Acta Cryst. (1990). A46, 585-593

Slow-Cooling Protocols for Crystallographic Refinement by Simulated Annealing

BY AXEL T. BRÜNGER* AND ANTON KRUKOWSKI

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

AND JOHN W. ERICKSON

Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

(*Received* 30 *October* 1989; *accepted* 12 *February* 1990)

Abstract

An improved protocol for crystallographic refinement by simulated annealing is presented. It consists of slow cooling starting at high temperatures. Tests of refinements of aspartate aminotransferase and porcin pepsin show that the slow-cooling protocol produces lower R factors and better geometry than other protocols previously published. The influence of the temperature-control method, weighting, cooling rate and duration of the heating stage on the success of the slow-cooling protocol is studied. Analysis of the time course of the potential-energy fluctuations indicates no global changes in the state of order of the system. Fluctuations of the potential energy are interpreted as localized conformational changes during the course of the refinement.

Abbreviations

CPU, central processing unit; mAATase, K258A mutant of aspartate aminotransferase; MIR, multiple isomorphous replacement; r.m.s., root-mean-square; SA refinement, crystallographic refinement by simulated annealing with molecular dynamics.

Introduction

Crystallographic refinement of macromolecules can be viewed as a nonlinear optimization problem which consists of minimizing a hybrid energy function (Jack & Levitt, 1978; Hendrickson, 1985; Brünger, Kuriyan & Karplus, 1987)

$$E_{\text{pot}} = E_{\text{chem}} + wE_{\text{X-ray}}, \qquad (1)$$

where E_{chem} comprises empirical information about equilibrium covalent-bonding geometry, hydrogen-bonded and nonbonded interactions, and $E_{\text{X-ray}}$ typically consists of the crystallographic residual. In general, $E_{\text{X-ray}}$ may consist of any target function which becomes zero if the model matches the data. For example, $E_{\text{X-ray}}$ can be set to $(1-c)$ where c is the linear correlation coefficient between observed and computed data (Brünger, 1990). The weighting factor w relates $E_{\text{X-ray}}$ to E_{chem}. Restrained least-squares refinement (Hendrickson, 1985) represents a special case of (1), in which all terms in E_{chem} are represented as squared differences between ideal data and calculated data. In the general case, E_{chem} consists of an empirical energy function (Burkert & Allinger, 1982).

For macromolecules, gradient descent minimization or restrained least-squares refinement of (1) will fail unless the model is fairly close to optimal. Better convergence is achieved if instead of restricting search directions to be energetically downhill, the optimization is allowed to go uphill as well. This process has been referred to as simulated annealing (SA) by Kirkpatrick, Gelatt & Vecchi (1983) who implemented it as a Monte Carlo algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953). SA was first applied to crystallographic refinement by Brünger *et al.* (1987) who implemented it as a molecular dynamics algorithm (Verlet, 1967). The latter method is now commonly referred to as SA refinement or molecular-dynamics refinement. A number of tests on crambin (Brünger, Karplus & Petsko, 1989), aspartate aminotransferase (Brünger, 1988*a*), myohemerythrin (Kuriyan, Brünger, Karplus & Hendrickson, 1989), and phospholipase A2 (Fujinaga, Gros & van Gunsteren, 1989) have shown the superiority of SA refinement over restrained least-squares refinement. SA refinement has already been used to refine a number of new structures, including foot-and-mouth disease virus (Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989), the human immunodeficiency virus-encoded protease (Navia *et al.*, 1989; Wlodawer *et al.*, 1989), the membrane-pore-forming fragment of colicin A (Parker, Pattus, Tucker & Tsernoglu, 1989), aconitase (Robbins & Stout, 1989), tumour necrosis factor (Jones, Stuart & Walker, 1989), thermitase complexed with eglin-c (Gros, Fujinaga, Dijkstra, Kalk & Hol, 1989), human renin (Sielecki *et al.*, 1989), streptavidin (Weber,

* To whom correspondence should be addressed

0108-7673/90/070585-09$03.00

Ohlendorf, Wendoloski & Salemme, 1989), HLA-Aw68 (Garrett, Saper, Bjorkman, Strominger & Wiley, 1989), triosephosphate isomerase (Robert Davenport, personal communication), influenza virus hemagglutinin (Weiss, Brünger, Skehel & Wiley, 1990), aspartate carbamoyltransferase (Gouaux & Lipscomb, 1988), glycogen phosphorylase b (Barford & Johnson, 1989), and the nucleic acid dodecamer CGCATATATGCG (Yoon, Prive, Goodsell & Dickerson, 1988).

Despite the success of this method there still remain questions of how to choose the optimal 'annealing' schedule for SA refinement of a particular macromolecule which determines how the temperature or weights are modified during the simulation. Kirkpatrick *et al.* (1983) kept the temperature initially very high and then 'annealed' the system by slowly reducing the temperature. In other words, a coarse search is carried out at high temperatures and a local minimum is approached during the cooling stage. This procedure may have to be repeated to reach the global minimum. It should be pointed out that the 'temperature' of the system is not a physical temperature but rather a control parameter that determines whether the system can escape certain local minima. Thus, very high temperatures may have to be introduced if the barriers between local minima are large.

In this report, we first introduce a general classification scheme for annealing schedules that are applicable to SA refinement. We then investigate the influence of the temperature-control method, cooling rate, duration of the heating stage, and weighting of $E_{\text{X-ray}}$ on the outcome of SA refinement. We also analyze the time course of conformational changes and E_{pot} during SA refinement. This is of interest since the fluctuations of E_{pot} are used by Kirkpatrick *et al.* (1983) to control the annealing schedule in SA for combinatorial optimization problems. Finally, we present further examples for the large radius of convergence of SA refinement in terms of conformational changes as well as improvements of electron density maps. As test cases for these studies we have chosen the K258A mutant of aspartate aminotransferase that was studied previously by Brünger (1988*a*) and the complex of porcine pepsin with a renin inhibitor (A63218) whose structure will be reported elsewhere (Erickson *et al.*, in preparation).

