). Similar results were obtained for a test case at lower (2.8\AA) resolution and for a

real case involving a new crystal structure starting with an initial model built into a poor electron-density map [10].

Cross-validation is essential in the calculation of the maximum-likelihood target [9,10,40]. Maximum-likelihood refinement without cross-validation gives much poorer results, as indicated by higher free R values, higher $R_{\text{free}}-R$ differences, and higher phase errors [10]. It should be noted that the normal R value in general 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).

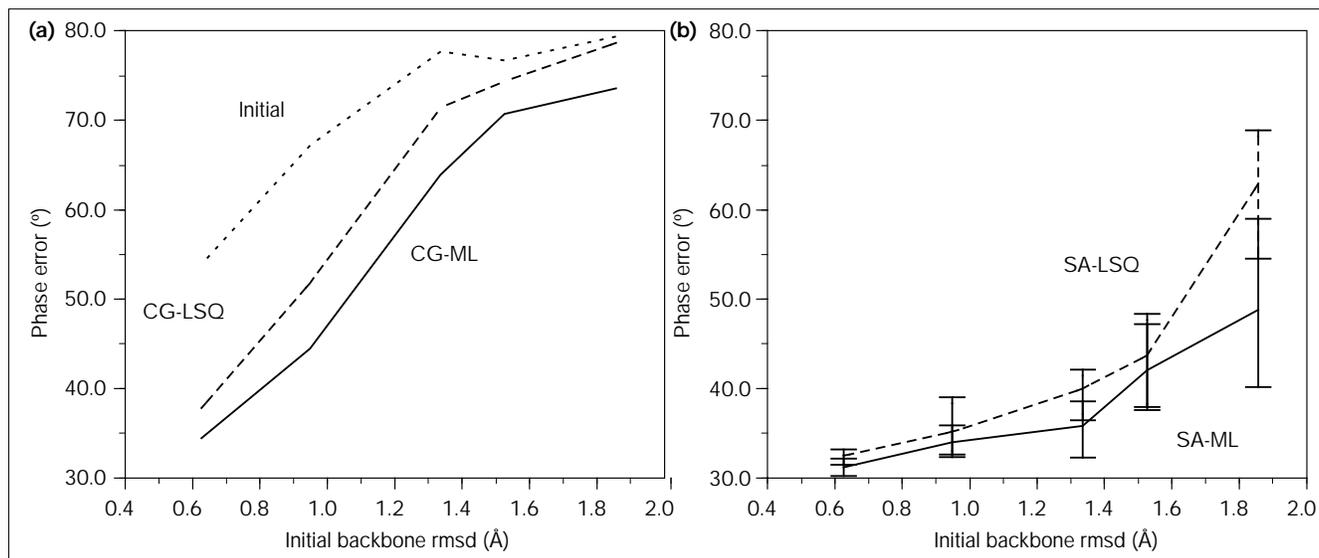
NMR structure calculation

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 interatomic 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 [13]. Furthermore, molecular dynamics-based simulated annealing can be used to automate the NOE assignment process [45]. 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 [15]. 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 the 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 [14] failed to produce the correct structure whereas

Figure 3



The radius of convergence. The final phase error for the least-squares residual target, Equation 3 (dashed line) and the maximum-likelihood target, Equation 4 (solid line) for refinements of the scrambled α -amylase inhibitor structures against diffraction data at $d_{\min} = 2.0 \text{ \AA}$, with respect to the published crystal structure [80]. (a) The results for conjugate gradient minimization for the least-squares target (CG-LSQ) and the maximum-likelihood target (CG-ML). (b) The results for ten

simulated-annealing runs with torsion-angle dynamics with the least-squares target (SA-LSQ) and the maximum-likelihood target (SA-ML). The simulated-annealing graphs shows the average of ten refinements with different initial velocities. Error bars indicate the maximum and minimum phase errors obtained with ten simulated-annealing runs. The best structures are readily identified by the lowest free R value. Initial phase errors are shown by the upper dotted line (marked 'initial') in (a).

