of fact, this is truly only the beginning in terms of the new chemistry that can be discovered within van der Waals clusters.

Thus, while the study of reactive processes in clusters may be utilized as a conceptual bridge between the "bimolecular" gas-phase collisions and the "solvated multimolecular" world of chemical reactions in solution, we sense that this bridge has in fact turned into a crossroads. Novel chemical reactions and unexpected dynamics can be observed that occur within the stabilizing environment of a molecular cluster by (1) providing novel chemical pathways (as in the cases of $\{NH_3\}_n^+$, $\{C_2H_4\}_n^+$, and $\{CH_3OCH_3\}_n^+$) and (2) stabilizing unstable reagents (as in $\{CH_3CHF_2\}_n^+$). New experimental directions we hope to employ within the near future are two-fold. The first involves exploring negative ion chemistry within clusters via techniques similar to those described in this paper. The second is the use of laser-induced fluorescence, to probe the internal states of the neutral radical product generated via these exoergic ion–molecule cluster reactions (NH_2 produced via reaction 3, for example).

This research was supported by the Office of Naval Research, which is hereby gratefully acknowledged. We also acknowledge the work of Gopal Vaidyanathan on the new olefinic cluster systems. This paper is dedicated to the memory of Richard Bernstein, whose boundless enthusiasm and inspiration will be sorely missed.

Molecular Dynamics Simulations with Experimental Restraints

AXEL T. BRÜNGER*

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

MARTIN KARPLUS*

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received February 2, 1990 (Revised Manuscript Received October 16, 1990)

Molecular dynamics simulations of biological macromolecules have many applications.[1] In addition to providing dynamical information (e.g., the magnitude and time scale of fluctuations; the relaxation after a perturbation such as ligand photodissociation), they can serve as a tool for exploring the conformational space of the molecule. One can distinguish between two different cases where this feature of molecular dynamics simulations is used. One of these requires the generation of Boltzmann-weighted ensembles to compute equilibrium thermodynamic properties and free energy changes in the system.[2] The other depends only on the generation of a representative sample of conformational space without the need for proper Boltzmann weighting. This has been used in studies of protein folding[3] and, more recently, has been applied to three-dimensional structure determination and refinement of macromolecules based on experimental information. It is the latter that is the focus of this Account.

Molecular dynamics simulations of macromolecules have a demonstrated utility in providing an atomic model that is in accord with the experimental information. A case in point is the determination and refinement of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance (NMR) spectroscopy.[4] Another example is the refinement of three-dimensional structures of macromolecules based on crystallographic diffraction data.[5] The experimental information is incorporated into the simulation by using a hybrid energy function,[6]

$$E^{\text{pot}} = E_{\text{chemical}} + wE_{\text{experimental}} \qquad (1)$$

where E_{chemical} is an empirical energy function that provides information about equilibrium covalent bonding geometry, vibrations, hydrogen-bonded interactions, and nonbonded interactions,[7] and $E_{\text{experimental}}$ consists of an expression for the difference between observed and computed data which is equal to zero if the model matches the data perfectly; the form used for

Axel T. Brünger was born on November 25, 1956, in Leipzig, Germany. He received his Physics Diploma degree at the University of Hamburg and his Ph.D. degree from the Technical University of Munich. He held a NATO postdoctoral fellowship and subsequently became a research associate at the Department of Chemistry, Harvard University. In 1987 he was jointly appointed Assistant Professor at the Department of Molecular Biophysics and Biochemistry, Yale University, and Assistant Investigator of the Howard Hughes Medical Institute. His research has focused on molecular dynamics studies of protein structure and function and on methods in protein crystallography and nuclear magnetic resonance spectroscopy.

Martin Karplus was born in Vienna, Austria, on March 15, 1930, and came to the United States at the beginning of World War II. He received his B.A. from Harvard College in 1950 and his Ph.D. from Caltech in 1953. He worked at Oxford University as an NSF postdoctoral fellow from 1953 until 1955, when he joined the faculty of the University of Illinois. In 1960 Karplus became a professor at Columbia University and, in 1966, a professor at Harvard University. He was named Theodore William Richards Professor of Chemistry at Harvard in 1979. Early in his career Karplus developed theories of magnetic resonance parameters, and he made fundamental contributions to the theory of reactive collisions. More recently, he has employed molecular dynamics simulation methods for the elucidation of the internal motions and thermodynamic properties of proteins and nucleic acids and for the study of enzymatic reactions.

(1) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. *Advances in Chemical Physics*; John Wiley & Sons: New York, 1988.
(2) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281–282.
(3) Levitt, M. *J. Mol. Biol.* **1983**, *170*, 723–754.
(4) Clore, G. M.; Gronenborn, A. M. *CRC Crit. Rev. Biochem.* **1989**, *24*, 479–563.
(5) Brünger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458–460.
(6) Jack, A.; Levitt, M. *Acta Crystallogr.* **1978**, *A34*, 931–935.
(7) Brooks, B. R.; Broccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

0001-4842/91/0124-0054$02.50/0

$E_{experimental}$ depends on the data being analyzed and is described in the appropriate application section below. The weight w scales $E_{experimental}$ such that the gradients of the chemical and experimental terms in eq 1 are of the same order. The goal of three-dimensional structure determination and refinement by molecular dynamics is to find a configuration of the macromolecule that corresponds to a minimum close to the global minimum of E_{total}; that is, with the assumption that the data and empirical energy function are sufficient to determine a unique structure, to find the structure that satisfies both the empirical and the experimental information. This is an example of a highly nonlinear optimization problem. In such a case, conventional gradient descent or least-squares optimization will fail unless the model is fairly close to the solution. Better convergence is achieved if, instead of restricting search directions to be against the energy gradient, the optimization is allowed to go uphill as well. Such a procedure, implemented as a Monte Carlo algorithm,[8] has been referred to as simulated annealing (SA).[9]