Materials and methods

Aspartate aminotransferase

The particular system studied is a single-site mutant of the protein aspartate aminotransferase from the aspC gene product of *E. coli* (Smith, Ringe, Finlayson & Kirsch, 1986) consisting of 396 amino acids. The active-site lysine has been substituted by alanine (K258A). The space group is $C222_1$. Diffraction data to 2·8 Å resolution were available. An initial atomic model was fitted by Smith, Almo, Toney & Ringe (1989) to an electron density map obtained by multiple isomorphous replacement (MIR). 8786 reflections to 2·8 Å above 2σ were included during the SA refinement. This resulted in about 70% of the theoretically observable reflections. MIR phases were available up to 3 Å resolution. No weighting scheme was applied, *i.e.* all weights were set to unity. Isotropic atomic temperature factors were used and set to 12 Å^2. In addition to the protein, a pyridoxamine phosphate cofactor and a sulfate molecule were included in the atomic model. The parameters for the pyridoxamine phosphate cofactor were inferred from the nucleic acid parameters developed by Nilsson & Karplus (1986).

Pepsin–A63218 complex

Porcine pepsin, a 326-amino-acid aspartic proteinase, was co-crystallized with the renin inhibitor A63218 in space group $P2_12_12_1$ (Erickson, Abad-Zapatero, Rydel & Luly, 1987). Diffraction data to 2·5 Å resolution were collected by diffractometry. 8028 reflections above 2σ representing about 90% of the theoretically unique data were used for SA refinement. The structure was solved by molecular replacement using a partially refined model of native porcine pepsin. The crystal structure of the complex was refined initially by performing two rounds of least-squares refinement using the program *PROLSQ* (Konnert & Hendrickson, 1980) with one round of map fitting using *FRODO* (Jones, 1978) to an R factor of 35·7% for data from 6·0 to 2·5 Å resolution. This structure was then used for the SA refinement 'experiments' described below.

Temperature control for molecular dynamics

Molecular-dynamics simulations involve the simultaneous solution of the classical equations of motion for all atoms i of a macromolecule where the forces are derived from the potential energy, E_{pot},

$$m_i \partial^2 r_i / \partial t^2 = -\nabla_i E_{\text{pot}}. \qquad (2)$$

The quantities r_i and m_i are the coordinates and masses of atom i, respectively. The solution of the classical equations of motion is carried out numerically by the Verlet (1967) algorithm. The initial velocities are assigned to a Maxwellian distribution at the appropriate temperature. There are a number of ways to control the temperature during a molecular-dynamics simulation. Three specific methods will be discussed below: velocity scaling, Langevin dynamics and T coupling.

Velocity scaling. Velocity scaling consists of periodic uniform scaling of the velocities v_i, *i.e.*

$$v_i^{\text{new}} = v_i^{\text{old}} (T/T_{\text{curr}})^{1/2} \quad (3)$$

for all atoms i where T is the target temperature and T_{curr} is the current temperature. T_{curr} is defined as

$$T_{\text{curr}} = 2E_{\text{kin}}/3nk \quad (4)$$

where n is the number of atoms and k is Boltzmann's constant. Each time velocities are scaled, the Verlet (1967) algorithm has to be restarted since it propagates positions without reference to velocities.

Langevin dynamics. Langevin dynamics incorporates the influence of a heat bath into the classical equations of motion,

$$m_i \frac{\partial^2 r_i}{\partial t^2} = -\nabla_i E_{\text{pot}} - m_i \gamma_i \frac{\partial r_i}{\partial t} + R(t), \quad (5)$$

where γ_i specifies the friction coefficient for atom i, and $R(t)$ is a random force. $R(t)$ is assumed to be uncorrelated with the positions and velocities of the atoms. It is described by a Gaussian distribution with mean of zero and variance

$$\langle R(t)R(t')\rangle = 2m_i\gamma_i kT\delta(t-t') \quad (6)$$

where k is Boltzmann's constant and $\delta(t-t')$ is the Dirac delta function. The finite-difference algorithm of Brünger, Brooks & Karplus (1984) can be used to obtain a numerical solution of (5) which is valid in the limit of $\gamma_i \Delta t \ll 1$. Pastor, Brooks & Szabo (1988) showed that this algorithm is preferable to other finite-difference algorithms to propagate positions.

T coupling. The weak temperature-coupling method of Berendsen, Postma, van Gunsteren, DiNola & Haak (1984), which we refer to as T coupling, is related to Langevin dynamics except that it does not use random forces and it applies a global scale factor to the friction coefficient which is proportional to the ratio of T_{curr} and T.

$$m_i \frac{\partial^2 r_i}{\partial t^2} = -\nabla_i E_{\text{pot}} - m_i \gamma_i v_i \left(1 - \frac{T}{T_{\text{curr}}}\right). \quad (7)$$

The second term on the right-hand side of (7) is equivalent to a positive friction coefficient if $T_{\text{curr}} > T$, thus lowering the temperature; it produces a negative friction coefficient if $T_{\text{curr}} < T$, thus increasing the temperature. Equation (7) can be integrated numerically using the algorithm of Brünger *et al.* (1984).

The computer program

All calculations were carried out with the program *X-PLOR*, version 1.5 (Brünger, 1988*b*). Details of the parameterization and functional form of E_{pot} have been published previously (Brünger *et al.*, 1989; Brünger, 1989). Requests for *X-PLOR* should be made to ATB.

Simulated annealing strategies

The success of SA critically depends on the choice of the annealing schedule (Bounds, 1987), that is, the sequence of and numerical values for control parameters, such as temperature and relative weighting factors of the various components of E_{pot}. We introduce below a classification scheme for the control parameters of SA that is specific for minimization problems involving a hybrid energy function E_{pot} [(1)].

SA of type one

The basic idea of SA using molecular dynamics is to increase the temperature T of the system so as to ensure that the atoms can overcome local energy barriers of E_{pot}. This approach was first applied to crystallographic refinement by Brünger *et al.* (1987), who heated the system and then cooled it while maintaining a constant weight w.

