Biophysical Chemistry, 43 (1992) 221–238
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01663

Review

Exploration of the phase space of molecular systems: Assessment of established and new methods

Jin Li, Eric Platt, Bohdan Waszkowycz, Rodney Cotterill [1] and B. Robson [2]

Proteus Molecular Design Ltd., Proteus House, 48 Stockport Road, Marple, Cheshire, SK6 6AB (UK)

(Received 24 January 1992; accepted 3 February 1992)

Abstract

The problem of efficiently exploring molecular phase space stands at the centre of the major problems in biophysical chemistry, from calculation of probabilities, to protein folding, protein engineering, and drug design. The state of the art of *ab initio* approaches, (i.e. not depending fundamentally on experimental data) is assessed, and some new developments are proposed. Some of the techniques also lend themselves naturally to the inclusion of new classes of constraint, under new conservation laws, which can also exploit experimental data.

Keywords: Phase space; Conformation; Molecular dynamics; Free energy

1. Background

In this review we re-identify a key problem in the calculation by computer of a matter of considerable medical and industrial importance: the form and properties of biological molecules. The key problem relates particularly to the efficiency of the calculation of those all-important quantities, the free energies, and to the matter of whether they have been adequately evaluated. It lies at the heart of difficulties in calculating physicochemical properties, modelling protein conformation, engineering proteins, and designing drugs. We follow discussions of this with a review of the present state of the art, and finish with some recent advances. The recent advances in particular represent a more personal view based on developments in our own laboratories and are not intended to provide an exhaustive account of current work in the field.

What the authors have always had in common, which determines the flavour of the last half of this review, is not only an interest in better calculating free energies of systems but also an interest in new classes of molecular dynamics simulations of molecules. These depart from established types

Correspondence to: Dr. B. Robson, Proteus Molecular Design Limited, Proteus House, 48 Stockport Road, Marple, Cheshire SK6 6AB, UK. Tel: (061) 426-0191.

[1] Also, Department of Biophysics and Materials Science, Technical University of Denmark.

[2] Also, Department of Biochemistry and Molecular Biology, University of Manchester, UK, and Department of Biophysics and Material Science, Technical University of Denmark.

of simulation in a fundamental way. The "Newtonian rules" are broken in order to more efficiently search the conformational possibilities of a molecule. To consider why this bizarre departure should be necessary one must first take a look at what the problem is, and why the current approaches are accepted.

First principle or ab initio calculations.

A key feature of early twentieth century science was the realisation, usually attributed to Dirac, that experimentation was not the only route to the discovery of the physicochemical properties of systems. In principle at least, such properties might be determined by calculation from fundamental theoretical principles. The difficulty in practice resides in the complexity of the calculations which have to be performed. In modern terms, this essentially means the speed of the calculation in the computer. In part, this question of speed (in order to obtain results in reasonable time) is a question of levels of approximation used, and, relatedly, the quality of parameters used as input in empirical or semi-empirical approaches. However, these problems are not insurmountable at least to the extent that useful studies can be performed. For small organic molecules, these aspects represent the bulk of the difficulties, and small organic molecules are routinely used to assess the state of the art in regard to adequate levels of approximation, and the development, testing and refinement empirical and semi-empirical parameters.

Specific problems in biomolecular calculations.

Biological molecules and particularly biopolymers such as proteins provide a further, and in some ways more fundamental challenge. If it was simply a matter of the size of the molecule (specifically, time for calculation of the energy of a single conformation rises as the square of the number of interacting atoms) then a complete study of the complex biomolecules would be obtainable in reasonable amounts of computer time. However, it is also a matter of the escalating complexity of the *phase space* which underlies the physicochemical properties of the molecule. The phase space of a molecule is that mathematical space in which a single point represents a complete description of the positions and velocities of all considered particles in the system, and a single trajectory for such a point represents the history of the molecular system.

From this trajectory or consideration of its approximate description (such as density of trajectories in a given region of phase space, and particularly the *ergodic approximation* which we consider below) the tools of statistical mechanics can extract a variety of interesting properties. These include most obviously the free energy, enthalpy, entropy, heat capacity, and any other observable average properties such as average Nuclear Overhauser distances, circular dichroism spectra, viscometric and light-scattering properties. However, any property which is an average over population or a single molecule in time can be evaluated in principle, and one may even calculate biological properties such as catalytic activity and pharmaceutical potency and toxicity, providing, of course, that the complex formula relating properties to conformation can be specified. In any event, the evaluation of the free energy itself already plays an important contemporary role in drug design and protein engineering by comparing the calculated binding strength of different agonists and antagonists, and in calculating the preferred conformations of proteins, nucleic acids, sugars, lipids and other biomolecules.

Phase space complexity.

Where then is the problem, or more precisely, wherein lies the complexity of the phase space? There are most generally $6N - 12$ dimensions to this phase space. The $6N$ relates to the three Cartesian coordinates (x, y, z) to describe the position of each of N atoms, plus three coordinates P_x, P_y, P_z to describe the momenta of the particles conjugate to the positional coordinates (the -12 relates to the changes of the net position and overall velocity of the system). It is worth noting that neglect of the conjugate momenta leads to a reduced phase space well known to stereochemists as the conformational space.

This $3N$-6 dimensional space is already complex since it defines the potential energy surface, that is, the potential energy as a function of $3N$-6 parameters. Below, we associate such a specific function of the phase space coordinates with the mechanical properties of the systems, as opposed to the thermal properties such as enthalpy which are based on statistical mechanical averaging. For a complex biomolecule such as a protein, this potential energy surface is not a simple smooth monotonic surface, but is filled with many maxima, minima and saddle points. Collectively associated with these features in the fuller phase space description are conceptually even more monstrous features: topologically complex and not necessarily continuous isoenergy manifolds, these being regions of phase space in which the point describing the molecular system can move so as to preserve its total (potential plus kinetic) energy under Hamiltonian (isolated, energy conserving) conditions. The manifold description is at least relatively simple in the region of a potential energy minimum too deep to allow escape without borrowing further energy. It is a hyper-dimensional torus (ring doughnut). The trajectory in phase space would, for an isolated energy conserving system, be confined to winding round this torus. However, in most interesting applications excepting the studies of the vibrational modes of a specified locally stable or metastable state, the trajectory is wound round the surface which is topologically far more complex. The point is, that there is in general no means of predicting such a trajectory even in gross terms such that the observable behaviour of the system can in turn be predicted. The behaviour has to be followed by simulating it in what amounts to, in effect, an experiment within the computer. An example is simulating the folding up of a protein to predict its final three dimensional structure, a notoriously difficult problem which exemplifies that the pursuit of the trajectory through phase space is a task which rapidly becomes more formidable as the molecular system becomes bigger and the manifolds more complex.

Our knowledge of a physical system can be categorised as mechanical properties and thermal properties. The mechanical properties are the explicit functions of the phase space coordinates of a system, for example, the potential energy $V(q^N)$, or the virial $B(q^N) = \sum r_i F_i$. A number of useful thermodynamic properties of a system, for example, the internal energy E, the pressure P and the heat capacity C_v, can be calculated by the ensemble or time average of the mechanical properties, which are usually readily available in molecular dynamics (MD) simulation. However, the thermal properties of a system can not be obtained directly from a typical simulation. The thermal properties of a system depend on the total volume of its phase space which is accessible to the system's condition. Due to the finite time and speed of computer simulation, and the limitations in the simulation algorithms, the sampling over the total phase space is almost an impossible task. The practical researchers have been concentrating on the improvement of sampling over representative regions of a system's phase space. The entropy S and free energy F of a system are typical examples of thermal properties.

Although there are apparent difficulties in estimating free energy of a system, many research efforts have been put into it in the past decade [1–3], and as the speed of computational facilities have improved many-fold, some practical methods for estimating free energy have been established and proved to give indicative results for experimental work. The following sections give a brief review of these techniques.