Monte Carlo and molecular dynamics simulations are two ways of generating conformations of the molecule that are consistent with a Boltzmann distribution appropriate to the specified temperature. For large biomolecular structures, the molecular dynamics algorithm is generally more efficient at generating equilibrium structures,[10] although Monte Carlo simulations are useful for sampling conformations distant from the starting conformation.[11] A molecular dynamics, rather than a Monte Carlo, algorithm was introduced to implement a SA procedure for refining protein structures with X-ray crystallographic diffraction data.[5] Similar techniques had already been used to determine and to refine three-dimensional structures of proteins and nucleic acids derived from solution NMR.[12,13] We refer to all these techniques simply as SA in spite of the fact that various other terms are used in the literature, such as restrained molecular dynamics[12-14] or dynamical simulated annealing.[4]

We focus in this Account on the utility of molecular dynamics simulations to carry out SA when they are employed in combination with experimental data. We briefly address special aspects of the simulations related to SA, including temperature control, the relative weighting between $E_{chemical}$ and $E_{experimental}$, and modifications of the empirical energy function $E_{chemical}$, followed by a discussion of SA strategies. Finally, we present applications of SA to structure determination and refinement based on NMR interproton distance data and to the crystallographic refinement of macromolecules. We presume that the reader is familiar with the methodology used in more standard applications of molecular dynamics. We therefore refer the reader to a comprehensive review of general molecular dynamics techniques that has been published elsewhere.[1]

Aspects of Refinement Methodology

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

$$m_i\frac{\partial^2 r_i}{\partial t^2} = -\nabla_i E^{pot} \qquad (2)$$

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 determined most commonly by numerical methods based on the Verlet algorithm.[15] In the present applications, the initial coordinates may be those for an extended chain, as in NMR structure determinations, or those of a low-resolution model based on isomorphous replacement data or a homologous protein, as in X-ray structure refinement. The initial velocities are assigned from a Maxwellian distribution at the appropriate temperature. In the SA applications considered here, a number of points that distinguish the calculations from ordinary molecular dynamics simulations need to be addressed.

Molecular Dynamics and Temperature Control. Control of the temperature during the molecular dynamics simulation can be achieved by uniform rescaling of the velocities v_i, i.e.,

$$v_i^{new} = v_i^{old}\sqrt{\frac{T}{T_{curr}}}$$

for all atoms i, where T is the target temperature and T_{curr} is the current temperature. Rescaling is performed typically every 25 to 50 integration steps. As the rescaling of velocities interrupts the standard Verlet algorithm, this method tends to produce oscillations of the temperature. An alternative temperature control procedure is to couple the system to a heat bath,[16]

$$m_i\frac{\partial^2 r_i}{\partial t^2} = -\nabla_i E^{pot} - m_i\gamma_i v_i\left(1 - \frac{T}{T_{curr}}\right) \qquad (3)$$

where γ_i specifies the "friction" coefficient for atom i. This method, which we refer to as T coupling in the following, suppresses excessive temperature oscillations. T coupling is preferable to Langevin dynamics in the context of SA since the friction term of the latter slows the atomic motions whereas the friction term in eq 3 is small if T is close to T_{curr}. The T-coupling method is being employed in the most recent applications of SA.[17-19]

Relative Weighting of $E_{chemical}$ and $E_{experimental}$. The weight w (eq 1) is needed in order to relate the artificial effective energy term $E_{experimental}$ to the "physical" energy term $E_{chemical}$. The choice of w can be critical: if w is too large, the refined structure shows large deviations

(8) Metropolis, N.; Rosenbluth, M.; Rosenbluth, A.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
(9) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671–680.
(10) Northrup, S. H.; McCammon, J. A. *Biopolymers* **1980**, *19*, 1001–1016.
(11) Bassolino, D.; Hirata, F.; Kitchen, D. B.; Kominos, D.; Pardi, A.; Levy, R. M. *Int. J. Supercomput. Appl.* **1988**, *2*, 41–61.
(12) Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. *J. Mol. Biol.* **1985**, *182*, 179–182.
(13) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3801–3805.
(14) Clore, G. M.; Gronenborn, A. M.; Brünger, A. T.; Karplus, M. *J. Mol. Biol.* **1985**, *186*, 435–455.
(15) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–105.
(16) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(17) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *239*, 129–136.
(18) Brünger, A. T.; Krukowski, A.; Erickson, J. W. *Acta Crystallogr.* **1990**, *A46*, 585–593.
(19) Fujinaga, M.; Gros, P.; van Gunsteren, W. F. *J. Appl. Crystallogr.* **1989**, *22*, 1–8.

from ideal geometry; if w is too small, the refined structure does not satisfy the experimental information. Jack and Levitt[6] proposed to set w so as to make the gradients of $E_{chemical}$ and $E_{experimental}$ of the same magnitude *for the current structure*. Thus, w would have to be periodically readjusted during the course of the refinement. We have found it possible to choose w so that no adjustment is needed except at the very final stages of refinement.[20,21] Our method consists of a short molecular dynamics run with w set to 0, followed by a comparison of the magnitude of the gradients of $E_{chemical}$ and $E_{experimental}$ of the molecular dynamics structure. The value of w is selected, as in the Jack–Levitt approach, to equalize the gradient of $E_{chemical}$ and $wE_{experimental}$. The dynamics step is used to randomize the structure so as to avoid unrealistically low gradients that could occur if the initial structure were in a local minimum of $E_{experimental}$ or $E_{chemical}$. This method has been used for many SA refinements of crystal structures.[20,21] It appears that the success of SA refinements is not sensitive to small changes in w except for the very final stages of refinement.[22] For structure determinations and refinements of solution NMR structures, because of the much larger conformational space under consideration, there is no simple rule for determining w, and it is obtained by trial and error procedure.[4]