Equivalent to the modification of T is scaling of E_{pot} by an overall scale factor c with $0 \le c \le 1$ while maintaining a constant temperature T. This can be seen as follows. Suppose Newton's equations are solved using the scaled energy cE_{pot} at constant T:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = -\nabla_i c E_{\text{pot}} \quad (8)$$

$$T = \frac{2E_{\text{kin}}}{3nk} = \frac{2}{3nk} \sum_i \frac{1}{2} m_i \left(\frac{\partial r_i}{\partial t}\right)^2 = \text{constant}. \quad (9)$$

This is equivalent to

$$m_i \frac{\partial^2 r_i}{\partial t'^2} = -\nabla_i E_{\text{pot}} \quad (10)$$

$$T' = \frac{2E'_{\text{kin}}}{3nk} = \frac{2}{3nk} \sum_i \frac{1}{2} m_i \left(\frac{\partial r_i}{\partial t'}\right)^2 = T/c \quad (11)$$

with the new time scale $t' = \sqrt{c}\,t$ and the scaled temperature T'. We will refer to procedures that scale the temperature T or, equivalently, E_{pot}, as SA of the first type.

SA of type two

In addition to the overall scaling of E_{pot}, the weight w of $E_{\text{X-ray}}$ [(1)] or weights of individual residual terms of $E_{\text{X-ray}}$ may be modified during the course of SA. In particular, w may be lowered while keeping the original E_{chem} term and maintaining a constant temperature at 300 K. We refer to weighting procedures involving $E_{\text{X-ray}}$ alone as SA of the second type. This approach was first applied to crystallographic refinement by Fujinaga *et al.* (1989). The approach is based on the assumption that it is the $E_{\text{X-ray}}$ term that introduces large energy barriers which would normally require the use of very high temperatures, such as 1000–4000 K, to overcome them.

Table 1. *Comparison of temperature-control methods for mAATase*

The cooling stage was preceded by minimization of the initial MIR structure which consisted of 40 conjugate gradient steps with a soft repulsive potential followed by 120 conjugate gradient steps with a van der Waals nonbonded potential, C^α restraints at 84 kJ (mol Å^2)$^{-1}$, $W_A = 544\,180$ kJ mol^{-1}, $W_P = 50\,232$ kJ (mol rad^2)$^{-1}$, $B = 12{\cdot}0$ Å^2, $\Delta_F = 0{\cdot}05$ Å. The minimization stage was followed by a 5 ps heating stage at $T = 4000$ K using velocity scaling. During the cooling stage $\Delta_F = 0{\cdot}2$ Å. The time step was set to 0·5 fs during the molecular-dynamics stages. The cooling stage was followed by 120 conjugate gradient steps as above but without C^α restraints, $W_P = 0$ kJ (mol rad^2)$^{-1}$ and $\Delta_F = 0{\cdot}05$ Å. Data between 8 and 2·8 Å resolution were used. [See Brünger (1988*a*) for definition of W_A, W_P, Δ_F.] 'Fast' cooling was carried out over a period of 0·25 ps with the target temperature set to 300 K. 'Slow' cooling was carried out over a period of 1·85 ps with the target temperature starting at 4000 K, temperature decrements of $\Delta T = 25$ K every 25 steps, and a final temperature of 300 K. 'Velocity scaling' was carried out every 25 fs using equation (3). Langevin dynamics [equation (5)] and T coupling [equation (7)] were carried out with the friction coefficient β set to 100 ps^{-1} for all atoms.

Cooling mode	Cooling rate	R factor (%)	Δ_{bonds} (Å)	Δ_{angles} (°)
Initial		41·8	0·037	4·8
Velocity scaling	Fast	22·8	0·018	4·3
Velocity scaling	Slow	21·7	0·018	4·1
Langevin	Fast	23·1	0·018	4·3
Langevin	Slow	22·2	0·018	4·2
T coupling	Fast	22·4	0·018	4·2
T coupling	Slow	21·9	0·018	4·0

SA of type three

The third alternative is to control the weights of components of both E_{chem} and $E_{X\text{-}ray}$ individually in order to ensure that the energy barriers for each component are lower than the kinetic energy of the atoms. Examples of energy components are bond lengths, bond angles, planarity restraints, or diffraction data belonging to different resolution shells. This third type of SA has been applied in a different context to structure determination based on interproton-distance nuclear magnetic resonance data starting from a random array of atomic coordinates (Nilges, Clore & Gronenborn, 1988). It has not yet been applied to crystallographic SA refinement.

Results and discussion

The three types of temperature control (velocity scaling, Langevin dynamics and T coupling) are compared in Table 1 for SA refinement of mAATase using two different cooling rates starting from the same initial structure. Listed are the R factor and the deviations of bond lengths and bond angles from ideality. For all three types of temperature control, slow cooling is superior to the fast-cooling protocols. Overall, the T-coupling method produces the best results, followed by the velocity-scaling method, and then the Langevin method. T coupling is preferable to Langevin dynamics in the context of SA since the friction term of the latter slows the atomic motions [(5)] whereas the friction term for T coupling becomes zero if T_{curr} agrees with T [(7)], thus speeding up convergence. T coupling is preferable to velocity scaling as the latter is typically performed at intervals of 25 to 50 integration steps and, thus, large temperature oscillations can occur. Experience has shown that these oscillations cause instabilities of the numerical integration at high temperatures. Therefore, we conclude that slow cooling with T coupling is the preferred temperature-control method for SA refinement.

The influence of the cooling rate and other parameters on the success of SA refinement has been investigated by carrying out different SA refinement 'experiments' starting from the same initial structure (see Tables 2 and 3). The results can be summarized as follows. The slower the cooling rate the better the R factor and geometry (compare jobs 3, 4, 5 in Table 2 and jobs 3, 4 in Table 3). An extended heating period prior to slow cooling does not necessarily improve the R factor (compare jobs 1, 2, 5 in Table 2). It is therefore more efficient to use a slower cooling rate than to heat the structure for a long period of time. Moreover, the slow-cooling process can be started from the beginning without equilibration of the system at the initial high temperature (compare jobs 2 and 3 in Table 2). The slow cooling rate ensures that the system quickly reaches equilibrium at the starting temperature, so that equilibration at the initial temperature is not necessary. A starting temperature of 4000 K is better than 2000 K, yet starting at 6000 K is not significantly better (compare jobs 5, 6 and 7 in Table 2). The lower temperature limit listed in the cooling-range column in Tables 2 and 3 is the temperature at which slow cooling is switched to conjugate gradient minimization. It appears that a low-temperature limit of 0 K is not better than 300 K, and a low-temperature limit of 800 K is only slightly worse than 300 K (compare jobs 5, 8, 9 in Table 2). The subsequent execution of two slow-cooling cycles improves results, but execution of a third slow-cooling cycle does not improve results further in the case of mAATase (compare jobs 5, 11, 12 in Table 2). Neutralizing charges on ionizable side chains (Asp, Glu, Lys, and Arg) improves the R factor (compare jobs 1 and 2 in Table 3) which was noted previously by Kuriyan *et al.* (1989). Increasing the tolerance for the structure-factor calculation update during molecular dynamics [Δf; see Brünger *et al.* (1989) for definition] improves the R factor but is computationally expensive (compare jobs 9 and 10 in Table 2).