2. Free energy estimation methods

2.1 Thermodynamic integration [4–7]

In statistical mechanics, the Helmholtz free energy A for a molecular system is calculated by

$$A = -k_B T \ln \frac{Z}{V} \tag{1}$$

where Z is the canonical ensemble configurational integral

$$Z = \int \cdots \int e^{-\beta V(r_1,\ldots,r_N)} \, dr_1, \ldots, dr_N \tag{2}$$

where $\beta = 1/k_B T$ and $V(r_1, \ldots, r_N)$ is the potential energy of the system then the free energy difference ΔA between state 0 and 1 is

$$\Delta A = A_1 - A_0 = -k_B T \ln \frac{Z_1}{Z_0} \tag{3}$$

where Z_0 *and* Z_1 are the configurational integrals of the systems at the states 0 and 1 respectively. The idea central to all the methods used so far for the estimation of free energy difference ΔA is how to change a system from state 0 to state 1, and obtain ΔA as a result of accumulating data during this process. The idea of a coupling parameter has been used for this purpose. According to this idea, the potential energy of the system V is written as a function of a coupling parameter λ, $V(r, \lambda)$, and the variation of λ from 0 to 1 smoothly converts the system configuration V_0 to V_1. Hence, the free energy A of the system can be described as the function of the coupling parameter λ

$$A(\lambda) = -k_B T \ln Z(\lambda) \tag{4}$$

The thermodynamic integration (TI) method then integrates $A(\lambda)$ over λ to calculate ΔA

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} d\lambda \tag{5}$$

Next, it is necessary to express this integral as the ensemble average of the mechanical property of the system which is easy to evaluate during a MD or MC simulation. Because $A(\lambda) = -k_B \ln Z(\lambda)$, then

$$\frac{\partial A(\lambda)}{\partial \lambda} = -k_B T \left[\frac{\partial \ln Z(\lambda)}{\partial \lambda} \right] = \frac{-k_B T}{Z(\lambda)} \frac{\partial Z(\lambda)}{\partial \lambda} \tag{6}$$

Using eq. (2), we have

$$\frac{\partial Z(\lambda)}{\partial \lambda} = -\beta \int \cdots \int \frac{\partial V(r, \lambda)}{\partial \lambda} \times e^{-\beta V(r, \lambda)} \, dr_1, \ldots, dr_N \tag{7}$$

Substituting eq. (7) into (6),

$$\frac{\partial A(\lambda)}{\partial \lambda} = \frac{1}{Z(\lambda)} \int \cdots \int \frac{\partial V(r, \lambda)}{\partial \lambda} \times e^{-\beta V(r, \lambda)} \, dr_1, \ldots, dr_N \tag{8}$$

Or more concisely,

$$\frac{\partial A(\lambda)}{\partial \lambda} = \left\langle \frac{\partial V(r, \lambda)}{\partial \lambda} \right\rangle_\lambda \tag{9}$$

For the free energy difference between state 1 and state 0,

$$\Delta A = \int_0^1 \left\langle \frac{\partial V(r, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{10}$$

With this expression, it is then possible to implement the calculation numerically during a MD or MC simulation. The integration variable λ can be a system's thermodynamic variable, such as temperature, or structural parameters which can introduce conformational changes to the system.

The implementation of TI in practice is as following: various simulations are carried out at different discrete values of λ over the interval [0, 1] to determine the ensemble average values for the term in eq. (10). Then numerical integration for these values is done over the λ values to yield the free energy difference ΔA. Usually, simulations are carried out in both forward and backward direction with respect to the state change (i.e., 0 to 1 and 1 to 0), and the corresponding free energy differences ΔAs are calculated to compare the accuracy of simulation. The difference (hysteresis) between two ΔAs represent the statistical uncertainty in the numerical integration over the configurational space. In order to obtain a more accurate result, the incremental $\Delta\lambda$ should be very small so that the process of configurational change from state 0 to state 1 or vice versa is smooth. Usually configurational sampling near the two end points (where $\lambda = 0$ and $\lambda = 1$) can become very difficult. This difficulty may be overcome by using a further smaller $\Delta\lambda$.

There has been a number of successful application of TI. For example, the ΔA calculated for

water, -18.0 ± 0.3 kJ/mol, agrees well with the experimental value of -24.0 kJ/mol, and the agreement between respective entropies (-59.9 ± 0.4 kJ/mol and -58.4 kJ/mol) is even better. However, the realisation of TI requires knowledge of certain nuances, for instance, the continuous distribution of discrete λ is not always optimal when integrating through summation. The insufficient sampling could easily lead to poor convergence.

2.2 Perturbation method [8–10]

The calculation of free energy difference between state 0 and 1 can also be approached from the following perspective:

$$\frac{Z_1}{Z_0} = \frac{\int \cdots \int e^{-\beta V_1(r_N)}\, dr_N}{\int \cdots \int e^{-\beta V_0(r_N)}\, dr_N} \tag{11}$$

As always, the next step is to transform this expression into some sort of ensemble or time average of a system's mechanical property. Equation (11) can be seen as

$$\frac{Z_1}{Z_0} = \frac{\int \cdots \int e^{-\beta V_1(r_N)}\, e^{\beta V_0(r_N)}\, e^{-\beta V_0(r_N)}\, dr_N}{\int \cdots \int e^{-\beta V_0(r_N)}\, dr_N}$$

$$= \int \cdots \int e^{-\beta \Delta V(r_N)} P_0(r_N)\, dr_N \tag{12}$$

where $\Delta V(r_N) = V_1(r_N) - V_0(r_N)$ is the potential energy difference between reference state 0 and perturbed state 1, and P_0 is the Boltzmann probability function for state 0:

$$P_0 = \frac{e^{-\beta V_0(r_N)}}{\int \cdots \int e^{-\beta V_0(r_N)}\, dr_N} \tag{13}$$

It is clear that eq. (12) is already an ensemble expression of potential energy difference between the two states:

$$\frac{Z_1}{Z_0} = \langle e^{-\beta \Delta V(r_N)} \rangle_0 \tag{14}$$

Substituting eq. (14) into (3), we have

$$\Delta A = -k_B T \ln \langle e^{-\beta \Delta V(r_N)} \rangle_0 \tag{15}$$

or, perturbing from state 1 to state 0

$$\Delta A = -k_B T \ln \langle e^{\beta \Delta V(r_N)} \rangle_1 \tag{16}$$

Equations (15) and (16) are the core of the perturbation method. It is valid only when state 1 and 0 differ very little. In fact, a necessary condition is that $\Delta A < 2k_B T$. In practice, this obviously poses a restriction on the direct use of this method. The following two methods seek to resolve this restriction by using a series of small perturbations to change the system from state 0 to state 1 or vice versa.

2.2.1 Window method

For real systems, the state change usually involves much greater free energy change than $2k_B T$. It is then necessary to differentiate this process of state change into a sequence of small perturbations, and the above perturbations method can be used to calculate the free energy difference, that is, the whole region of the states 0 and 1 is divided into discrete intervals so the ΔA for each interval is less than $2k_B T$. These discrete intervals are called thermodynamic windows. The overall free energy difference is then obtained by summing over all the intervals

$$\Delta A = \sum_{i=0}^{n-1} \Delta A_i(\lambda_i \rightarrow \lambda_{i+1}) \tag{17}$$

where ΔA_i is the free energy difference between intermediate states i and $i+1$, which are constructed by mixing up the initial system state 0 and the final state 1

$$H_{\lambda_i} = \lambda_i H_0 + (1 - \lambda_i) H_1, \qquad 0 \le \lambda_i \le 1 \tag{18}$$

where H is the Hamiltonian for the system, λ is the perturbing parameter. Usually, the intermediate state for a molecular system is generated by mixing the interaction potential parameters or coordinates. The direct use of eq.(18) is generally known as "double-ended" sampling, which is in effect a direct summation over the passes λ_i to λ_{i+1}. This may not be the optimal way. It has

been suggested that the free energy differences for the passes λ_i to λ_{i+1} and λ_i to λ_{i-1} are both calculated when averaging over the ensemble for the i^{th} state. This procedure is know as "double-wide" sampling, and it has been shown to reduce the computation by half. The basis of this procedure is the equality of the free energy differences for the state changes λ_i to λ_{i-1} and λ_{i-1} to λ_i. The equality is also usually used to check the accuracy of simulation.