Modifications of the Empirical Energy Function. Several modifications[20,23,24] of the parameters of the empirical energy function are required to avoid unphysical conformation changes due to close contacts in the initial structure, or due to the effects of high temperature during SA. The force constants of the improper torsion angles that maintain the handedness of chiral centers and the planarity of aromatic rings are increased to prevent distortions of these angles that result from the use of standard parameters. Similarly, the force constant specifying the dihedral torsion angle ω for the peptide bonds of all amino acids except proline is increased to prevent formation of cis peptide bonds. The force constant for the proline ω angle is initially kept at a small value to allow cis–trans isomerizations; during the final stages of refinement, a standard value for this force constant is necessary to avoid large deviations from planarity. We found it necessary to apply these modifications even for room-temperature calculations (300 K) or for straight minimizations. Otherwise, crystallographic refinement produced distortions of idealized geometry in regions of crystal disorder. Similar problems occurred in solution NMR structures[4] since the interproton distance restraints affected the geometry in the vicinity of hydrogen atoms that are involved in nuclear Overhauser enhancement (NOE) measurements.

The techniques of SA refinement are not restricted to empirical energy functions.[17] Different choices of force fields tend to affect regions that are not well determined by the diffraction information, such as disordered regions of crystal structures or surface loops in solution NMR structures. Comparison of refinements with different forms for $E_{chemical}$ is thus a way of evaluating the precision of the final structure.

Simulated Annealing Strategies. Gradient descent[26] or least-squares minimization[25,26] methods restrict the search direction to be energetically downhill.[26] The principle of SA consists of allowing search directions to be energetically uphill, as well. In the Monte Carlo formulation, the probability of such an uphill move is e^{-E/k_bT}, where k_b is Boltzmann's constant and T is an effective temperature used as a control parameter. The success of SA is dependent on the "annealing" schedule defined as the sequence of temperatures and number of configurations sampled at each temperature during the search. The temperature is usually kept very high initially, and the system is then "annealed" by reduction of the temperature.[9] In other words a coarse search is carried out at a high temperature, and a local minimum is approached during the cooling stage. The local minimum is usually lower than one that can be reached by gradient descent methods. 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 values of T may have to be introduced if the barriers between local minima are large. Corresponding considerations apply to the implementation of SA by molecular dynamics, as described here.

In applying SA to specific problems, the cooling rate[27] is, a priori, an unknown parameter. For example, it was recently shown[18] that crystallographic SA refinement produces a lower R factor and better geometry when a slow cooling rate is used. Apart from the cooling rate, the relative weighting between $E_{chemical}$ and $E_{experimental}$ terms can be modified during SA. It is possible to introduce a classification scheme for the control parameters of SA.[18] For minimization problems involving a hybrid energy function, the different approaches include overall scaling of E^{pot}, scaling of partial terms of $E_{experimental}$, and scaling of partial terms of both $E_{experimental}$ and $E_{chemical}$.

Application to NMR Structure Determination and Refinement in Solution

Although NMR has long been applied to structure determinations of peptides and a wide range of organic molecules, the availability of high-field NMR spectrometers and the advent of two-dimensional NMR spectroscopy have led to significant progress in its application to macromolecules.[29] With the development of sequential resonance assignment strategies based on the delineation of through-bond and through-space (<5 Å) connectivities, it has become possible to obtain complete or virtually complete and unambiguous resonance assignments for small proteins and oligonucleotides and to derive a large set of approximate

(20) Brünger, A. T.; Karplus, M.; Petsko, G. A. *Acta Crystallogr.* **1989**, *A45*, 5–61.
(21) Brünger, A. T. *J. Mol. Biol.* **1988**, *203*, 803–816.
(22) Weis, W.; Brünger, A. T.; Skehel, J. J.; Wiley, D. C. *J. Mol. Biol.* **1990**, *212*, 737–761.
(23) Brünger, A. T. X-PLOR Manual, Version 1.5, Yale University, 1988.
(24) Kuriyan, J.; Brünger, A. T.; Karplus, M.; Hendrickson, W. A. *Acta Crystallogr.* **1989**, *A45*, 396–409.
(25) Konnert, J. H.; Hendrickson, W. A. *Acta Crystallogr.* **1980**, *A36*, 344–349.
(26) Press, W. H.; Flannery, B. P.; Teukolosky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986.
(27) Bounds, D. G. *Nature (London)* **1987**, *329*, 215–219.
(28) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *239*, 317–324.
(29) Enrst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Claredon Press: Oxford, 1986.

interproton distances by means of pre-steady-state NOE measurements.[30] These interproton distance estimates comprise the data for three-dimensional structure determination. For larger proteins, where the proton spectrum is not well resolved, heteronuclear (^{13}C, ^{15}N) NMR techniques and higher dimensional spectra can be used to assign the NOE cross peaks. Because of the limitation in the number and the range of the available NOE distances (generally <5 Å), the structure determination and refinement of a macromolecule is not straightforward. The short-range character of the NOE distance information implies that a globular structure, such as that of many proteins, can be determined more accurately than an extended structure, such as DNA.