SA of the first and second type of E_{pot} are compared in Table 2 for mAATase. The protocols for the second type of SA are similar to the ones used by Fujinaga *et al.* (1989) and Gros *et al.* (1989). The first type of SA produced somewhat lower R factors and it

Table 2. *Comparison of slow-cooling protocols for mAATase*

The heating and cooling stages were preceded and followed by minimization stages as described in Table 1. The time step was set to 0·5 fs for all cases that start at 4000 or 2000 K, 1 fs for all cases that start at 600 or 300 K, and 0·25 fs for all cases that start between 6000 and 4000 K. The *T*-coupling method was used in all cases.

Job	Heating	Cooling range (K)	Cooling rate (K fs^{-1})	*R* factor (%)	Δ_{bonds} (Å)	Δ_{angles} (°)	CPU time (h)*
Initial				41·8	0·037	4·8	—
1	5 ps at T = 4000 K	4000-300	25/12·5	21·9	0·018	4·0	172
2	1 ps at T = 4000K	4000-300	25/12·5	21·6	0·017	4·1	64
3	None	4000-300	25/12·5	21·8	0·018	4·0	38
4	None	4000-300	25/6.25	23.0	0·018	4·3	24
5	None	4000-300	25/25	21·5	0·017	3·9	74
6	None	2000-300	25/25	22·6	0·019	4·2	28
7	None	6000-300	25/25	21·4	0·017	3·8	106
8	None	4000-0	25/25	21·5	0·017	3·9	76
9	None	4000-800	25/25	21·7	0·017	4·0	64
10†	None	4000-800	25/25	20·9	0·017	4·0	250
11	None	4000-300	25/25				
		4000-300	25/25	21·1	0·017	3·9	147
12	None	4000-300	25/25				
		4000-300	25/25				
		4000-300	25/25	21·1	0·017	3·9	221
13	5 ps at T = 300 K (3·5 Å)						
	5 ps at T = 300 K (2·8 Å)	None	—	23·0	0·019	4·2	84
14	5 ps at T = 300 K (3·5 Å)						
	2 ps at T = 300 K (2·8 Å, wa/4)						
	2 ps at T = 300 K (2·8 Å, wa/2)						
	2 ps at T = 300 K (2·8 Å, wa)	None	—	22·0	0·018	4·0	94

* CPU time on a Convex-C1.
† Δ_F = 0 Å.

Table 3. *Comparison of SA refinement protocols for pepsin*

The heating stage was preceded by minimization of the initial pepsin A-63218 structure which consisted of 30 conjugate gradient steps with van der Waals nonbonded potentials, C^α restraints at 84 kJ (mol Å^2)$^{-1}$, W_A = 506 297 kJ (mol rad^2)$^{-1}$, W_P = 0 kJ (mol rad^2)$^{-1}$, B = 12·0 Å^2, Δ_F = 0·05 Å. During the heating and cooling stages, Δ_F = 0·2 A and the time step was set to 1·0 fs. The cooling stage was followed by 120 conjugate gradient steps as above but without C^α restraints, and Δ_F = 0 Å. 8087 reflections between 6·0 and 2·5 Å resolution were used. The velocity scaling method [equation (3)] was used for all cases. [See Brünger (1988*a*) for definition of W_A, W_P, Δ_F.]

Job	Heating	Cooling range (K)	Cooling rate (K fs^{-1})	*R* factor (%)	Δ_{bonds} (Å)	Δ_{angles} (°)	Remarks
Initial				35·7	0·028	5·3	
1	4 ps at T = 2000 K	2000-300	1700/250	27·6	0·022	5·2	
2	4 ps at T = 2000 K	2000-300	1700/250	26·8	0·025	5·2	Asp, Glu, Lys, Arg neutral
3	4 ps at T = 4000 K	4000-300	1700/250	27·7	0·026	5·4	Asp, Glu, Lys, Arg neutral
4	4 ps at T = 4000 K	4000-300	25/25	25·3	0·024	5·1	Asp, Glu, Lys, Arg neutral

required less CPU time (compare jobs 5, 13 and 14 in Table 2).

The time course of the *R* factor, $E_{pot}(t)$, and the r.m.s. fluctuations of $E_{pot}(t)$ are shown in Figs. 1 and 2 for the SA refinements of mAATase (job 10) and pepsin-A63218 (job 4), respectively. The r.m.s. fluctuations of $E_{pot}(t)$ are defined by

$$r[E_{pot}(t)] = \{\langle [E_{pot}(t) - \langle E_{pot}(t)\rangle]^2\rangle\}^{1/2} \quad (12)$$