2.2.2 Slow growth method

The slow growth method is a variant of the window method. Instead of dividing the whole region of state change into a set of discrete windows, the slow growth method treats the coupling parameter λ as a function of simulation time, and changes continuously from 0 to 1. In practice, the slow growth method differs from the window method in that it changes λ in very small increments. In fact, this interval can be made sufficiently small to allow the strictest calculation of ΔA. The inevitable shortcoming of the slow growth method is the slow convergence.

As can be seen from the above discussions, the key of the perturbation method is a small system state change. This can either be achieved by the window method or slow growth method, or by designing a thermodynamic cycle. In molecular simulation, the state change is usually achieved by changing the force field parameters to represent a molecular structural change.

The perturbation method is probably the most widely used free energy estimation procedure in molecular simulation and solution study. Its application ranges from the study of solute substitution in solution to the relative binding of different ligands to an enzyme. For example, the free energy difference for the transformation $CH_3OH \rightarrow CH_3CH_3$ was calculated by the window method to be 28.0 ± 0.9 kJ/mol, which is in excellent agreement with the experimental value 29.0 kJ/mol.

2.3 Finite difference thermodynamic integration [11]

Finite difference thermodynamic integration (FDTI) is the combination of thermodynamic integration and the perturbation method (PM). Equations (15) and (17) are the basis for the perturbation method. Because of its conceptual simplicity, it has been widely adopted as the protein engineer's free energy calculation procedure. Unfortunately, the direct summation eq. (17) seems to lack the basic physical insight into the system that thermodynamic integration has, as expressed in eq. (5). The idea behind FDTI is how to combine the two together. From the basic principle of the calculus, we know that.

$$\int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} \mathrm{d}\lambda = \lim_{n\to\infty, \Delta\lambda\to 0} \sum_{i=0}^{n-1} \frac{\Delta A_i(\lambda_i \rightarrow \lambda_{i+1})}{\Delta\lambda} \tag{19}$$

Therefore, by choosing arbitrarily small $\Delta\lambda$ in the perturbation method, the integration in the thermodynamic integration method, eq. (5), can be achieved. In practice, FDTI is realised by calculating ΔA_i by the perturbation method, dividing ΔA_i by $\Delta\lambda$, and integrating this result over λ numerically. FDTI was used to calculate the free energy of aqueous hydration of acetone and dimethylamine. It was found that FDTI converges significantly faster than both TI and PM.

2.4 Umbrella sampling [12–14]

From the description presented in the above sections, it can be seen that transforming the calculation of the configuration integral into the ensemble or time average of a system's internal property is the key to the calculation of statistical mechanical properties as expressed by

$$\frac{Z_1}{Z_0} = \langle \mathrm{e}^{-\beta\Delta V(r_N)} \rangle_0 \tag{20}$$

However, the applicability of the equation is limited by the fact that, wherever V_0 and V_1 differ to a certain degree, the probability $P_0(V)$ of sampling a point in the configuration space of system 0 for which $\mathrm{e}^{-\beta\Delta V}$ is appreciable tends to be very small. New methods are needed to improve the sampling of part of configurational space where the product $P_0(V)\,\mathrm{e}^{-\beta\Delta V}$ is large by carrying out

a biased (non-Boltzmann) sampling. One of these methods is known as umbrella sampling (US).

The basic idea behind umbrella sampling is that the relative probability to sample a point is no longer based on $e^{-\beta\Delta V}$, but $w(\boldsymbol{r})\, e^{-\beta\Delta V}$, where $w(\boldsymbol{r})$ is a weighting function chosen to ensure that both the region where $P_0(V)$ is large and the region where $P_0(V)\, e^{-\beta\Delta V}$ has its maximum are adequately sampled. Therefore, the desired average in eq. (20) is related to the average obtained by using biased sampling by

$$\langle e^{-\beta\Delta V}\rangle_0 = \int \frac{e^{-\partial\Delta V_w}}{w} e^{-\beta\Delta V}\, d\boldsymbol{r}_N \Big/ \int \frac{e^{-\beta\Delta V_w}}{w} d\boldsymbol{r}_N \tag{21}$$

$$\langle e^{-\beta\Delta V}\rangle_0 = \left\langle \frac{e^{-\beta\Delta V}}{w} \right\rangle_w \Big/ \left\langle \frac{1}{w} \right\rangle_w \tag{22}$$

where the angular brackets with subscript indicate an average over the biased sampling. Equation (22) can also be written as

$$\langle e^{-\beta\Delta V}\rangle_0 = \left\langle \frac{e^{-\beta\Delta V}}{w} \right\rangle_w \langle w\rangle_0 \tag{23}$$

This equation shows clearly that, in order to get reliable results, w should be chosen such that $w\, e^{-\beta\Delta V}$ is reasonably large in the region of configuration space accessible to system state 1 (otherwise the first term on the right hand side of eq. (23) vanishes), but w should also have appreciable value wherever $e^{-\beta\Delta V_0}$ is large (otherwise the second term on the right vanishes). The name (umbrella sampling) for this method is based on the fact that $w\, e^{-\beta\Delta V_0}$ is designed to cover both $P_0(V)$ and $P_1(V)$.

Obviously, the key in umbrella sampling is the choice of the weighting function w. Unfortunately, there is no general guidance for this purpose. This is usually the criticism of umbrella sampling. In practice, satisfactory choices can be efficiently obtained by trial and error. Attempts to predict reliable general analytical forms for w have been unsuccessful. This is not surprising, because knowledge of such forms would imply the solution of the general problem of evaluating statistical mechanical partition functions.

However, this does not mean that umbrella sampling has not been used at all. In fact, the following transformation has been used to achieve better sampling.

$$w\, e^{-\beta\Delta V} = e^{-\beta(v_0+v_1)/2}$$

In other cases, the intuition about the density of state of configuration space as a function of V_0, is used to choose a form for w which guarantees a uniform a sampling of the relevant part of configuration space as possible. As the system size becomes larger, it becomes increasingly difficult to find a weighting function that does the job. This is indeed the down-side of the umbrella sampling method, i.e., it uses a non-physical weighting function, and it needs to alter the internal mechanism of the program to experiment on a different choice. However, we have studied several polymers (including proteins) in which the weighting is based on the conformational energy surfaces of the individual units, which are readily calculated and stored in advance. This has some meaning in terms of the dominant interactors seen in "Theta conditions", i.e. solvent conditions when interactions *between* units cancel *in toto*.

In such applications, the advantages of umbrella sampling are very attractive. Firstly, given a reasonable choice of weighting function w, the free energy difference ΔA can be obtained for a range of different Hamiltonians and temperatures. Secondly, umbrella sampling is a very general way to probe a system. Instead of $e^{-\beta\Delta V}$, any other functions $F(\boldsymbol{r}^N)$ can be investigated. For instance, consider the case where we are interested in the probability that the quantity $X(\boldsymbol{r}^N)$ has its value in the interval dx around x. Let us assume that this value of X is very unlikely in the original ensemble, it could correspond to an improbable molecular conformation. In that case, simple Boltzmann sampling will produce very poor statistics on the desired probability. But with umbrella sampling., the weighting function can be chosen such that the relevant region of configuration space is sampled adequately. The probability density to observe X around x is then given by

$$P_X(x) = \left\langle \frac{\delta(X-x)}{w} \right\rangle_w \Big/ \left\langle \frac{1}{w} \right\rangle_w \tag{24}$$

In this way, umbrella sampling can be used to estimate the probability of rare events, or to construct the Landau free energy associated with a particular type of order parameter fluctuations.