Since, in general, the available NOE distances do not fully determine the three-dimensional structure, the experimental data has to be augmented by information about the geometry and nonbonded interactions of the macromolecules. In other words, a hybrid energy function such as that given in eq 1 is introduced and one tries to locate the global minimum. Several methods have been developed to determine one or more structures that are located close to the global minimum.[31] One class of methods operates in distance space[32] by employing methods of distance geometry. A second class of methods operates in torsion angle space through gradient descent minimization[33,34] or Monte Carlo simulation.[11] A third class of methods operates in Cartesian coordinate space. The most widely used method employs molecular dynamics to refine protein[12,14] or nucleic acid structures[35] or, more ambitiously, to determine protein structures starting from extended strands[13] or random atomic coordinates.[17] Energy minimization is also employed but provides useful results only for a structure very close to the minimum. Interproton distance restraints may be incorporated as biharmonic effective potentials[12,14,35] when the error ranges associated with the measured distances are small, or as well potentials[4] for larger error ranges. To improve the convergence behavior of the simulated annealing protocol, the well potential can be combined with a soft asymptote for large deviations between current and target distances.[57] Torsion angle restraints derived from coupling constant measurements[37] can also be incorporated in the form of well potentials.[36]

For distances involving either methylene protons without stereospecific assignments or methyl protons, single $(\langle r_{ij}^{-6}\rangle)^{-1/6}$ mean distances are used as this is the quantity that can be related directly to the experimental NOE. It is important that this quantity is heavily weighted toward the distance with the smaller value. An alternative approach is to compute geometric center averaged distances ($\langle r_i\rangle - \langle r_j\rangle$) and to increase the allowed distance range for the NOEs between the i and j groups of protons. This approach, which is equivalent to the pseudo-atom method,[30] has the advantage of avoiding local minima for alternate conformations of groups involving protons without stereospecific assignments, but clearly some information is lost. Recently, stereospecific assignments of methylene groups have become feasible,[38] in which case no averaging is required; this appears to result in a significant improvement of the final structures.[39,40]

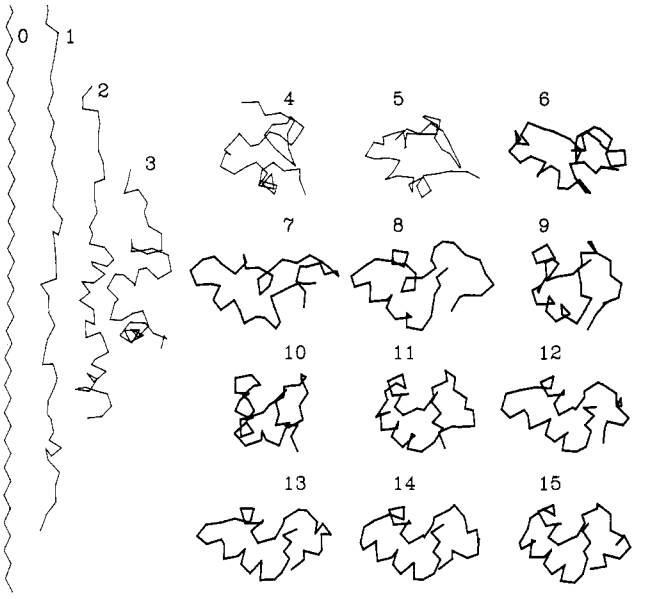

Figure 1. Pathway of SA for the structure determination of crambin starting with an extended β-strand using interproton distance data.[13] Simulated interproton distance data were used that could realistically be obtained by two-dimensional NOE spectroscopy. Snapshots of the C^α backbone are shown at 1-ps intervals. During the first 5 ps (thin lines), only interproton distance restraints between residues i, $i \pm 1$, $i \pm 2$, $i \pm 3$, $i \pm 4$, and $i \pm 5$ were taken into account, while all interproton distances were used during the following stages (thick lines).

Use of SA to refine folded structures obtained by distance geometry calculations is normally carried[28] out at a low temperature (300–1000 K). The weights for all terms of E_{total} are kept constant during the refinement, i.e., SA of the first type is used. If partial or complete folding or refolding of the structure is required, one can either reduce the weights for individual energy terms with large initial gradients or increase the simulation temperature, or both. The first such published SA protocol was of the second type.[13] It used a two step-procedure where during the first step only secondary structure interproton distance restraints were included, and in the second step the remaining interproton distance restraints were included, while the weights for all $E_{chemical}$ terms were kept constant. This protocol had a success rate of 50%. A number of new protocols have been developed to improve the SA success rate.[4] The most promising approach[17] consists of controlling the weights for all partial $E_{experimental}$ and $E_{chemical}$ terms (such as bond lengths, bond angles, and classes of interproton distances) individually to ensure that the energy barriers for each class of partial energies

(30) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(31) Braun, W. *Q. Rev. Biophys.* **1987**, *19* (3/4), 115–157.
(32) Crippen, G. M. *J. Comput. Phys.* **1977**, *24*, 96–107. Havel, T. F.; Wüthrich, K. *Bull. Math. Biol.* **1984**, *46*, 281–294.
(33) Braun, W.; Go, N. *J. Mol. Biol.* **1985**, *186*, 611–626.
(34) Billeter, M.; Havel, T. F.; Wüthrich, K. *J. Comput. Chem.* **1987**, *8*, 132–141.
(35) Nilsson, L.; Clore, G. M.; Gronenborn, A. M.; Brünger, A. T.; Karplus, M. *J. Mol. Biol.* **1986**, *188*, 455–475.
(36) Clore, G. M.; Nilges, M.; Sukumaran, D. K.; Brünger, A. T.; Karplus, M.; Gronenborn, A. M. *EMBO J.* **1986**, *5*, 2729–2735.
(37) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11–15.
(38) Hyberts, G.; Märki, W.; Wagner, G. *Eur. J. Biochem.* **1987**, *164*, 625–635.
(39) Discroll, P. C.; Gronenborn, A. M.; Clore, G. M. *FEBS Lett.* **1989**, *243*, 223–233.
(40) Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1988**, *204*, 675–724.
(41) This footnote was deleted on revision.

