Journal of Computer-Aided Molecular Design, 8 (1994) 67–82
ESCOM

J-CAMD 231

Novel algorithms for searching conformational space

D. Byrne[a], J. Li[a], E. Platt[a], B. Robson[a,*] and P. Weiner[a,b]

[a]*Proteus Molecular Design Ltd., Proteus House, Lyme Green Business Park, Macclesfield, Cheshire SK11 0JL, U.K.*
[b]*Amber Systems Inc., 28 Tower Street, Somerville, MA 02143, U.S.A.*

Received 27 September 1993
Accepted 29 October 1993

Key words: Molecular simulation; Conformational search; Dynamic algorithms

SUMMARY

The modelling of biological molecules by molecular dynamics is beset by a range of problems. The most important of these is the multiple-minima problem. The deep metastable minima can cause difficulties in proper equilibration of a molecular system and result in the simulated system being trapped in a long-lived metastable state. One way to overcome these problems is to re-engineer the 'Newtonian Rules' in order to more efficiently search conformational space. Re-engineering of the 'Newtonian Rules' implies a redesign of the physical laws arising from them. This is done in various ways by the RUSH, Hybrid Monte Carlo and PEACS algorithms. This paper explores applications of these algorithms, and compares them to a traditional molecular dynamics method.

INTRODUCTION

Molecular dynamics (MD) is an established simulation technique in the prediction, analysis and design of complex molecules [1,2]. These techniques typically apply the Newtonian Laws to the motion of all the atoms in the system, once the forces have been defined as a function of atom type, bond type, dihedral type and interatomic distance. There are many software packages that implement these techniques very efficiently on computers ranging from workstations to large parallel supercomputers.

Unfortunately, the primary difficulty in modelling complex molecules is not overcome by attention to details of hardware and software implementation. This difficulty relates to the presence of multiple minima in the $(3N-6)$-dimensional conformational space that defines the potential energy surface. These minima can cause the simulations to become locally trapped or delayed. Dynamical simulations are run in the $(6N-12)$-dimensional phase space which may have multiple attractors that can also locally trap or delay trajectories. Thus, it is difficult to insure

*To whom correspondence should be addressed.

adequate sampling to accurately estimate thermodynamic properties, such as free energy, or even to insure that the structures reached include all of the biologically relevant ones.

This paper explores the properties of three distinct methods which may be helpful alone or as part of a more complex protocol in addressing some of these issues. These methods include RUSH dynamics [3–5], Hybrid Monte Carlo [6,7] and PEACS [8]. The methods are non-Newtonian dynamics in the sense that the Newtonian laws are highly modified or Newtonian motion is in some sense interrupted.

The methods are tested on *N*-acetylalanyl-*N'*-methylamide (alanine dipeptide) which has a potential energy surface that can be expressed graphically in two dimensions (see Fig. 1). The major degrees of flexibility arise from rotation about the dihedral angles ϕ and ψ. By following changes in the sign of these dihedral angles, most interesting processes (transitions between minima) on the surface can be categorized.

METHODS

Force field

The force field of Robson and Platt [9] was used. This force field has been carefully calibrated to reproduce quantum-mechanical and experimental data for a variety of model peptide systems. This approach first calculates the distances in a distorted space known as r'. This avoids loss of precision in the derivatives $\partial E/\partial r$ at very close interatomic approaches, models physical features of the interactions considered desirable at short and long range, and avoids use of a time-consuming square-root calculation, even when odd powers are used (the nonbonding potential used is the 9-6-1 form). In dynamical simulations, this also allows efficient evaluation of force because, since the Jacobian dr'/dr is known, the true force $\partial E/\partial r$ can be evaluated.

RUSH dynamics

RUSH dynamics [3–5], like the hybrid Monte Carlo method, has been adapted from the realm of physics for service in chemistry. The original application was to study dislocations in two-dimensional matter [3], where the original interest was fueled by demonstration that, under some conditions, dislocations could move one to ten million times faster than in classical particle simulations. The first application to molecular systems was demonstrated by a collaboration of the laboratories of Robson and Cotterill at the Schloss Ringberg Symposium in 1988 [3]. It was not immediately obvious that a transfer to molecular applications was possible, since the Newtonian laws are drastically modified, and there is the obvious requirement to retain the connectivity and bond geometry of any molecule, without 'particle anarchy'.