2.5 *Thermodynamic cycle*

From the above discussion, it can be seen that umbrella sampling is probably the most powerful method for achieving better sampling in configurational space. But due to its lack of rule for choosing the weighting function, it still only remains a 'theoretical darling' rather than a practical recipe. In molecular simulation, thermodynamic integration, perturbation method and their combination, i.e. result–finite difference thermodynamic integration are the major methods for free energy estimation. Both thermodynamic integration and the perturbation method can be implemented in either MD or MC. Naturally, most of the limitations existed for MD or MC apply to the free energy calculation as well. Reliable calculation of relative free energy requires adequate sampling of configurational space. For complex systems, it is difficult to know whether sampling is adequate or not. The usual hysteresis check used in free energy calculation is a necessary, but not a sufficient criterion for sampling adequacy. A small hysteresis indicates only that similar configurations have been collected in both forward and backward directions, but still many significant configurations may have not been sampled at all in both directions, perhaps because the perturbation is proceeding so fast that the system cannot adjust. In addition to configurational sampling problems, many other factors also affect the calculation, for example, truncation of long-range potential interaction, boundary condition effects, force field parameters, etc. One approach to circumvent these difficulties uses the so-called "thermodynamic cycles in free energy" calculation. The basis on which the thermodynamic cycle approach rests is the fact that free energy ΔA is a state function. This means that as long as a system is changed in a reversible way the change in free energy A will be independent of path. Therefore, along a closed path or cycle, $\Delta A = 0$. This result implies that there are two possible ways to obtain the ΔA for a specific process: one may calculate it directly by using the above method along a path corresponding to the process, or one may design a cycle of which the specific process is only a part and calculate the ΔA for the remaining part of the cycle. For example, in order to compare the binding of two different substrates S and S′ to an enzyme E, a cycle is designed as following

$$\begin{array}{ccc} \mathrm{E}+\mathrm{S} & \xrightarrow{\Delta A_1} & \mathrm{ES} \\ \Big\downarrow{\scriptstyle \Delta A_3} & & \Big\downarrow{\scriptstyle \Delta A_4} \\ \mathrm{E}+\mathrm{S}' & \xrightarrow{\Delta A_2} & \mathrm{ES}' \end{array}$$

The desired free energy difference is $\Delta\Delta A = \Delta A_2 - \Delta A_1$. Because of the thermodynamic cycle, $\Delta\Delta A = \Delta A_4 - \Delta A_3$. The calculation of $\Delta\Delta A$ may be more reliable than the calculations of individual As, as any error due to inadequate force field parameters, etc., will approximately cancel for similar pairs of paths. However, the sampling problems still remain.

3. Novel methods

The above techniques address sampling from the point of view of restarting "classical" simulations in a structured way from different points, or by imposing the view that a broad consideration of conformational space is not essential. In particular, the perturbation method assumes that the molecular system retains a conformation and behaviour which is similar to a high level of detail even after a chemical change to one of its components. Experimentalists have long perceived the innate conformational flexibility of molecules such as proteins and indeed the functions of proteins usually depend on that conformational flexibility. Specifically, protein molecules and other complex biomolecules are rarely if ever confined to simple quadratic potential energy wells in which they show simple periodic or quasi-periodic behaviour.

Can the underlying simulation algorithm itself be tailored to provide a more comprehensive treatment? Previously, the principle approaches

involving departure from reality "broke the rules" in regard to the unphysical nature of the force rather than in regard to the Newtonian laws themselves. As further information, either experimental data specific to the molecule or data pertaining to the general class of molecules, is often introduced. "Pseudo-energy" functions in energy minimisation or "pseudo-force" functions in molecular dynamics may represent the effect of this further information and replace the classical description of energy and force which are based on evaluation of interatomic interactions. To impart some degree of stereochemical reasonableness, such functions may in fact be (or imply) the sum of the classical energy function plus the new component as a "target function" or "penalty function". A simple and common example arises in energy refinement of X-ray coordinates of a protein, or in modelling a protein by reference to a homologous protein of known conformation [15,16]. Here one may use the function

$$F = aE + (1-a)\sum_{ij} d_{ij}^2, \quad 0 < a < 1 \qquad (25)$$

and vary the conformation Q in the direction downhill in the effective (internal) potential energy surface E as dictated by the pseudo force $\mathrm{d}F/\mathrm{d}Q$ [17]. The problem is that one must find an optimal value for the coefficient "a" because, since the components are not cometric, and since neither the X-ray coordinates nor the energy calculation are perfect, a mathematical "catastrophe" occurs in which one obtains a good energy with poor fit or vice versa.

In contrast, nature itself adheres well to certain properties without being explicitly "drawn" towards them by penalty or target functions. These are the conserved properties, notably total energy, total linear and angular momentum, and Liouville phase space volume. The problem of choosing the reasonable value of "a" does not arise, nor does the question of a catastrophe. The origin of these conservation phenomena or *conservation laws* is inherent in the equations of motion employed. This raises the possibility of changing the basic equations of motion to introduce novel conservation laws. We here develop as a starting point the approach reported by Cotterill and Madsen for a materials science application [18]. We have been the first to use this to study the conformational possibilities of molecules. This approach is colloquially known as 'rush dynamics". With the exception of our earlier brief report concentrating on a small molecule [19], no previous written account has been given of the application to molecular structures. Indeed other workers may have been discouraged from the application of the method to chemical compounds since it may not have been obvious that a method which changes the laws of motion for the atoms can do so while safely retaining the integrity of molecular structure.

Essentially, the Cotterill–Madsen technique in both conformational studies and in its original materials science applications [18] essentially causes the system to conserve potential energy in a computationally efficient manner. This is sometimes picturesquely referred to as "contour tracing", since the effect on an easily visualised two-dimensional potential energy surface would be to run round the one dimensional potential energy contour. This useful visual appreciation of the problem makes it clear that hunting out the initial contour level of choice will be an important first step, but it also gives a misleading impression of the freedom available to the system. A phase space of $6N$ dimensions (arising from the particle coordinates x, y, z and their conjugate momenta, for each particle of an N-particle system) has a $3N$ potential surface and the regions of constant potential energy is not in general a contour but a $3N$-1 dimensional manifold. For a typical small protein of approximately the size of hen egg while lysozyme (129 residues and approximately 2000 atoms), the manifold has some 6000 dimensions.

The importance of these present studies is seen by us to be primarily in further development of the method as we shall describe in detail elsewhere. Briefly, we have noted that imposition of the trajectory in phase space to lie on this manifold leaves open the possibility of introducing further artificial conservation laws. This is because the specific trajectory taken is specific but arbitrary as to its physical significance. Another choice of trajectory lying on the manifold

would, in general, serve the same purpose. A further conservation law of another type can be introduced so as to leave unperturbed the conservation of the potential energy. In fact, for a complex molecule such as a protein this allows the introduction of potentially many thousands of conservation laws, as will be discussed below.

Alternatively, though with the loss of some benefits of the above type of approach, the iso-potential energy function could in fact be an "iso-pseudo potential energy" function from the outset. This is to say that the potential surface results from the sum of energy and function value components as in eq. (25).

Finally, the Cotterill–Madsen approach provides the incentive to make up the jump to other potentially useful modifications of the Newtonian laws of motion which do not depend on potential energy conservation. In the present study, we consider the Cotterill–Madsen algorithm for the first time as applied to organic molecules, and consider its relationship with other classes of conservation laws. Particularly we note briefly that there is a novel class of momentum-conservation law which does not depend on conservation of potential energy.