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

Averaging of independently refined structures

As mentioned above, multiple simulated-annealing refinements will generally produce somewhat different structures, some of which may be better than others (e.g. as assessed in terms of the free R value for X-ray structures, and NOE violations for NMR structures). This approach offers several advantages. Firstly, a more optimal 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. Secondly, 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. Thirdly, the structure factors of all structures of the family may be averaged in the X-ray crystallographic case [8]. 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 Figure 5. Each simulated-annealing run

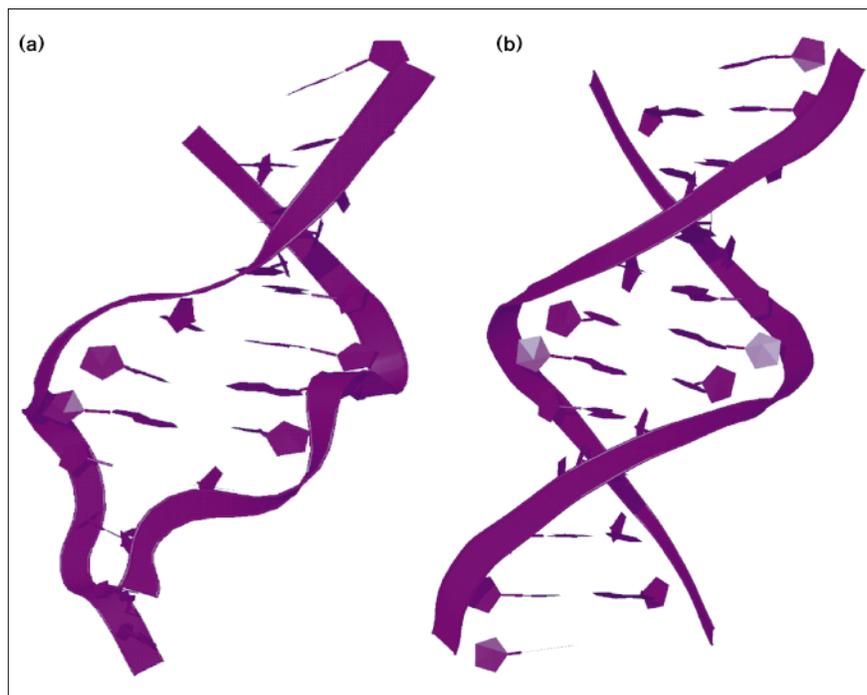
employed torsion-angle molecular dynamics with the maximum-likelihood target (Equation 4). The calculations were performed on human heterogeneous ribonucleoprotein A1 (hnRNP), a recently solved crystal structure [77]. Averaging produced the least model-biased map (as indicated by the lowest free R value and the lowest $R_{\text{free}} - R$ difference) with the polypeptide backbone being completely connected (LMR, Y Shamoo, ATB, unpublished results). This example is another demonstration that cross-validation of the R value is essential for assessing model correctness [7], as the normal R value decreases with increasing model bias of the electron-density maps whereas the free R value shows the correct behavior.

Ensemble models

In cases of conformational variability or discrete disorder, there is not a single correct solution to the optimization problem, Equation 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 modelled locally. Alternate conformations can be generalized to global conformations [17,18,20,78], that is the model is duplicated n -fold, the corresponding calculated structure

Figure 4

Average structures obtained from refinements of the DNA dodecamer (CGCGPATTGCG) [81]. No acceptable structures (i.e. without NOE violations) 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. The original structure [80] was obtained by restrained molecular dynamics refinement, starting from canonical A-form and B-form DNA. **(a)** The 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)** The structure produced using torsion-angle molecular dynamics starting from extended nucleotide strands.



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 [18,19]. However, this is difficult for solution NMR structures [21] because of the unfavorable observable to parameter ratio.

An advantage of a multiconformer 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 rms 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 [79] and, thus, the multiconformer model is a more faithful representation of the diffraction data. A disadvantage of the multiconformer model is that it introduces many more degrees of freedom. However, cross-validated maximum-likelihood refinement can reduce the danger of overfitting. For example, introduction of multiple conformers into refinement of a fragment of mannose-binding protein did not increase the amount of overfitting compared to the single-conformer case (FT Burling and ATB, unpublished results).