RUSH dynamics is seen as a special case in the engineering of novel conservation laws [4,5]. In the particular case of RUSH dynamics, it is the potential energy which is conserved. The isopotential energy surface of a 6N-dimensional phase space of an N-atom system is a $(3N-1)$-dimensional manifold. Adherence of the trajectory to the isopotential energy manifold is achieved by (in effect) rotating the normal force vector $\partial E/\partial Q$ 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:

$$r_1 = r_0 + v_0\,dt$$

$$v_1 = v_0 - \nabla U \cdot \frac{\nabla U \cdot v_0}{|\nabla U|^2}$$

$$\nabla U \cdot v_1 = 0$$

where r_0 is the instantaneous configuration of the N particles with the velocities v_0, r_1 and v_1 are the respective configuration and velocities after the pseudo time interval dt and $U(r_0)$ is the potential energy of the initial configuration.

In RUSH dynamics, the time interval dt has no physical significance. However, it does dictate, in any finite numerical integration, how well the simulation adheres to the ideal, continuous trajectory, using the finite numeric precision of the computer. In the present case, it principally relates the extent to which $U(r_1)$ and all subsequent values of the potential energy hold to $U(r_0)$, the initial potential energy.

Though formally the method avoids 'particle anarchy', 'blow-ups' do occur as a consequence of finite step size and finite precision, and supporting software is required to maintain or restore

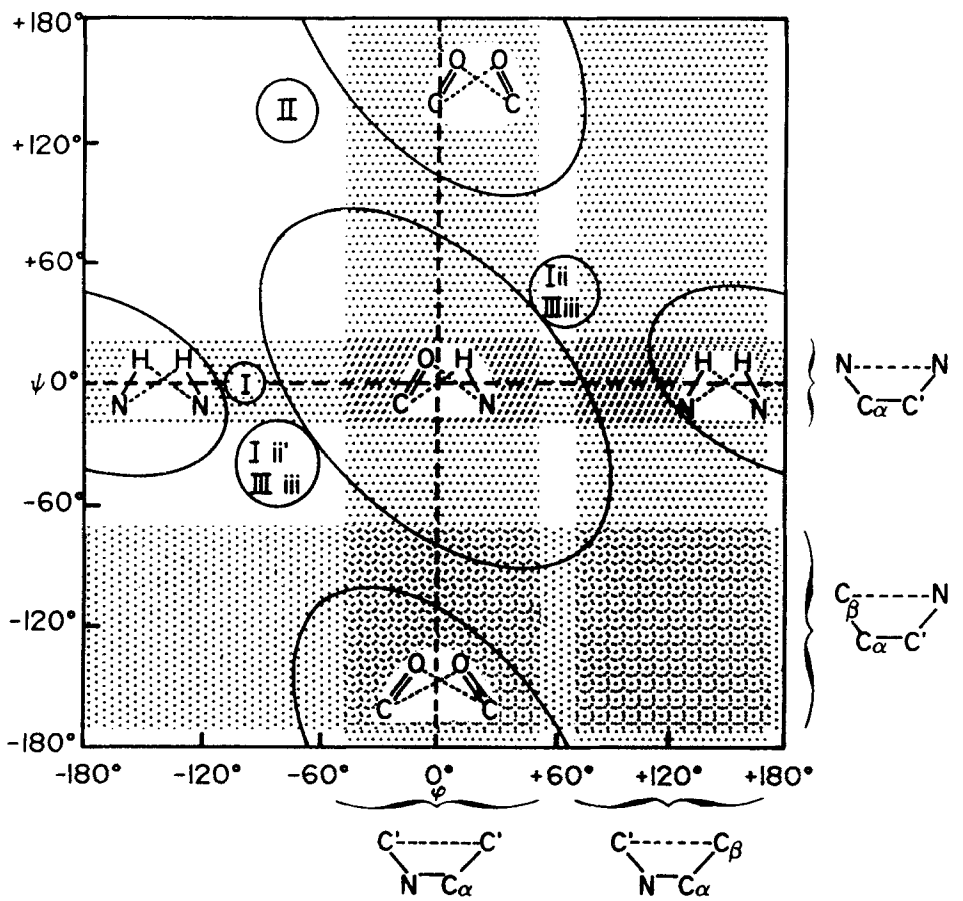


Fig. 1. Steric clashes between atom pairs in *N*-acetylalanyl-*N'*-methylamide. The grey areas and interiors of the ellipses are 'sterically forbidden' by at least one clash. Such a diagram is at the origin of the Ramachandran plot. Interactions are represented by dotted lines. The circles are allowed regions favored by β-turns. Adapted from Robson and Garnier [10].

the adherence to the isopotential surface. When such an explosion occurs, use of small dt and backtracking to restart from a previous stable condition are recommended. Potential instabilities seem a common feature of artificial, as opposed to actual, conservation laws, and researchers exploring 'physical law engineering' for molecular systems are advised to install check and recovery mechanisms.