3.1 Theory of rush dynamics

The Cotterill–Madsen method works by conserving potential energy, corresponding to a $3N - 1$ dimensional manifold in the $6N$ dimensional phase space of a system of N atoms. In effect, the normal force vector $(\mathrm{d}E/\mathrm{d}Q)$ is rotated maximally orthogonally with respect to the original direction. By definition such a rotation places the new vector in a direction of constant potential energy. In practice, it is the velocity vector which is rotated as follows:

$$\boldsymbol{r}_1 = \boldsymbol{r}_0 + \boldsymbol{v}_0 \, \mathrm{d}t \tag{26}$$

$$\boldsymbol{v}_1 = \boldsymbol{v}_0 - \nabla U \cdot \frac{\nabla U \cdot \boldsymbol{v}_0}{|\nabla U|^2} \tag{27}$$

$$\nabla U \cdot \boldsymbol{v}_1 = 0 \tag{28}$$

where $\boldsymbol{r}_0$ is the instantaneous configuration of the N particles with velocities $\boldsymbol{v}_0$; $\boldsymbol{r}_1$, $\boldsymbol{v}_1$ are the respective configuration and velocities after the pseudo time interval $\mathrm{d}t$, $U(\boldsymbol{r}_0)$ is the potential energy to the initial configuration. In rush dynamics, the time interval, $\mathrm{d}t$, has no physical significance.

The effect is easiest to envisage in a two-dimensional potential surface: the above transformation would cause the vector to point along the current contour of constant potential energy without significantly distorting the bonding geometry.

3.2 Simple approach to generalised conservation law dynamics

There are several indirect ways to conserve the specified property of a system. Two will be mentioned briefly (a) to provide a workable approach should an elegant direct approach not be found, and (b) because this less direct class of techniques is ideally suited to seeking out a required value of a function to which one will wish to adhere as the conserved property.

We have explored many techniques for maintaining constant the value of another (extrinsic) function Φ, while the energy is being minimised. These include (a) minimising

$$F = E + K(\Phi - \Phi^*)^2 \tag{29}$$

where Φ^* is the value of Φ to be conserved or (b) using a simulated annealing approach [20] where E is minimised through a combination of general downhill search and occasional uphill moves to avoid local minima, where the uphill move has to satisfy the following distribution:

$$p = \mathrm{e}^{-\Delta E/kT} \tag{30}$$

where ΔE is the difference between the current configuration and the previous one, T is the control variable usually referred to as temperature and k is the Boltzmann constant. These approaches can be many times faster when Φ^* is judiciously chosen.

The advantage of the more elegant approaches which directly manipulate the underlying laws of motion, is speed. We recall that the methods of this section are best suited to finding initial con-

formations satisfying Φ^* to which the simulation is then held by the "true" conservation techniques.

3.3 Underling relation between rush molecular dynamics and generalised conservation law dynamics

The above rush MD approach seems applicable only to conservation of potential energy. However, it is easy to reason that for a typical (fairly small) protein of some 2000 atoms (e.g. a lysozyme), this conservation of potential energy would still leave $3N - 1 = 5999$ dimensions which would conserve potential energy, but also allow up to 5999 new conservation laws to be introduced. In effect, these represent up to 5999 items of specific data (such as Nuclear Overhauser distances), or information about proteins in general, e.g. an expected amount of hydrophobic packing or the degree of intramolecular hydrogen bonding. We have described these aspects in terms of dynamical systems theory [21], specifically applied to the conservation problem [22].

3.4 Computational drawbacks of the rush dynamics algorithm

Most generally, the rush technique can be described as involving the rotation of any function gradient vector to be orthogonal to its original downhill direction, without considering first the conservation of potential energy. In many cases, such approaches are unstable, and rapidly lead to some kind of "explosion" of numbers to large absolute values. In comparison, the natural conservation laws may have a number of properties which render them intrinsically stable even as computer simulations expressing an algorithmic form.

A standard way [23] of assessing the stability of an MD program is to calculate the standard deviation of the total energy, E, for a certain number of steps while gradually increasing the integration size dt:

$$\sigma_E = \sqrt{\langle E^2 \rangle - \langle E \rangle^2} \tag{31}$$

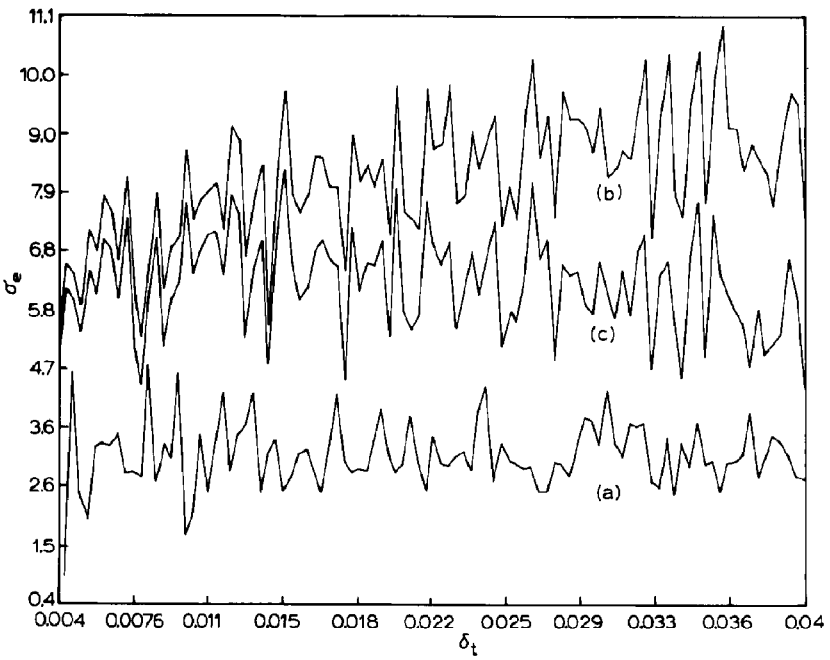

Fig. 1. Standard deviation of total energy for a system of 108 Lennard–Jonesian particles over 1000 steps in MD simulation: (a) for normal MD, (b) for rush MD with no feedback control, and (c) with feedback control.

We used a system of 108 particles interacting by a Lennard–Jonesian potential, $V(r_{ij})$

$$V(r_{ij}) = 4\varepsilon\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) \tag{32}$$

to compare the stability of normal MD and rush MD. Integration step size increases from 0.4–4.0 fs. For each step size, E is calculated over 1000 steps for both normal MD and rush MD. The results are shown in Fig. 1. It is clear that rush MD shows much larger fluctuations than normal MD, and it shows a much stronger tendency to instability than MD, i.e. E for rush MD steadily increases with dt.

From the description of rush MD algorithm, it is known that the dot product between the force vector and velocity vector $\boldsymbol{F} \cdot \boldsymbol{V}$ plays a key part in the adjustment of the new velocity, and $\boldsymbol{F} \cdot \boldsymbol{V} \neq 0$. If we calculate $\boldsymbol{F} \cdot \boldsymbol{V}$ in normal MD (see Fig. 2), it can be seen that $\boldsymbol{F} \cdot \boldsymbol{V}$ fluctuates symmetrically along the zero value line. The breaking of this symmetry in the behaviour of $\boldsymbol{F} \cdot \boldsymbol{V}$ may be the cause of instability in rush MD. Some adjustment to the pure rush MD form is needed for practical application. When the Cotterill–Madsen feedback control mechanism [18] is introduced into the rush MD simulation of 108 Lennard–Jonesian particles, it is much more stable, i.e. ε fluctuates about a steady mean (Fig. 1c) as does the normal MD calculation (Fig. 1a), although these fluctuations are much greater in the rush MD simulation than the normal MD simulation.

3.5 Alternative conservation approaches not related to rush dynamics

Some artificial conservation laws seem easier to implement and stabilise than others. One developed in the laboratories at Proteus is *momentum conservation dynamics* which is useful for emphasising large scale motions, as follows.