are lower than the kinetic energy of the atoms (SA of the third type). SA with NOEs has also been combined with distance geometry calculations to obtain the best features of both methods.[28]

Of considerable interest is the result that the radius of convergence of SA with NOEs is quite large. Figure 1 shows the pathway of SA with NOEs of crambin starting from an extended β-strand.[13] The conformational changes are very rapid. At the end of the first 2 ps of SA, the secondary folding is essentially established while the molecule is still in an extended conformation; at 9 ps, the tertiary structure is approximately correct. Convergence to the final structure is achieved rapidly when the weight w (eq 1) is raised to a sufficiently high level to ensure that all the restrained distances lie within their specified error ranges.

A very severe convergence test of SA with NOEs was performed by starting from a random array of atomic coordinates.[17] The SA protocol used in this case was of the third type, i.e., weights on classes of restraints were controlled individually to ensure that all energy barriers are below the kinetic energy of the system. It was possible, for example, to correct a misfolded structure of crambin which consisted of a mirror image of the crystal structure.[42]

The accuracy of NMR-structure determination is at present assessed by using different refinement protocols and/or different starting conditions that generate a family of conformers, each of which satisfies the experimental as well as the chemical information. It is then argued that the smaller the spread of the conformers, the more accurate is the structure.[4,43] However, it is necessary to assume that the sampling of conformational space is adequate; that is, a representative set of structures that satisfy the experimental information has to be sampled. The 74-residue protein α-amylase inhibitor tendamistat[40] and the 43-residue protein BDS-I[39] are examples of well-determined solution NMR structures including stereospecific assignments and measurements of coupling constants. For the α-amylase inhibitor, the average atomic root-mean-square difference between the individual structures and the mean structure is 0.57 Å for backbone atoms and 1.01 Å for all atoms. For BDS-I, the corresponding values are 0.67 and 0.90 Å. In marked contrast, the 1–29 fragment of human growth hormone releasing factor in trifluoroethanol shows root-mean-square differences for backbone atoms of up to 6 Å between individual conformers[44] determined by SA with NOEs and distance geometry (Figure 2). The structures satisfy all information available from NOE distance data and the empirical energy function. Although the NOE data are satisfied by unique secondary structural elements, they are not sufficient for global convergence to a unique three-dimensional geometry because no long-range NOEs were observed. The reason for the absence of observable tertiary NOEs could be either that the structures are extended, so that distances are too large, or that the NOE data set represents an average over a set of distinct folded conformations.

(42) Nilges, M.; Gronenborn, A. M.; Clore, G. M. *Proceedings of the 1989 Daresbury Study Weekend.*
(43) Wüthrich, K. *Acc. Chem. Res.* **1989**, *22*, 36–44.
(44) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. *Protein Eng.* **1987**, *1*, 399–406.

Application to Crystallographic Refinement of Macromolecules

The goal of protein crystallographic structure determination, as of solution NMR structure determination, is to obtain a three-dimensional atomic model of the macromolecule. The models are derived from electron density maps that are computed by using the observed structure factor amplitudes of the crystal and phase information obtained, for instance, through multiple isomorphous replacement.[45] The accuracy of the initial atomic model is usually limited, partly because the phase information is insufficient to show atoms individually and partly because of errors in the initial model building. The limited accuracy is insufficient for the full understanding of the chemistry of the macromolecule, and it is therefore important to improve the atomic models of macromolecules by crystallographic refinement.[46] Moreover, theoretical studies are best done by starting with refined high-resolution structures.

Conventional refinement of the X-ray structure of macromolecules involves a series of steps, each of which consists of a few cycles of least-squares refinement[25] with geometric and internal packing restraints that are followed by refitting the model structure to difference electron density maps with interactive computer graphics.[47] With currently available computing power, manual adjustment of the atomic positions using computer graphics to display the model in the electron density maps is the rate-limiting step in the refinement process. It can take several months to even years, so that there is a need for automation or at least reduction of the manual steps.

In X-ray crystallography as in NMR, the aim is to minimize E^{pot} (eq 1). For X-ray refinement, $E_{\text{experimental}}$ is the crystallographic residual, which is defined as the sum over the squared differences between the observed ($|F_o(h,k,l)|$) and calculated ($|F_c(h,k,l)|$) structure factor amplitudes. Other terms that may be included in E_{total}, in addition to E_{chemical}, include phase restraints or crystal packing terms.[21] During the final stages of refinement, ordered solvent molecules and alternate conformations for some atoms or residues in the protein may be introduced.

Several algorithms have been developed over the past 20 years to accomplish crystallographic refinement of macromolecules.[5,6,25,26,48,49] These algorithms can be generally classified into restrained least-squares procedures, minimization procedures, and the recently developed SA procedure. All of these procedures are being used by a number of crystallographic groups; we do not discuss their relative merits here.

Crystallographic SA refinement consists of minimizing E_{total} using molecular dynamics.[5] E_{chemical} is an empirical energy function[7] with certain modifications adapting it to use with high-temperature dynamics (see above). It has been shown[5,20,21,24] that crystallographic SA refinement starting from initial models obtained by multiple isomorphous replacement (MIR) or molecular replacement produces a structure (without human in-

(45) Watenpaugh, K. D. *Methods Enzymol.* **1985**, *115*, 227–234.
(46) Jensen, J. H. *Methods Enzymol.* **1985**, *115*, 227–234.
(47) Jones, T. A. *J. Appl. Crystallogr.* **1978**, *11*, 268–272.
(48) Sussman, J. L.; Holbrook, S. R.; Church, G. M.; Kim, S. H. *Acta Crystallogr.* **1977**, *A33*, 800–804.
(49) Tronrud, D. E.; Ten Eyck, L. F.; Matthews, B. W. *Acta Crystallogr.* **1987**, *A43*, 489–500.