Furthermore, combination in software with classical dynamics or hybrid Monte Carlo (see below) is also desirable: though the method satisfies one feature of the ergodic theorem by sampling the manifold $U(r_0)$ with equal probability, it cannot lead to a Boltzmann relationship, since access to other potential energy levels is expressly forbidden. In sampling phase space, there will always exist one manifold which will weave its way to the largest extent through phase space and an arbitrary choice of $U(r_0)$ may not best represent the optimal search power of the RUSH method. The other methods can be used to 'hunt out' a point in phase space with $U(r_0)$ satisfying the required energy.

The effect of RUSH dynamics is easiest to envisage in a two-dimensional potential surface. The above transformation of velocity would cause the vector to point along the current contour of constant potential energy without significantly distorting the bonding geometry. This cannot be seen in the maps in the figures presented here since, although the potential surface is represented as a two-dimensional slice through conformational space, there are many degrees of freedom at each point in that two-dimensional slice.

Hybrid Monte Carlo

Like RUSH dynamics, Hybrid Monte Carlo (HMC) [7] is borrowed from the physics domain, where it was recently developed by Pendleton and co-workers [6] for the quantum chromodynamics community. Again, it was not obvious that a method developed for such a different problem (i.e., the study of quark systems) is applicable to the molecular world. However, the resulting combination of MD and Monte Carlo (MC) seems a 'sensible' modus operandi. The combination is a specific and formal one. In particular, the time integration scheme used is time reversible: it retraces its steps on reversal of the sign of the integration step. This is in fact a stringent criterion, not met by the widely used Runge–Kutta and Beeman integration schemes.

HMC [6,7] carries out L dynamics iterations with an MC step between each series of L iterations. Hence, the L iterations correspond to the random perturbation of the configurations of particles in the classical MC approaches. For large L and vanishing MC contributions, the technique has the useful property of converging to classical dynamics.

The central idea of HMC [6,7] is to introduce MC sampling steps during an MD run. A stochastic element in the Langevin manner is inherent in the generation of new momenta with a distribution satisfying the prevailing temperature [7]. MC sampling is generally done through the Metropolis acceptance criterion:

$$P_{acc} = \min(1, \exp(\delta E / k_B T))$$

where δE is the energy change due to a configurational change, k_B is the Boltzmann constant and T is the simulation temperature. This method is nonstandard in that the total (as opposed to the potential) energy is used.

Though the Monte Carlo step constitutes a departure from Newtonian behavior, it has the

important effect of guaranteeing the Ergodic condition (in contrast to RUSH dynamics, this includes the Boltzmann distribution). Moreover, this guarantee is over relatively short time scales in integration. It might thus be expected that this is at the cost of losing the opportunity for extensive, breadth-first searching, but it should guarantee, a priori, a more rapid convergence of entropy and free energy than is the case in classical dynamics, and it is likely to be highly suited for perturbation methods.

PEACS

PEACS (Potential Energy Annealed Conformational Search) [8] performs molecular conformational searches through biasing dynamical simulation towards low-energy states. This is achieved by coupling the potential energy of the system with an external potential bath. Therefore, the potential energy changes in PEACS calculations according to

$$\frac{dU}{dt} = \frac{1}{\tau_u}(U_0 - U)$$

where U_0 is the potential energy value of the potential bath and τ_u is the potential energy relaxation time which controls the rate of exchange between the potential bath and the potential of the system. This coupling equation can be turned into constraints on the velocities of the dynamical simulation. The practical implementation of the PEACS algorithm involves calculating velocities by