The underlying philosophy of the approach is to tackle directly one of the central problems in MD applications: how to achieve long time scale simulation without increasing hardware cost (this is also a merit of rush dynamics, though the speeding effects were discovered somewhat indirectly). Tackling the problem from the viewpoint of the MD algorithm itself, the obvious choice is to increase integration step size. However, this can quickly lead the simulation into an unstable

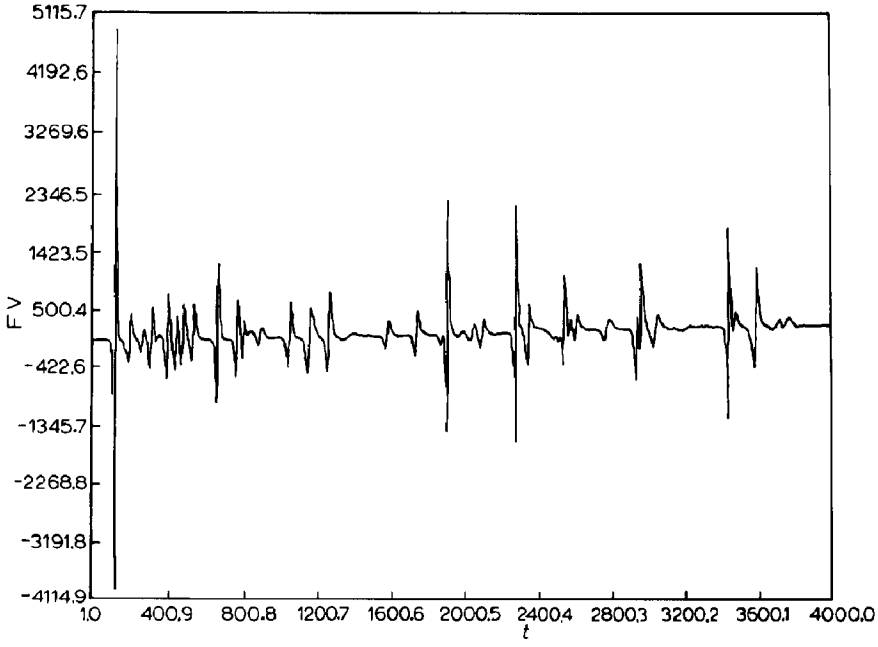

Fig. 2. The changing behaviour of $\boldsymbol{F} \cdot \boldsymbol{V}$ during normal MD simulation of 108 Lennard–Jonesian particles.

state and subsequent blow up due to the limitations inherent in the MD algorithm. Many methods have been tried with only limited success.

Nonetheless recently an idea from quantum field theory provides some inspiration for improving the MD algorithm [24]. This idea states that, when two particles interact, it can be imagined that a virtual particle is emitted from particle A to particle B, particle B then collides with the virtual image of A, and after the collision, particle B absorbs its virtual image. During this process, the momentum is required to be conserved. For a 2D system, it has been shown that this algorithm allows a ten-fold increase in integration step size. The central step of the algorithm is to modify the velocity increment in normal MD by

$$\Delta V = -\frac{a}{2b} + S\sqrt{\frac{a^2}{4b^2} - \frac{\Delta E}{bm}} \tag{33}$$

a, b are related to the previous velocities and S is a constant. Clearly, for this algorithm to work, the discriminant needs to be positive. If this condition is not satisfied, the system is in a so-called classically forbidden region. Again, we have been first in applying this idea to molecules. However, some further work is needed to make the algorithm robust. In some cases at present, simulations must be briefly paused and "backtracked" to recalculate the occasions when a "blow up" occurs.

3.6 Existing methods which introduce artificial conservation laws

The introduction of novel conservation laws is not wholly novel in itself. Some techniques already available and used actually imply some modification of Conservation Laws or the introduction of new laws. However, they are not always perceived in that light and are tailored to very specific problems. Techniques such as Nose's temperature bath dynamics [25] and constant pressure dynamics [26] can be considered as "artificial laws" in the sense that they model natural situations (e.g. conservation of temperature, conservation of pressure in a small volume in periodic boundary conditions) in a somewhat unnatural way, as a highly "localised" conservation law for a microscopic system. However, they differ from the approaches considered above in that they appear to have an affinity with the natural conservations.

3.7 General approach to recognition of new physical laws in the computer

Searching for conservation laws in Hamiltonian systems is best described by the Hamilton–Jacobi equation [27].

$$H\left(\frac{\partial S}{\partial q_1}, \frac{\partial S}{\partial q_2}, \ldots, \frac{\partial S}{\partial q_n}, q_1, q_2, \ldots, q_n\right) + \frac{\partial S}{\partial t} = 0 \tag{34}$$

Where $q_1, q_2 \ldots q_n$ are generalized coordinates, S is the generating function.

The solution of this equation for S will lead to the solution of Hamilton's equations of motion. For the Hamilton–Jacobi equation, a conservation law needs to satisfy the following equation [28]:

$$\frac{\partial T(S)}{\partial t} + \frac{\partial X(S)}{\partial q_i} = 0 \tag{35}$$

where $T(S)$ is called conserved density and $X(S)$ flux. The conserved density $T(S)$ has such a property that

$$\int T(S(\boldsymbol{q}, t))\, \mathrm{d}\boldsymbol{q} \tag{36}$$

is independent of time t, for all solutions of the Hamilton–Jacobi equation so that the integral converges.

When a system has an infinite number of nontrivial conservation laws, the system is formally integratable. For practical problems, complete integration is usually an extremely difficult task, and numerical techniques are used instead. The time invariant property of conserved density can be used to evaluate the numerical accuracy. This means in practice that a single universal approach to designing a "conservation law" is probably not possible. It requires a number of approaches and skills, and a variety of explorative

trial and error approaches, which we collectively refer to as "Physical Law Engineering".

3.8 Assessment of speeding

It seems self-evident that rush dynamics and/or a generalised conservation law approach must be faster for several intrinsic reasons:

(a) the method implicitly avoids going to an energy minimum;
(b) there is one less dimension for every constraint, i.e. there are less parameters;
(c) even though finding an improved solution to a model is difficult, it rapidly locates equally good solutions;
(d) the entropy and depth of minima can be evaluated from the local curvature of the iso-potential energy manifold. (Once two contour levels has been established in the region of a minimum, a quadratic interpolation will provide an estimator of the actual minimum and of the entropy relative to another minimum).

Unfortunately, it is intrinsically difficult to quantify the speeding in view of the fact that the "time step" in techniques of this class has no real meaning. Further, between normal and this type of dynamics, one is comparing phase space volumes of different dimensionality. However, one can assess that something of interest has been located in n iterations, or that in one million iterations, n "Planck" elements of phase space have been searched, or n minima located, for example.

For sometime our view has been that the most persuasive of these is the number of progressively deeper minima located, as used in the GLOBEX stack method [29]. However, even this should be treated with reservation, (a) since important local conformations might be missed by virtue of less ability to look "near at hand" (in effect, by virtue of concentrating on a "depth first" rather than "breadth first" search), or (b) many minima might be located rapidly, without further improvement over a long time, or (c) many minima of high energy might be explored, misleadingly giving an assessment similar to a run with many low energy minima.

A more quantitative measure can be obtained from the difference in the autocorrelation function for the RMS deviation changes between a normal MD run and a rush MD run.

3.9 Results

Until the present study, the Cotterill–Madsen algorithm had been applied only to materials science simulations, particularly regarding crystal dislocations. The original study [15] involved studies of the diffusion of a hole in a two-dimensional Lennard–Jones solid. We have repeated this study and reconfirm that it is indeed some million times faster than normal dynamics, in the sense of the relative rate at which the hole moves. We observed, however, that as in a diffusion process this system is one in which the potential energy stays approximately constant in any event. Interesting systems such as protein molecules spend most of their time in a localised, well defined energy well and indeed it is this drastic departure from iso-potential energy behaviour which gives rise to the characteristic structure. We would expect that the speed benefits in terms of searching conformational space would decrease as the potential energies for the configurational possibilities of the system go from a plateau-like to a valley-like description. In essence, this means as entropy makes a diminishing contribution to the dominant behaviour of the system, and will tend to proceed in conjunction with the appearance of strong interatomic forces as more complex systems are progressively studied.