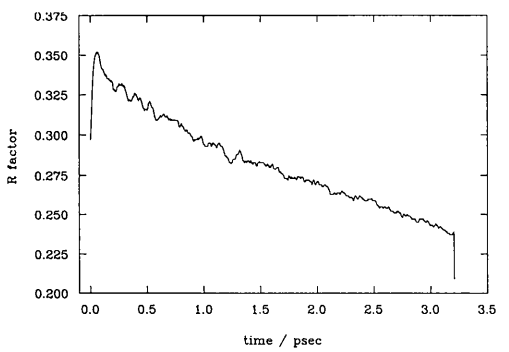
where the brackets $\langle\,\rangle$ denote the mean computed in 25 fs intervals of the simulation. The *R* factor, $E_{pot}(t)$, and the magnitude of the 'spikes' of $r[E_{pot}(t)]$ decrease more or less monotonically during the course of the cooling process (Fig. 1). The r.m.s. fluctuations $r[E_{pot}(t)]$ are related to the specific heat $c(t)$ by

$$c(t) = \{r[E_{pot}(t)]\}^2/kT^2. \quad (13)$$

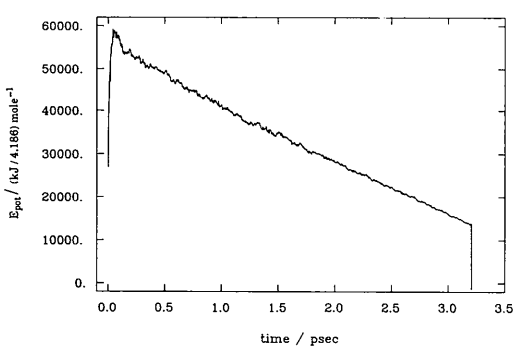
Both $r[E_{pot}(t)]$ and $c(t)$ indicate the rate of change of energy with respect to the temperature *T*. Large values of either quantity indicate a phase transition, that is, a global change in the state of order of the system, such as freezing or condensation. The spikes in Fig. 1(*c*) are not large enough to suggest a phase transition. In marked contrast, Kirkpatrick *et al.* (1983) obtained changes of the specific heat near phase transitions by almost an order of magnitude relative to ordered states of the system [*cf.* Fig. 4 of Kirkpatrick *et al.* (1983)].

The spikes of $r[E_{pot}(t)]$ are correlated with conformational changes that occur during SA refinement. We were able to correlate one of the largest spikes in Fig. 1(*c*) with a conformational change around residue 163 of mAATase. The r.m.s. fluctuations of backbone (C, C^α, N) atoms during the entire course

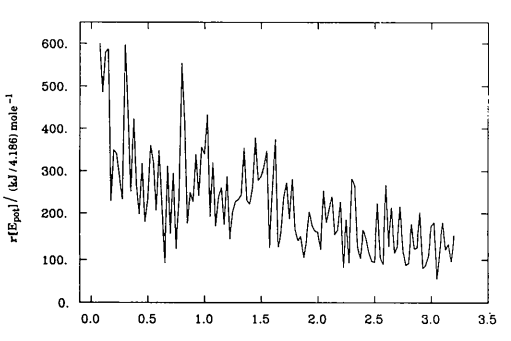
of the slow-cooling job 10 of mAATase are shown in Fig. 3. Several regions exhibit large fluctuations, such as residues 19, 31, 93 and 163. [Fig. 3 looks remarkably similar to Fig. 4(*d*) of Brünger (1988*a*) which shows the r.m.s. fluctuations of (C, C^{α}, N) atoms for a constant-temperature ($T = 2000$ K) heat stage.] To examine these fluctuations in more detail, structures were extracted every 12·5 fs from the slow-cooling job 10. The r.m.s. differences for backbone (C, C^{α}, N) atoms of residues 161–165 were computed for each adjacent pair of structures and plotted against time (Fig. 4). A particularly large conformational change around 0·7 ps coincides with the large spike of $r[E_{pot}(t)]$ (Fig. 1*c*). Similar temporal correlations between conformational changes and spikes of $r[E_{pot}(t)]$ were found in other cases as well, although the correlations were not as pronounced as in the case of residue 163.



(*a*)

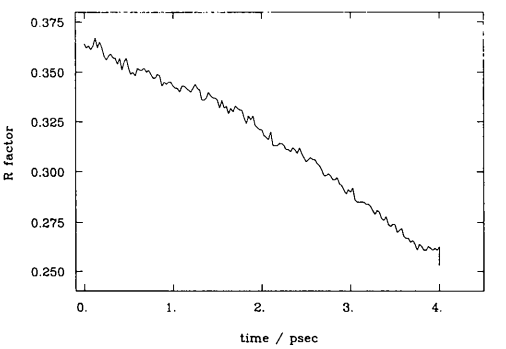


(*b*)

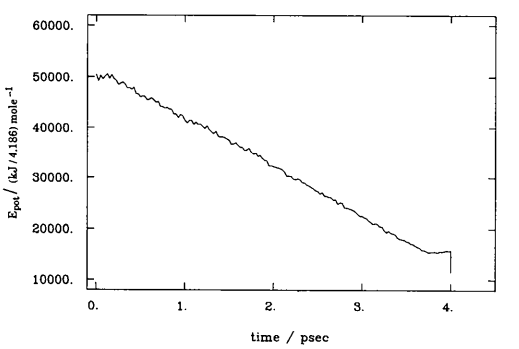


(*c*)

Fig. 1. Time course of (*a*) the *R* factor, (*b*) average of E_{pot}, and (*c*) r.m.s. fluctuations of E_{pot} during 25 fs intervals of the SA refinement of mAATase (job 10 in Table 2). The data points at $t = 0$ correspond to the initial MIR structure after minimization (*cf.* Table 2). The last data points at 3·2 ps correspond to the structure after the final minimization cycle.



(*a*)



(*b*)

Fig. 2. Time course of (*a*) the *R* factor and (*b*) E_{pot} during 25 fs intervals of the SA refinement of pepsin-A63218 (job 4 in Table 3). The data points at $t = 0$ correspond to the structure after the heating stage (4 ps at 4000 K, $R = 36{\cdot}4\%$) just prior to the start of the cooling stage. The last data points at 4·0 ps correspond to the structure after the final minimization cycle.

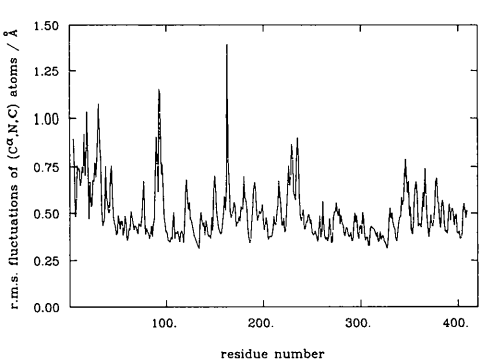


Fig. 3. R.m.s. fluctuations of backbone (N, C, C^{α}) atoms in Å during the time course of the SA refinement of mAATase (job 10 in Table 2) *vs* residue number.