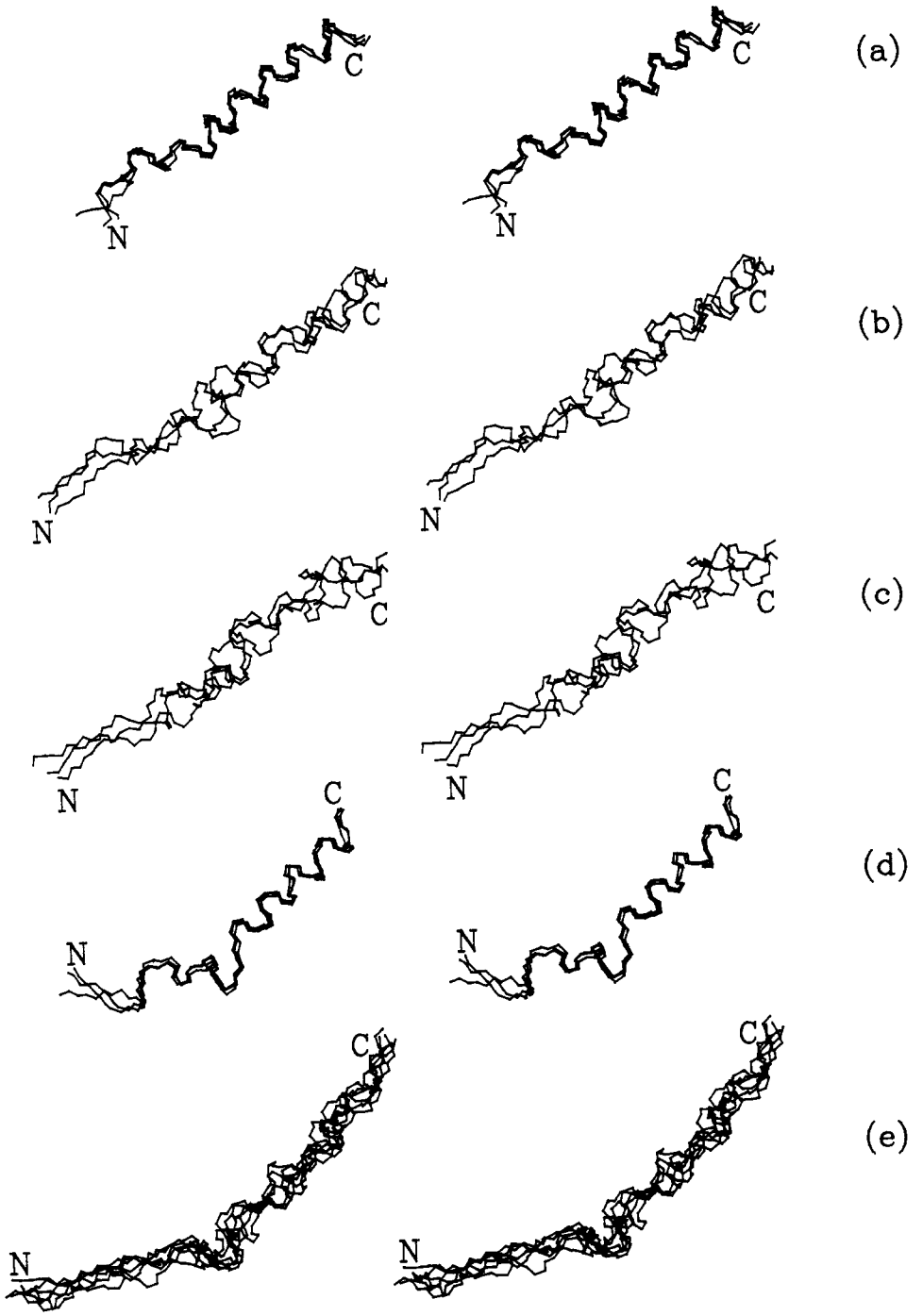

Figure 2. Stereo pictures showing best-fit superpositions of solution conformations[44] of the 1–29 fragment of human growth hormone releasing factor. Only the backbone (C, C^{α}, N) is shown: (a) SA structures starting from an extended α-helix, (b) SA structures starting from an extended β-strand, (c) SA structures starting from a polyproline helix, (d) SA structures starting from a mixed α/β structure, and (e) distance geometry structures using DISGEO.[32]

tervention) that has an overall R factor, geometry, and quality of fit of the model to the electron density map significantly improved in comparison with the use of conventional refinement programs without manual refitting of the structure to the electron density. This improvement is accomplished after a relatively short high-temperature SA refinement[20,21] or longer low-temperature SA refinements.[19] Slow cooling SA refinement[18] yields even better results but is computationally more expensive.