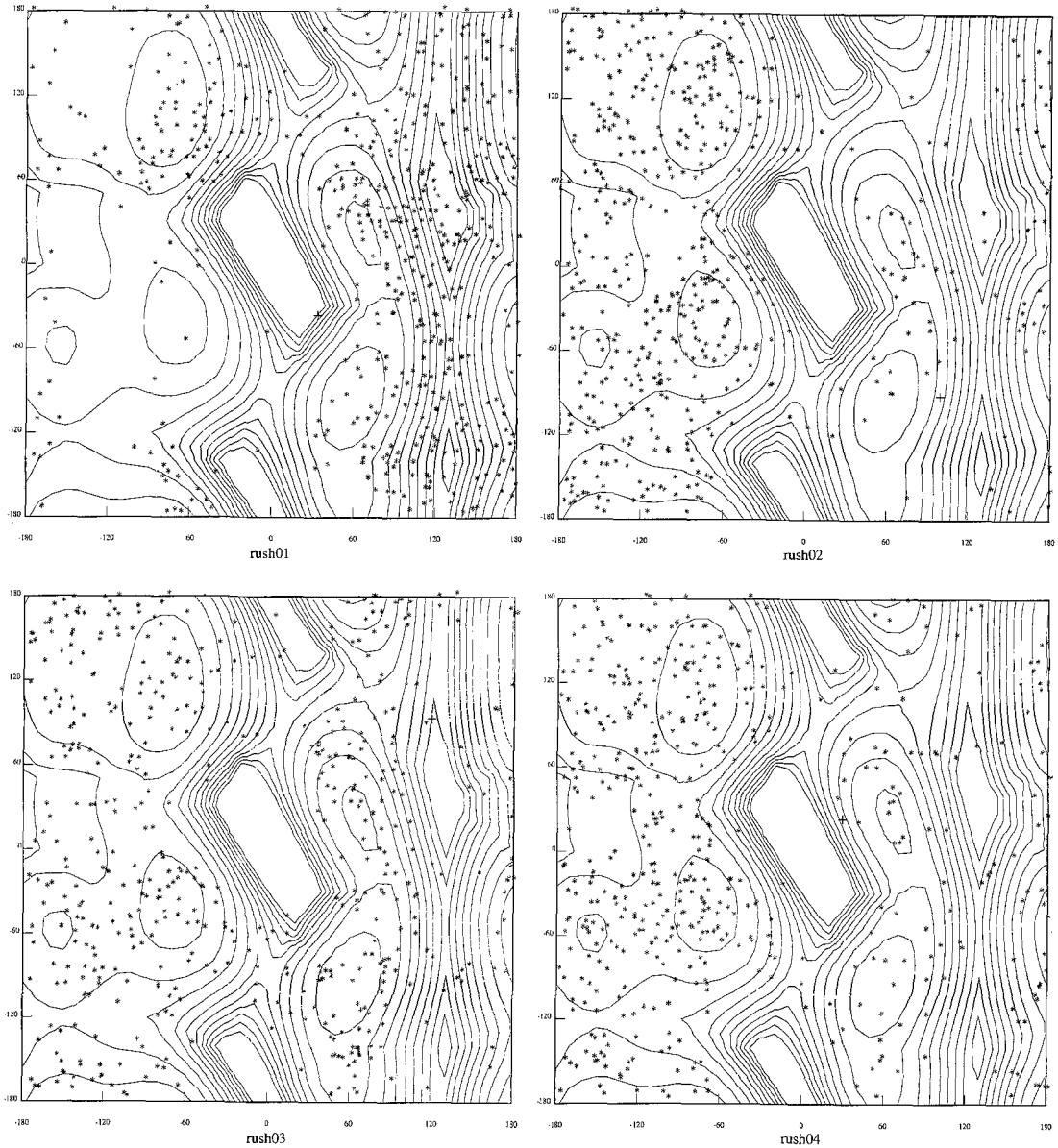
$$v_i(t+\delta t) = v_i\left(t - \frac{1}{2}\delta t\right) + \left(\varepsilon_i - \frac{\delta t}{m_i}\right)\nabla_i U$$

The correction factor ε_i can be calculated from

$$\varepsilon_i = \frac{U_0 - U}{\tau_u N \|\nabla_i V\|^2}$$

It is clear that the energy level at which the algorithm samples is determined by the reference potential energy value U_0 of the potential bath. This method thus has a certain affinity with RUSH dynamics. An energy level change scheme must be devised to achieve conformational searching. One practical approach is to make the energy level of the potential bath a function of simulation time, and the energy value is reduced according to the lowest energy values found in the simulation process.