The fit of the atomic model to corresponding $(2|F_{obs}| - |F_{calc}|) \exp(i\varphi_{calc})$ electron density maps around residue 163 of mAATase is compared before and after SA refinement in Fig. 5. The fit in this turn has improved; the electron density map shows more connectivity than before SA refinement. There are still some problems in this region, *e.g.* the density for some of the carbonyl oxygen atoms is not well defined.

Another example of the large radius of convergence of SA refinement is shown in Fig. 6 in the case of pepsin-A63218. The fit of the disulfide loop (Cys 206, Ser 207, Gly 208, Gly 209, Cys 210) to the electron density map was poor in the initial model after *PROLSQ* refinement (Fig. 6*a*). However, SA refinement improved the fit markedly (Fig. 6*b*), and affected changes as large as 7–8 Å in the positions of several side-chain atoms (*e.g.* Ser 207 O^{γ}) and 3–4 Å for several main-chain atoms (*e.g.* Gly 208 O). The SA-refined loop conformation in Fig. 6(*b*) agrees quite well with that found in the structure of native porcine pepsin (Abad-Zapatero, in preparation) and also in the zymogen precursor pepsinogen (Remington & Hartsuck, private communication). This example demonstrates that large corrections for an inaccurate model structure may in certain instances be obtained by using the SA-refinement procedure. In general, the outcome is difficult to predict, and depends upon data quality and on the nature of the errors in the model. In the case of the Cys 206–Cys 210 loop, the overall structural change can be visualized as a nearly 90° rotation of the loop outward from the plane of the diagram in Fig. 6(*a*), with the disulfide bond serving as an approximate hinge for the 'motion'. These changes, although large in magnitude, do not require the breaking of covalent bonds such as is often required during a manual rebuilding session.

In general, the improvement of the electron density map achieved in SA refinement is a consequence of conformational changes distributed throughout the molecule as opposed to a relatively few localized conformational changes. This is a reflection of the fact that the first derivatives of the crystallographic residual [(1)] with respect to the coordinates of a particular atom depend not only on the coordinates of that atom and its neighbors but also on the coordinates of all other atoms including solvent atoms in the crystal structure. This global feature of crystallographic refinement is illustrated for the mAATase case in Fig. 7 which shows the time course of the r.m.s. gradient of $E_{X\text{-}ray}$ with respect to atoms around residue 163 for two different conditions. Fig. 7(*a*)

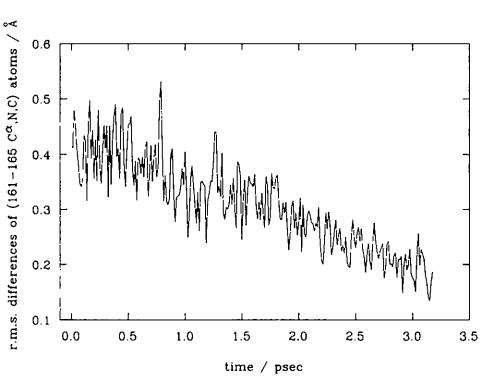


Fig. 4. R.m.s. differences of backbone (N, C, C^{α}) atoms in Å of residues 161–165 between subsequent structures *vs* time during the SA refinement of mAATase (job 10 in Table 2). The r.m.s. differences have been computed between structures of the slow-cooling trajectory that are 25 fs apart.

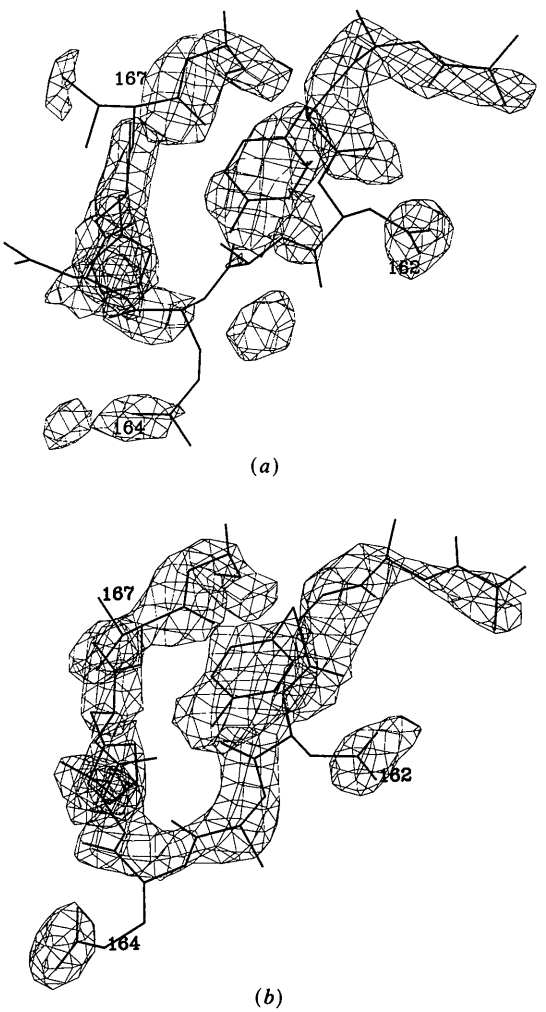


Fig. 5. Pictures showing a segment of mAATase consisting of residues 159–167 (Ala 159, Tyr 160, Tyr 161, Asp 162, Ala 163, Glu 164, Asn 165, His 166, Thr 167). In (*a*) the initial structure after minimization prior to slow cooling and in (*b*) the final SA-refined structure after slow cooling (job 10 in Table 2) and minimization are shown. The atomic models are shown as thick lines. Electron density maps are superimposed as thin lines. The maps were computed from $(2|F_{obs}| - |F_{calc}|)(\exp i\varphi_{calc})$ using φ_{calc} phases corresponding to the atomic models at 2·8 Å resolution. The maps were contoured at 1·5σ.

shows the r.m.s. gradient of $E_{X\text{-}ray}$ computed for the atoms of residues 161–165 only. Fig. 7(*b*) shows the r.m.s. gradient computed for the same atoms but based on an artificial trajectory in which the same set of atoms to compute the gradient were allowed to move according to the actual SA trajectory while all other atoms were held fixed at positions corresponding to the final structure of the slow-cooling trajectory (at 3·2 ps). In the latter case, the gradient is initially larger and it exhibits oscillations that are not present in the former case. The difference between the gradient profiles for the same subset of atoms arises from the contributions of the remainder of the structure to the first derivatives of $E_{X\text{-}ray}$ with respect to the atoms around residue 163.