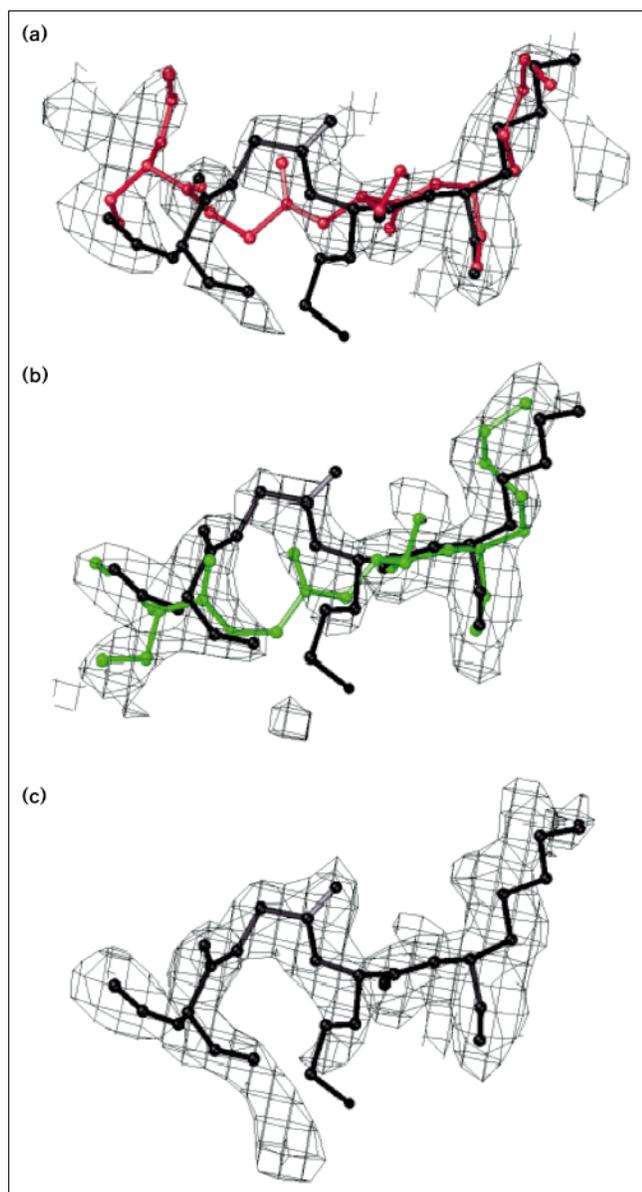
Although there are some similarities between averaging individually refined structures and multiconformer models

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, multiconformer 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 as the whole ensemble is refined against the data. Clearly, this method requires high-quality data and a high observable-to-parameter ratio.

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 [8,10]). However, simulated-annealing refinement alone is still insufficient to refine a structure automatically

Figure 5



A demonstration of the improvement provided by averaging the structure factors of several independently refined structures. Five torsion-angle molecular dynamics simulated-annealing refinements using the maximum-likelihood target were carried out for the recently solved structure of human heterogeneous ribonucleotide protein A1 (hnRNP) [10,77]. Cross-validated σ_A weighted electron-density maps, contoured at 1σ , are shown for: (a) the typical result of the refinements (red line, $R = 35.7\%$, $R_{\text{free}} = 41.3\%$); (b) the best result of the refinements (green line, $R = 36.2\%$, $R_{\text{free}} = 40.9\%$); and (c) the average map from averaging structure factors calculated from the four best models ($R = 36.6\%$, $R_{\text{free}} = 38.9\%$ for the average F_o). In all cases the refined structure is shown in black. (The figure was generated with the program O [82].)

without human intervention. For example, crystallographic refinement using simulated annealing typically cannot correct chain tracing errors, such as register shifts. In the

case of NMR, 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 [45]. Fully automatic structure determination remains a distant goal that will be likely to require significant new algorithmic developments.

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 multiconformer models which can represent molecular motion and discrete disorder, especially when combined with the acquisition of high-quality data [19]. Simulated annealing is thus a stepping stone towards the development of improved models of macromolecules both in solution and in the crystalline state.

Acknowledgements

The recent computational developments discussed in this review will become available on the internet (URL: <http://atb.csb.yale.edu>) in the near future. We thank Temple Burling, Michael Nilges and 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

1. Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1986). Modeling of data. In *Numerical Recipes*. pp. 498–546, Cambridge University Press, UK.
2. Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P., Jr. (1983). Optimization by simulated annealing. *Science* **220**, 671–680.
3. Laarhoven, P.J.M. & Aarts, E.H.L. (1987). *Simulated Annealing: Theory and Applications*. D Reidel Publishing Company, Dordrecht, Germany.
4. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460.
5. Brünger 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.
6. 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.
7. Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474.
8. Rice, L.M. & Brünger, A.T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19**, 277–290.
9. Pannu, N.S. & Read, R.J. (1996). Improved structure refinement through maximum likelihood. *Acta Cryst. A* **52**, 659–668.
10. Adams, P.D., Pannu, N.S., Read, R.J. & Brünger, A.T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. USA*, in press.
11. Brünger, 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.
12. Stein, E.G., Rice, L.M. & Brünger, A.T. (1997). Torsion angle molecular dynamics is a new, efficient tool for NMR structure calculation. *J. Magn. Resonance Ser. B* **124**, 154–164.
13. Braun, W. & Gö, N. (1985). Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J. Mol. Biol.* **186**, 611–626.
14. Braun, W. (1987). Distance geometry and related methods for protein structure determination from NMR data. *Quart. Rev. Biophys.* **19**, 115–157.

15. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988). 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.
16. Nilges, M., Kuszewski, J. & Brünger, A.T. (1991). Sampling properties of simulated annealing and distance geometry. In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*. (Hoch, J.C., Poulsen, F.M. & Redfield, C., eds), pp. 451–455, Plenum Press, NY, USA.
17. Kuriyan, J., Ösapay, K., Burley, S.K., Brünger, A.T., Hendrickson, W.A. & Karplus, M. (1991). Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins* **10**, 340–358.
18. Burling, F.T. & Brünger, A.T. (1994). Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Isr. J. Chem.* **34**, 165–175.
19. Burling, F.T., Weis, W.I., Flaherty, K.M. & Brünger, A.T. (1996). Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **271**, 72–77.
20. Bonvin, A.M.J.J. & Brünger, A.T. (1995). Conformational variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.* **250**, 80–93.
21. Bonvin, A.M.J.J. & Brünger, A.T. (1996). Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? *J. Biomol. NMR* **7**, 72–76.
22. Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270.
23. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. *Acta Cryst. A* **47**, 392–400.
24. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A.T. & Berman, H.M. (1996). New parameters for the refinement of nucleic acid containing structures. *Acta Cryst. D* **52**, 57–64.
25. Allen, F.H., Kennard, O. & Taylor, R. (1983). Systematic analysis of structural data as a research technique in organic chemistry. *Accounts Chem. Res.* **16**, 146–153.
26. Dauter, Z., Lamzin, V.S. & Wilson, K.S. (1995). Proteins at atomic resolution. *Curr. Opin. Struct. Biol.* **5**, 784–790.
27. Sevcik, J., Dauter, Z., Lamzin, V.S. & Wilson, K.S. (1996). Ribonuclease from *Streptomyces aureofaciens* at atomic resolution. *Acta Cryst. D* **52**, 327–344.
28. Stec, B., Zhou, R. & Teeter, M.M. (1995). Full-matrix refinement of the protein crambin at 0.83 Å and 130K. *Acta Cryst. D* **51**, 663–681.
29. 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.
30. Karplus, M. & Petsko, G.A. (1990). Molecular dynamics simulations in biology. *Nature* **347**, 631–639.
31. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324.
32. Brünger, 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.
33. Weis, W.I., Brünger, A.T., Skehel, J.J. & Wiley, D.C. (1989). Refinement of the influenza virus haemagglutinin by simulated annealing. *J. Mol. Biol.* **212**, 737–761.
34. Brünger, A.T., Krukowski, A. & Erickson, J. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst. A* **46**, 585–593.
35. Fujinaga, M., Gros, P. & van Gunsteren, W.F. (1989). Testing the method of crystallographic refinement using molecular dynamics. *J. Appl. Cryst.* **22**, 1–8.
36. Silva, A.M. & Rossmann, M.G. (1985). The refinement of southern bean mosaic virus in reciprocal space. *Acta Cryst. B* **41**, 147–157.
37. Read, R.J. (1990). Structure-factor probabilities for related structures. *Acta Cryst. A* **46**, 900–912.
38. 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.
39. Bricogne, G. (1993). Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. *Acta Cryst. D* **49**, 37–60.
40. Kleywegt, G.J. & Brünger, A.T. (1996). Cross-validation in crystallography: practice and applications. *Structure* **4**, 897–904.
41. Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, 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.
42. Nilges, M. (1996). Structure calculation from NMR data. *Curr. Opin. Struct. Biol.* **6**, 617–623.
43. Karplus, M. (1963). Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85**, 2870–2871.
44. Wüthrich, K. (1986). Three-dimensional protein structure by NMR. In *NMR of Proteins and Nucleic Acids*. pp. 176–199, Wiley, NY.
45. 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.
46. Kim, Y. & Prestegard, J.H. (1990). Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins* **8**, 377–385.
47. Garrett, D.S., et al., & 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. Resonance Ser. B* **104**, 99–103.
48. Mierke, D.F., Huber, T. & Kessler, H. J. (1994). Coupling constants again: experimental restraints in structure refinement. *Comput. Aided Mol. Des.* **8**, 29–40.
49. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. (1995). The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Resonance Ser. B* **107**, 293–297.
50. Kuszewski, J., Qin, J., Gronenborn, A.M. & Clore, G.M. (1995). The impact of direct refinement against ¹³C alpha and ¹³C beta chemical shifts on protein structure determination by NMR. *J. Magn. Resonance Ser. B* **106**, 92–96.
51. Oldfield, E. (1995). Chemical shifts and three-dimensional protein structures. *J. Biomol. NMR* **5**, 217–225.
52. Nilges, M., Gronenborn, A.M., Brünger, A.T. & Clore, G.M. (1988). 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.
53. 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.
54. Diamond, R. (1971). A real-space refinement procedure for proteins. *Acta Cryst. A* **27**, 436–452.
55. 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. *Protein Sci.* **5**, 1067–1080.
56. Brünger, 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.
57. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58.
58. Arnold, E. & Rossmann, M.G. (1988). The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. *Acta Cryst. A* **44**, 270–282.
59. Fujinaga, M. (1993). Crystallographic refinement. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*. (van Gunsteren, W.F., Weiner, P.K. & Wilkinsin, A.J., eds), pp. 371–381, ESCOM Science Publishers B.V., Leiden, Netherlands.
60. 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.
61. Verlet, L. (1967). Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–105.
62. Curro, J. (1974). Computer simulation of multiple chain systems—the effect of density on the average chain dimension. *J. Chem. Phys.* **61**, 1203–1207.
63. 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.
64. Xu, Y. & Krishna, N.R. (1995). Structure determination from NOESY intensities using a metropolis simulated-annealing (MSA) refinement of dihedral angles. *J. Mag. Resonance Ser. B.* **108**, 192–196.