It was early noted that the Cotterill–Madsen algorithm is fundamentally unstable for any practical computational step size. To overcome this, a check is made every 0.25 ps to discover the departure from a constant potential energy and a small correction is applied to the velocity vector. This represents additional computation per iteration, but still allows substantial speed gains which arise essentially from the reduction of the dimensionality of the problem. How then does the algorithm actually fare for conformational changes in organic molecules? Preliminary studies using the Cotterill–Madsen algorithm have been briefly reported by us [19] using *N*-acetyl

alanyl *N'*-methylamide, a simple analogue of a residue within the context of a protein chain. An extended analysis of that result is pertinent to a better understanding of the following studies. Because this relatively simple molecule has only two major degrees of freedom, the rotation of the N–C and C–C′ bonds, the trajectory can be easily displayed on a two-dimensional diagram. To casual inspection the trajectory is not grossly different from that in normal dynamics. It does not adhere to contours of iso-potential energy as plotted for the two dominant degrees of freedom with the other angles held constant, minimised or statistically mechanically averaged. This is because the other potential variables are indeed true degrees of freedom (bond stretching, valence angle bending, and rotation of bonds with at most partial double bond character) which cannot be visualised in such a representation. Particularly, energy can be exchanged with these hidden variables to overcome energy barriers, and the hidden conformational variables play a role not entirely different to the thermal environment of a molecule simulated under "rigid geometry" conditions in which only N–C and C–C′ rotations are allowed to vary. Note that a direct comparison is possible because we have developed techniques in which the calculation of the angular momenta associated with bond rotations can be performed in "dihedral angle space". Similarly, to casual inspection, the dynamic motions perceived at a graphics terminal are not grossly different from those seen in the application of normal molecular dynamics. One important difference is that in a given time as estimated from the trajectory diagram in particular, more conformational space is covered with greater density in the case of the Cotterill–Madsen algorithm. However, it was clear that though the method of estimating speed gain was crude (being simply based on comparing the time taken for normal dynamics to achieve comparable spread and density), the speed gain was at best of the order of hundreds-fold, not millions-fold.

This test system may nonetheless fail to reveal the Cotterill–Madsen algorithm at its best performance because the molecule has so few major degrees of freedom. The studies were extended to the folding a simple model polypeptide *N*-acetyl-(Ala)5–(Gly)2–(Ala)5-*N'*-methylamide (A5G2A5). Simulations using normal dynamics were first "equilibrated" or "relaxed" by 10000 × 0.005 fs steps, and then simulations were performed over 50 ps, using a 0.5 fs step size. As is customary, temperature regulation at 310 K was imposed by rescaling of the velocities, and this was applied in both the initial "equilibration" phase and the subsequent phase. Both simulations (normal and Cotterill–Madsen) were started with identical velocities and identical structures (an extended conformation with all backbone angles at 180°). This molecule is significantly longer than the *N*-acetyl alanyl-*N'*-methylamide previously studied, and hence of polymer character: it thereby lends itself more naturally to more objective measures of conformational variation which has been developed largely for polymers as opposed to compact organic compounds. These natural methods include descriptions which essentially show the extent to which a polymeric chain is folded up into a relatively compact state. Figure 3 depicts qualitatively the folding of the peptide. The root mean square (rms) fit between each "snapshot" and the initial extended conformation is plotted. The rms fit measure is in this case the differences between interatomic distances for the two conformations, measured between pairs of alpha-carbon atoms within each conformation. The classical crystallographers' rms measure, based on super positioning and orientating of the two conformers and examining the mean square distance of the corresponding atoms in the two separate structures, is inappropriate (and indeed meaningless) for widely differing conformers.

Figure 3 shows that during normal dynamics A5G2A5 folds at a steady rate and spends the remainder of the simulation vibrating and locally liberating in that state. It is, in effect, trapped. In contrast, the Cotterill–Madsen algorithm causes the molecule to fold more quickly initially, but more importantly the structure never remains truly compact but constantly unfolds and refolds during the course of the simulation. The subsequent history reveals a much greater and richer variety of conformations than is accessible to the

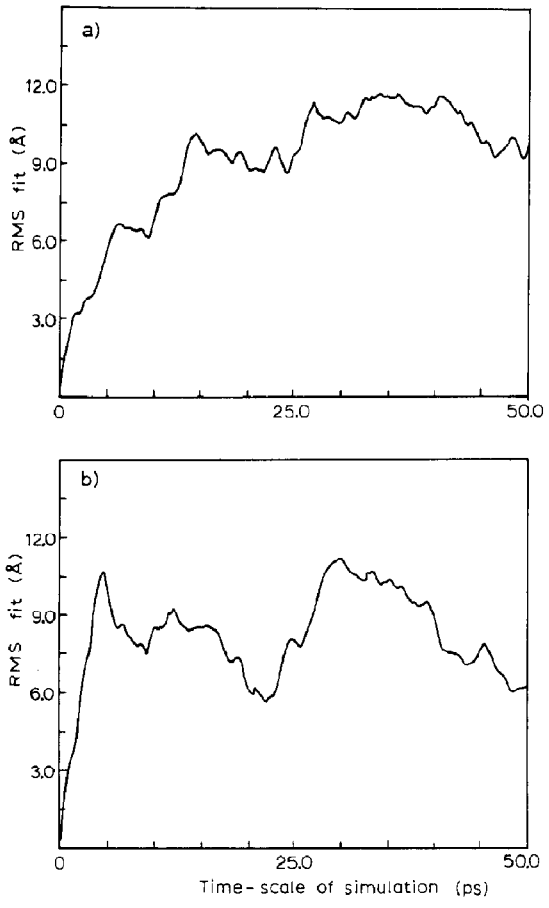

Fig. 3. The rms fit deviations between the initial extended conformation and the configurations obtained during (a) and normal MD and (b) a rush MD run.

normal dynamics algorithm. It is important to note that the method of analysis used distinguishes only states by virtue of their compactness and in fact a large number of different conformers will be accessed which have the same degree of compactness. That is to say, the estimate obtained will be a minimal estimate. By comparing the autocorrelation functions for the A5G2A5 studied by normal and Cotterill–Madsen dynamics, this minimal estimate of the ratio of the number of distinct conformations generated is 150.

The notion of "minimal" has no formal absolute significance here and even lower minimal estimates of speed enhancement can be obtained by using methods which are poorer at distinguishing a greater variety of states. Computer experiments reveal that the above approach provides at least a more workable estimate than other approaches derived from polymer theory. A more formal approach is based on evaluating the partition function for the mean squared end-to-end distance of the polymer, or between any two specified points on the polymer. Such a partition function represents the abundance of conformers as a function of the mean square end-to-end distance. Effectively, conformations classified by having different end-to-end distances are counted and some conformations are more abundant for some mean square end-to-end distances than others. From such a viewpoint normal and Cotterill–Madsen dynamics do not differ so significantly at the sampling levels studied so far: though the ratio is in favour of the Cotterill–Madsen algorithm by about ten-fold.

These difficulties of assessment arise in that the number of iterations performed is of the same order per unit computation time in both the normal and Cotterill–Madsen methods, and each iteration effectively represents a new conformation. The advantage of the Cotterill–Madsen approach is not that it generates a larger number of conformations, but a larger number of *significantly different* conformations. The above lower minimal estimate presumably arises became information based only on end-to-end distance distinguishes conformations even less than a method based on compaction. To attempt to achieve a more reasonable estimate than either of the above "minimal" estimates, this and similar systems have been studied using a simple program developed by one of us (BR) for classifying protein loops as similar or dissimilar. Basically, conformations are considered identical if the root mean squared deviations of their dihedral angles (in this case, the N–C and C–C′ rotation angles of each residue unit), divided by the number of dihedral angles, does not exceed a critical value. Using $60 \times 60 = 3600$ square degrees angular measure gives an adequate density of sampling for comparison of the numbers of conformations generated, at least as far as an order of magnitude. It is also natural stereochemical limit measure in that conformers in polypeptides tend to