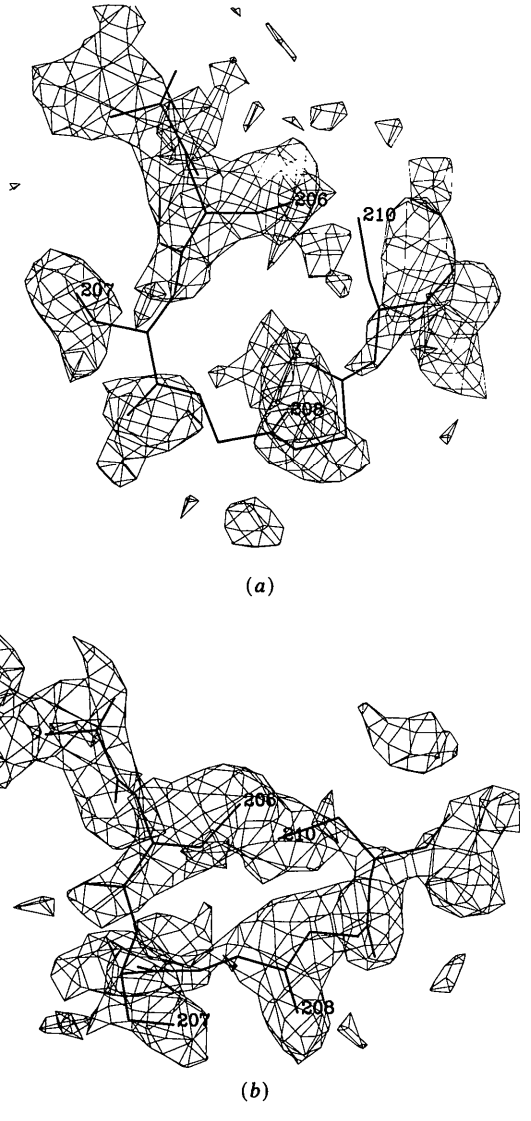


Fig. 6. Pictures showing a segment of pepsin consisting of residues 205–210 (Ala 205, Lys 206, Ser 207, Gly 208, Gly 209, Cys 210). This segment comprises the disulfide loop between Cys 206 and Cys 210. In (*a*), the initial structure derived from a round of *PROLSQ* refinement ($R = 35{\cdot}7\%$) prior to slow cooling, and, in (*b*), the final SA-refined structure after slow cooling (job 4 in Table 3) and minimization ($R = 25{\cdot}3\%$) are shown. The atomic models are shown as thick lines. Electron density maps are superimposed as thin lines. The maps were computed from $(2|F_{obs}| - |F_{calc}|)(\exp i\varphi_{calc})$ using φ_{calc} phases corresponding to the atomic models at 2·5 Å resolution. The maps were contoured at approximately $1{\cdot}0\sigma$, where σ is the r.m.s. electron density in the unit cell.

Concluding remarks

We conclude that *T* coupling is more efficient than either velocity scaling or Langevin dynamics. Slow-cooling protocols produce lower *R* factors than faster-cooling protocols. The first type of SA which was introduced by Brünger *et al.* (1987) produces somewhat lower *R* factors than SA of the second type which was introduced by Fujinaga *et al.* (1989). If the cooling is sufficiently slow, the initial heating stage is not required. Repetitions of the slow-cooling process improve the results somewhat. Based on these results, we recommend the slow-cooling protocol

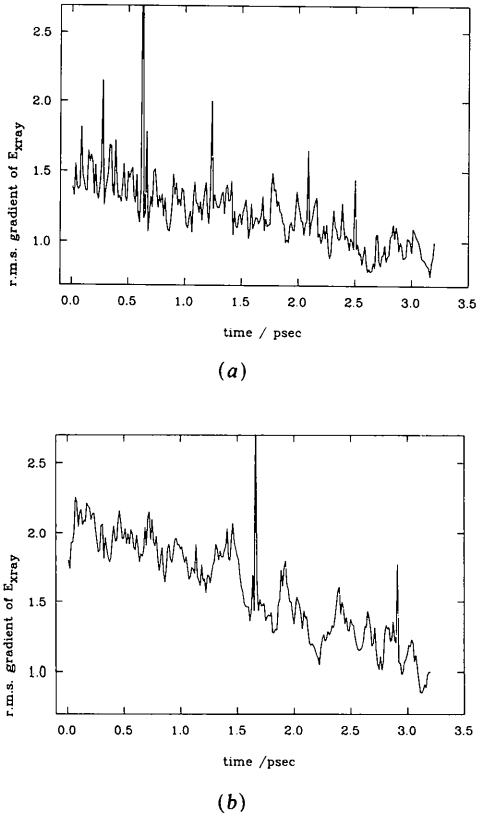


Fig. 7. (*a*) R.m.s. gradient of $E_{X\text{-}ray}$ with respect to atoms of residues 161–165 *vs* time during the SA refinement of mAATase (job 10 in Table 2); the conformations of the atoms are obtained from the slow-cooling trajectory at every 12·5 fs. (*b*) Same as (*a*) except atoms which are not in residues 161–165 are kept fixed in the conformation of the last structure of the actual slow-cooling trajectory (at 3·2 ps).

Table 4. *Recommended slow-cooling protocol*

Stage	Description
1	Determination of weights; sidechains neutral
2	Minimization, 40 conjugate gradient steps with soft repulsive potential followed by 120 conjugate gradient steps with van der Waals nonbonded potential, C^{α} restraints at 84 kJ(mol Å^2)$^{-1}$, $\Delta_F = 0{\cdot}05$ Å
3	Slow-cooling molecular dynamics, 3·7 ps, $T = 4000$-300 K (or lower initial temperature should instabilities occur), time step = 0·5 fs, T coupling, temperature decrement 25 K every 25 fs, $\Delta_F = 0{\cdot}2$ Å
4	Minimization, 80 conjugate gradient steps, $\Delta_F = 0{\cdot}0$ Å

described in Table 4 to refine macromolecular structures by the SA refinement method.