65. Saunders, M. (1987). Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum. *J. Am. Chem. Soc.* **109**, 3150–3152.
66. 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.
67. Abagyan, R. & Argos, P. (1992). Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J. Mol. Biol.* **225**, 519–532.
68. Goldstein, H. (1980). *Classical Mechanics 2nd Edition*. Addison-Wesley Publishing Co., Reading, MA.
69. Bae, D.-S. & Haug, E.J. (1987). A recursive formulation for constrained mechanical system dynamics: Part I. Open loop systems. *Mech. Struct. Mach.* **15**, 359–382.
70. Bae, D.-S. & Haug, E.J. (1988). A recursive formulation for constrained mechanical system dynamics: Part II. Closed loop systems. *Mech. Struct. Mach.* **15**, 481–506.
71. Jain, A., Vaidehi, N. & Rodriguez, G. (1983). A fast recursive algorithm for molecular dynamics simulation. *J. Comp. Phys.* **106**, 258–68.
72. 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* **20**, 227–247.
73. Abramowitz, M. & Stegun, I. (1968). Numerical interpolation, differentiation and integration. In *Handbook of Mathematical Functions, Applied Mathematics Series, Vol. 55*. pp. 877–899, Dover Publications, NY.
74. 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.
75. Hoppe, W. (1957). Die Faltmolekülmethode — eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannter Molekülstruktur. *Acta Cryst.* **10**, 750–751.
76. Rossmann, M.G. & Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst. A* **15**, 24–51.
77. 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.*, in press.
78. 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.
79. 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.
80. Pflugrath, J.W., Wiegand, G., Huber, R. & Vertesey, L. (1986). Crystal structure determination, refinement and the molecular model of the alpha-amylase inhibitor Hoe-467A. *J. Mol. Biol.* **189**, 383–386.
81. Clore, G.M., *et al.*, & Gronenborn, A.M. (1988). Refinement of the solution structure of the DNA dodecamer 5'd(CGCGPATTGCG)2 containing a stable purine–thymine base pair: combined use of nuclear magnetic resonance and restrained molecular dynamics. *Biochemistry* **27**, 4185–4197.
82. 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.