The essential element in the SA refinement is its larger radius of convergence, relative to other refinement programs. This is due to the SA annealing protocol which permits the model to search over a larger region of configuration space. SA refinement is able to move side-chain atoms by more than 2 Å and, in some cases, change backbone conformations or flip peptide bonds. Figure 3 shows two representative portions of the enzyme aspartate aminotransferase where SA refinement (which consumed 4 CRAY-XMP CPU h in this case[21]) has essentially converged to a refined structure obtained by several cycles of manual rebuilding and PROLSQ refinement. In the case of His-193 (Figure 3a), the imidazole ring has undergone a 90° rotation around the χ_1 bond during SA refinement. The rotation was accompanied by significant

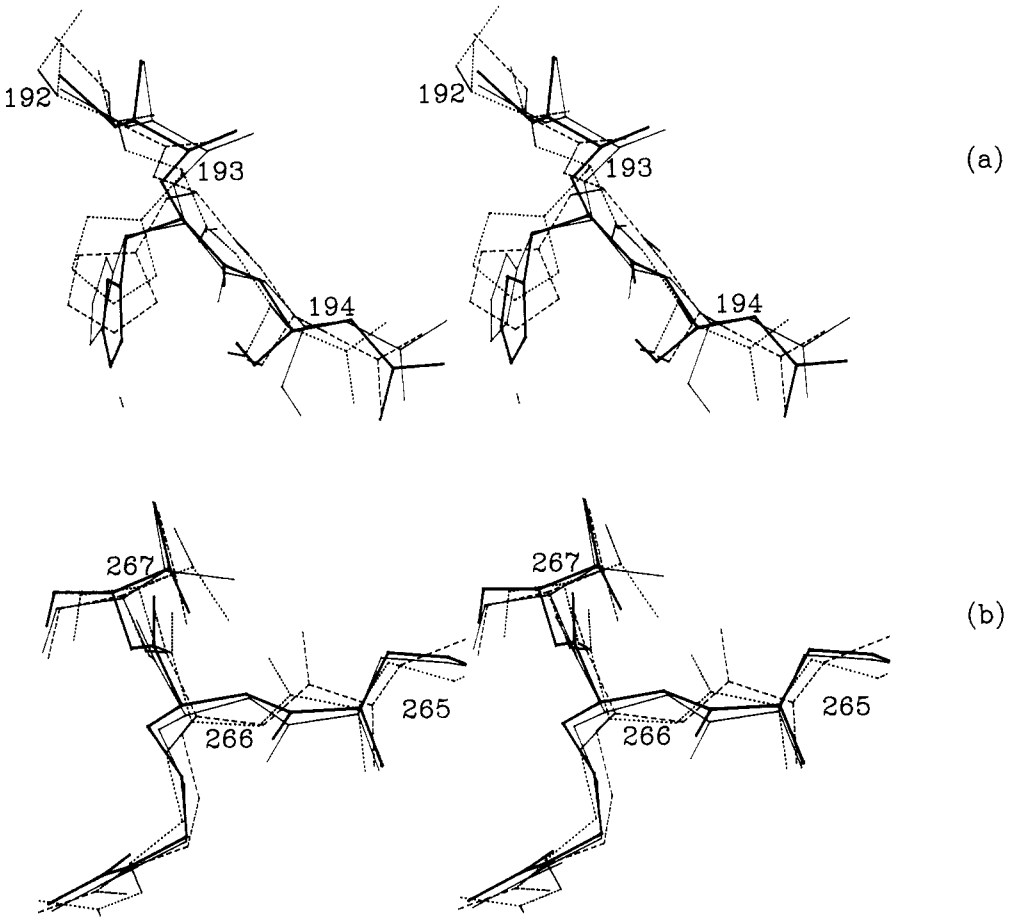

Figure 3. Stereo pictures showing segments of the K258A mutant of *Escherichia coli* aspartate aminotransferase:[56] (a) Cys192, His193, Asn194; (b) Glu265, Arg266, Val267. Superimposed are a manually refined structure (thick lines), the SA-refined structure (thin lines), the restrained least-squares refined structure (broken lines), and the initial structure (dotted lines). All refinements[21] started from the same initial structure obtained by fitting to a multiple isomorphous replacement map.

structural changes of the backbone atoms, and the SA-refined structure converged to the manually refined structure in this region. In the case of Glu-265 (Figure 3b), SA refinement has flipped the peptide bond by 180°, similarly to the result obtained by manual refitting. These structural changes were not obtained when restrained least-squares refinement without refitting was done. SA refinement in its present form is most effective for structures with errors in the atomic positions on the order 2–3 Å. For starting structures of significantly better quality, the benefit of using SA refinement relative to conventional refinement is less clear.

It should be noted that SA refinement is not able to correct chain traces with large errors (Figure 4). In this case, as in others[24] involving sequence errors, the improvement of the final SA-refined map is sufficient to simplify the model refitting in this region, thus reducing the overall amount of human intervention required during crystallographic refinement.

In spite of the success of the SA methodology, manual examination of the electron density maps at various stages of the refinement is important. It is essential for the placement of surface side chains and solvent molecules and for checking regions of the protein where large deviations from idealized geometry occur.

Concluding Remarks

This Account demonstrates that molecular dynamics, in addition to its use for studying the dynamics and thermodynamics of macromolecules,[1] is an important addition to the arsenal of methods available to structural biologists working with X-ray crystallographic or NMR spectroscopic data. The utility of the SA approach is confirmed by its widespread use in structural laboratories even though it was introduced only five years ago for NMR[12] and three years ago for X-ray crystallography.[5] More generally, the hybrid energy function approach (eq 1) can be applied to any situation where fitting of an atomic model to measured data or other restraints is required. The methodology described here is being extended in a number of directions. One example is molecular replacement[50] to solve the crystallographic phase problem based on homologous or similar macromolecules. In this case "Patterson" refinements of a large number of the highest peaks of a rotation search are carried out. If the root-mean-square difference between the search model and the crystal structure is within the radius of convergence of Patterson refinement, the correct orientation can be identified by having the lowest value of E^{pot} after refinement. Several X-ray structures, which could not have been solved by conventional approaches, have already been solved by this method.[51,52,58] A second example

(50) Brünger, A. T. *Acta Crystallogr.* **1990**, *A46*, 46–57.
(51) Brünger, A. T.; Milburn, M. V.; Tong, L.; DeVos, A. M.; Jancarik, J.; Yamaizumi, Z.; Nishimura, S.; Ohtsuka, E.; Kim, S.-H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4849–4853.
(52) Brünger, A. T.; Leahy, D.; Fox, R. O. *J. Mol. Biol.*, submitted.
(53) Kuriyan, J.; Burley, S.; Brünger, A. T.; Karplus, M. Proteins, submitted.