It should be pointed out that even with the best annealing protocol it will not be possible to correct grossly incorrect chain tracing or connectivities. This stresses the need for careful examination of the electron density maps even after SA refinement has produced a relatively low R factor.

Computations were carried out at the Pittsburgh Supercomputer Center of the National Science Foundation (ATB, Grant DMB 870007P), on the Convex C1 in the Center for Structural Biology at Yale University, and on a Cray XMP2 at Cray Research Inc. We wish to acknowledge the generosity and support provided by R. Eades, J. Mertz and E. Wimmer of Cray Research Inc.

References

ACHARYA, R., FRY, E., STUART, D., FOX, G., ROWLANDS, D. & BROWN, F. (1989). *Nature (London)*, **337**, 709-716.

BARFORD, D. & JOHNSON, L. N. (1989). *Nature (London)*, **340**, 609-616.

BERENDSEN, H. J. C., POSTMA, J. P. M., VAN GUNSTEREN, W. F., DINOLA, A. & HAAK, J. R. (1984). *J. Chem. Phys.* **81**, 3684-3690.

BOUNDS, D. G. (1987). *Nature (London)*, **329**, 215-219.

BRÜNGER, A. T. (1988*a*). *J. Mol. Biol.* **203**, 803-816.

BRÜNGER, A. T. (1988*b*). *X-PLOR Manual*, Version 1.5. Yale Univ., New Haven, USA.

BRÜNGER, A. T. (1989). *Acta Cryst.* A**45**, 42-50.

BRÜNGER, A. T. (1990). *Acta Cryst.* A**46**, 46-57.

BRÜNGER, A. T., BROOKS, C. L. III & KARPLUS, M. (1984). *Chem. Phys. Lett.* **105**, 495-500.

BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50-61.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458-460.

BURKERT, U. & ALLINGER, N. L. (1982). *Molecular Mechanics*. Am. Chem. Soc. Monogr. No. 177. American Chemical Society, Washington, DC, USA.

ERICKSON, J. W., ABAD-ZAPATERO, C., RYDEL, J. J. & LULY, J. (1987). *Acta Cryst.* A**43**, C-35.

FUJINAGA, M., GROS, P. & VAN GUNSTEREN, W. F. (1989). *J. Appl. Cryst.* **22**, 1-8.

GARRETT, T. P. J., SAPER, M. A., BJORKMAN, P. J., STROMINGER, J. L. & WILEY, D. C. (1989). *Nature (London)*, **342**, 692-696.

GOUAUX, J. E. & LIPSCOMB, W. N. (1988). *Proc. Natl Acad. Sci USA*, **85**, 4205-4208.

GROS, P., FUJINAGA, M., DIJKSTRA, B. W., KALK, K. H. & HOL, G. J. (1989). *Acta Cryst.* B**45**, 488-499.

HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252-270.

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931-935.

JONES, E. Y., STUART, D. I. & WALKER, N. P. C. (1989). *Nature (London)*, **338**, 225-228.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268-272.

KIRKPATRICK, S., GELATT, C. D. JR & VECCHI, M. P. (1983). *Science*, **220**, 671-680.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-349.

KURIYAN, J., BRÜNGER, A. T., KARPLUS, M. & HENDRICKSON, W. A. (1989). *Acta Cryst.* A**45**, 396-409.

METROPOLIS, N., ROSENBLUTH, M., ROSENBLUTH, A., TELLER, A. & TELLER, E. (1953). *J. Chem. Phys.* **21**, 1087-1092.

NAVIA, M. A., FITZGERALD, P. M. D., MCKEEVER, B. M., LEU, C.-T., HEIMBACH, J. C., HERBER, W. K., SIGAL, I. S., DARKE, P. L. & SPRINGER, J. P. (1989). *Nature (London)*, **337**, 615-620.

NILGES, M., CLORE, G. M. & GRONENBORN, A. M. (1988). *FEBS Lett.* **239**, 129-136.

NILSSON, L. & KARPLUS, M. (1986). *J. Comput. Chem.* **7**, 591-616.

PARKER, M. W., PATTUS, F., TUCKER, A. D. & TSERNOGLU, D. (1989). *Nature (London)*, **337**, 93-96.

PASTOR, R. W., BROOKS, B. R. & SZABO, A. (1988). *Mol. Phys.* **65**, 1409-1419.

ROBBINS, A. H. & STOUT, C. D. (1989). *Proc. Natl Acad Sci. USA*, **86**, 3639-3642.

SIELECKI, A. R., HAYAKAWA, K., FUJINAGA, M., MURPHY, M. E. P., FRASER, M., MUIR, A. K., CARILLI, C. T., LEWICKI, J. A., BAXTER, J. D. & JAMES, M. N. G. (1989). *Science*, **243**, 1346-1350.

SMITH, D. L., ALMO, S. C., TONEY, M. D. & RINGE, D. (1989). *Biochemistry*, **28**, 8161-8167.

SMITH, D. L., RINGE, D., FINLAYSON, W. L. & KIRSCH, J. F. (1986). *J. Mol. Biol.* **191**, 301-302.

VERLET, L. (1967). *Phys. Rev.* **159**, 98-105.

WEBER, P. C., OHLENDORF, D. H., WENDOLOSKI, J. J. & SALEMME, F. R. (1989). *Science*, **243**, 85-88.

WEISS, W. I., BRÜNGER, A. T., SKEHEL, J. J. & WILEY, D. C. (1990). *J. Mol. Biol.* **211**. In the press.

WLODAWER, A., MILLER, M, JASKÓLSKI, SATHYANARAYANA, B. K., BALDWIN, E., WEBER, I. T., SELK, J. M., CLAWSON, L., SCHNEIDER, J. & KENT, S. B. H. (1989). *Science*, **245**, 616-621.

YOON, C., PRIVE, G. G., GOODSELL, D. S. & DICKERSON, R. E. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6332-6336.