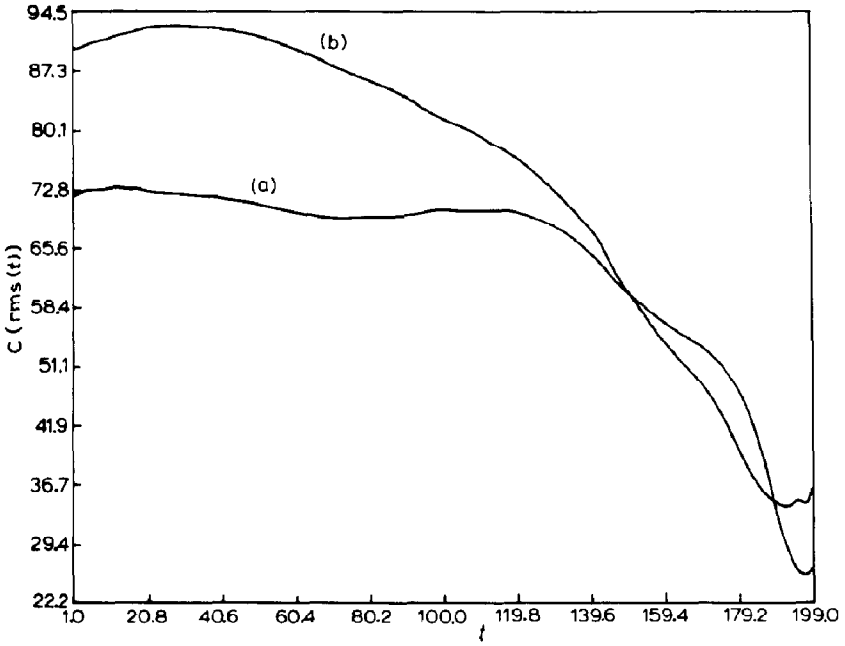

Fig. 4. C (rms(t)) is the autocorrelation function for the rms deviation changes during (a) normal MD, and (b) rush MD for A5G2A5 simulations.

lie, as they do for hydrocarbons, in energy wells of some 120° width. This method suggests closer to a thousand-fold enhancement in the case of 3600 square degrees, but is dependent on the critical value chosen. A more formal approach depends on calculating the partition function for the mean square deviation between the ends of a polymer.

4. Summary

The dynamical system properties under "rush dynamics" are novel, complex, and under study. These simulations have been analysed for both periodicity and quasi-periodicity. No conformation with specific velocities was re-encountered in this study, nor was any closely comparable conformation with closely comparable velocities encountered. Thus there is no evidence of any consistent return of trajectories close to the starting point within the lifetime of the simulation, though some gross properties, and specifically compactness may show overall periodic behaviour. The simulations show some evidence of chaotic behaviour which could well be advantageous in searching the manifold of iso-potential energy most exhaustively. "Fingerprints" for chaotic behaviour as analysed by us will be described elsewhere. With the exception of the approximate periodic repeat in "compactness" and related properties, there is as yet no evidence of an attractor on any specific continuous iso-potential energy surface, though from the point of view of phase space as a whole, the manifold represents a clear, well bounded attractor.

Studies have been carried out on a number of smaller proteins and in particular oxytocin has been well studied. Note, however, that the inclusion of a disulphide loop makes assessment by the method of Fig. 3 more difficult.

The practical application of this approach is seen as requiring interaction with classical molecular dynamics. At regular but not necessarily frequent intervals, the Cotterill–Madsen dynamics can be temporarily interrupted and normal dynamics applied. This is to find the nearest significant potential minima and to evaluate the

free energy in the nearest low energy region. Methods of doing this efficiently will be described elsewhere, but the general principle should be self-evident and the details are not believed to be critical. An analogous approach using classical energy minimisation instead of classical dynamics has be used to assess the number of minima close to a trajectory [30]. The results in terms of new minima found is of the order of 100–200-fold.

References

1 G. Ciccotti and W.G. Hoover (eds.), Molecular dynamics simulation of statistical mechanical systems, (North-Holland, Amsterdam, 1986).
2 W.F. van Gunsteren and P. Weiner (eds.), Computer simulation of biomolecular systems (ESCOM, Leiden, 1989).
3 B.J. Alder, Theory of Liquids, (Gordon and Breach, New York, NY, 1968).
4 D.A. McQuarrie, Statistical mechanics, Harper and Row, New York, NY, 1976).
5 J.J.C. Berendsen, J.P.M. Postma and W.F. van Gunsteren, Statistical mechanics and molecular dynamics: The calculation of free energy, in: Molecular dynamics and protein structure ed. J. Hermans, (Polycrystal Book Service, Western Springs, NC, 1985).
6 M. Mezei and D.L. Beveride, Free energy simulation, Ann. N.Y. Acad. Sci. 482 (1986) 1–23.
7 M. Mezei, S. Swaminathan and D.L. Beveridge, J. Am. Chem. Soc. 100 (1978) 3255.
8 W.L. Jorgensen and C. Ravimoham, J. Chem. Phys. 83 (1985) 3050.
9 U.C. Singh, F.K. Brown, P.A. Bash and P.A. Kollman, J. Am. Chem. Soc. 109 (1987) 1607.
10 I.I. Sheykhet and B. Ya. Simkin, Comp. Phys. Rep. 12 (1990) 67.
11 M. Mezei, J. Chem. Phys. 86 (1987) 7085.
12 J.P. Valleau and G.M. Torrie, in: Statistical mechanics Part A: Equilibrium techniques, ed. B.J. Berne (Plenum Press, New York, NY, 1977).
13 D. Frenkel, in: Molecular dynamics simulation of statistical mechanical systems, eds. G. Ciccoti and W.G. Hoover (North-Holland, Amsterdam, 1986).
14 P.A. Bash, U.C. Singh, R. Langridge and P.A. Kollman, Science 236 (1987) 564.
15 B. Robson and E. Platt, J. Mol. Biol. 188 (1986) 259.
16 B. Robson and E. Platt, J. Computer-Aided Mol. Des. 1 (1987) 17.
17 B. Robson and D.J. Osguthorpe, J. Mol. Biol. 132 (1979) 19.
18 J.U. Madsen and R.M.J. Cotterill, in: Potential energy contour tracing, ed. H. Bohr, Characterizing complexity (World Scientific, Singapore, 1990) p. 177.
19 R.M.J. Cotterill, E. Platt and B. Robson, in: Rapid conformational investigations of organic molecules, eds. M. Soumpasis and T. Jovin, Computation on biomolecular structures: Achievements, problems, perspectives; Schloss Ringberg Symp. 1991 Proc. (Springer, Berlin, 1991).
20 S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, Science 220 (1983) 671.
21 D.J. Ward, A.M. Brass, J. Li, E. Platt, Y. Chen and B. Robson, in: Theoretical approaches to peptide drug design, ed. D.J. Ward, Peptide pharmaceuticals (Wiley, Chichester, 1991) p. 83.
22 B. Robson, E. Platt and J. Li, Computer-aided bimolecular design, in: eds. D.L. Beveride and R. Lavery, Theoretical biochemistry and molecular biology (Adenine Press, Guilderland, NY, 1991) p. 207.
23 H.J.C. Berendsen and W.F. van Gunsteren, Practical algorithms for dynamic simulations, in: eds. G. Ciccotte and W.G. Hoover, Molecular-dynamics simulations of statistical-mechanical systems (North-Holland, Amsterdam, 1986) p. 43.
24 A.T. Emren, Phys. Scr. T33 (1990) 77.
25 S. Nose, Mol. Phys. 52 (1984) 255.
26 S. Nose and M.L. Klein, Mol. Phys. 50 (1983) 1055.
27 V.I. Arnold, Mathematical methods in classical mechanics (Springer-Verlag, New York, NY, 1989).
28 J.A. Cavalcante and K. Tenenblat, J. Math. Phys. 29 (1988) 806.
29 B. Robson, E. Platt, R.V. Fisheigh, A. Marsden and P. Millard, J. Mol. Graph. 5 (1987) 17.
30 D.J. Ward, Y. Chen, E. Platt and B. Robson, J. Theor. Biol. 148 (1991) 193.