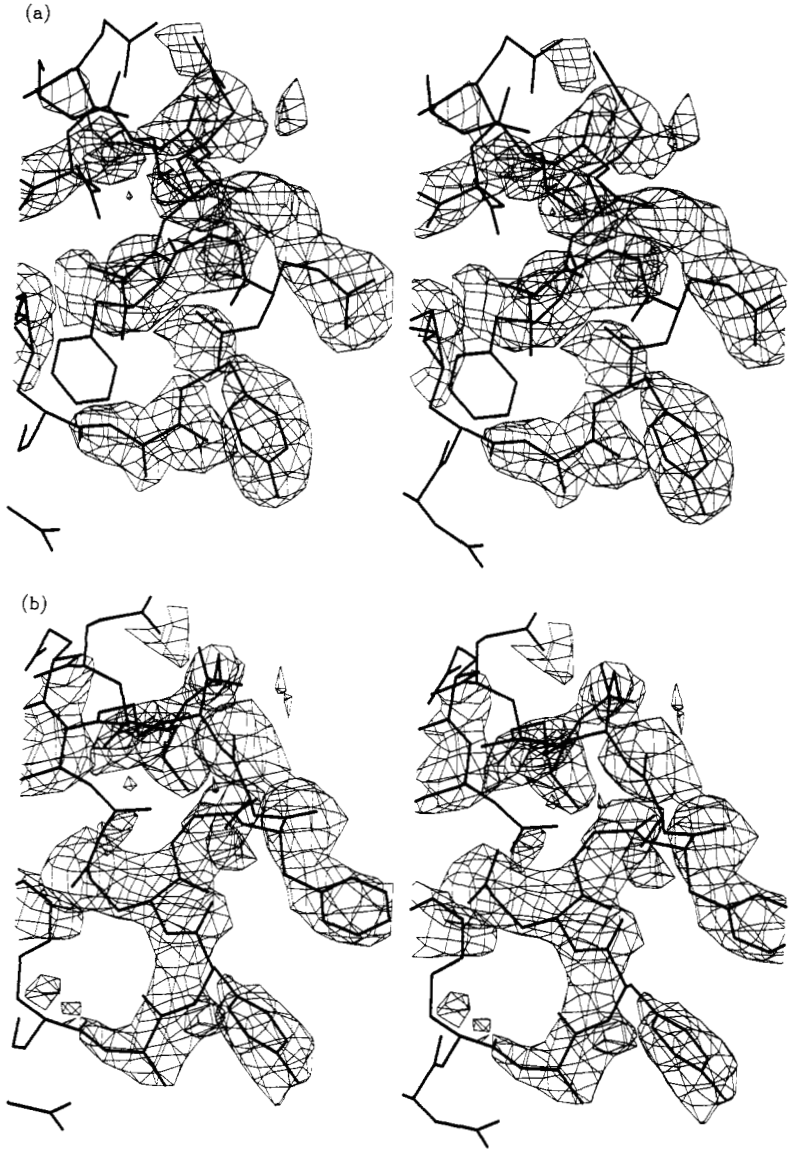

Figure 4. Stereo pictures showing a segment of aspartate aminotransferase,[56] which consists of residues 233–229 (Phe223, Ala224, Tyr225, Gln226, Gly227, Phe228, Ala229). (a) The SA-refined structure and (b) the manually refined structure are shown. The atomic models are shown as thick lines. Electron density maps are superimposed as thin lines. The maps were computed from $(2F_o - F_c)$ amplitudes using F_c phases corresponding to the atomic models at 2.8-Å resolution. The maps were contoured at 1.5σ where σ is the root-mean-square electron density in the unit cell. In the case of segment Phe223–Ala229, the backbone conformation of (a) the final SA-refined structure has not converged to (b) the manually refined structure. The chain trace of the initial structure contained a gross error which interchanged residues Gln226 and Phe228, presumably because of the bad quality of the MIR map. The intercharge of the two residues is too large a motion to obtain from a 3-ps SA refinement. However, the electron density map of (a) the final SA-refined structure clearly shows the density of the correct positions of (b) the *manually* refined structure. In marked contrast, the electron density map of the structure obtained by conventional restrained least-squares refinement is very difficult to interpret.[21]

involves the use of multiple structures in the refinement against high-resolution data. The resulting deviation between the two structures provides information concerning crystal disorder and the anisotropy of local motions.[53] A third example is the application of the hybrid energy function approach to the refinement of solution NMR structures against the observed NOE intensities from two-dimensional nuclear Overhauser correlation experiments.[54,55] In this case the effects of network relaxation and spin diffusion are taken into account by a complete matrix relaxation treatment.

(54) Yip, P.; Case, D. A. *J. Magn. Reson.* **1989**, *83*, 643–648.
(55) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1988**, *79*, 493–512.
(56) Smith, D. L.; Ringe, D.; Finlayson, W. L.; Kirsch, J. F. *J. Mol. Biol.* **1986**, *191*, 301–302.
(57) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27–38.
(58) Brünger, A. T. *Acta Crystallogr., Sect. A*, in press.

Registry No. Aspartate aminotransferase, 9000-97-9; human growth hormone releasing factor, fragment 1–29, 90830-28-7.