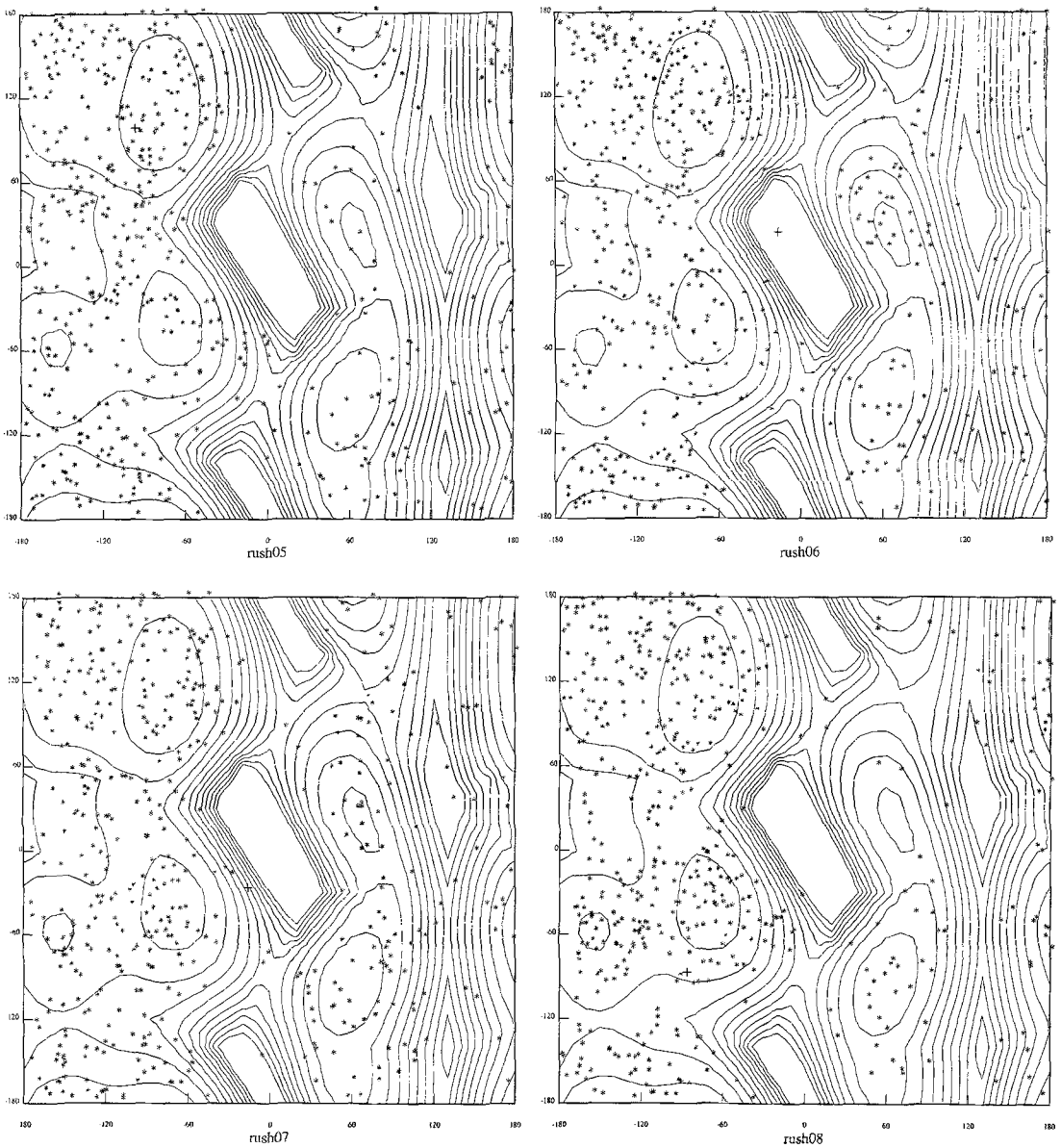
It is worth noting that this algorithm is non-Newtonian, i.e., the usual conserved quantities (e.g., total energy) of a Newtonian system no longer exist. Hence, as with the other procedures described above, it is more appropriate to regard the algorithm as a conformational searching algorithm, rather than a dynamical simulation algorithm. The behaviour of this algorithm is closely related to the parameters used in coupling the potential energy of the system and that of the external bath, and the descending energy level of the potential bath during conformation searching. Certain experimentations in the choice of parameters are needed to achieve optimal results.



Figs. 2.1–2.4. RUSH dynamics (points 1–4).

Simulation

Eight simulations at a constant temperature of 310 K were run for each of the three methods described above, as well as for a traditional Newtonian MD procedure. Points of both high and low potential energy were selected as starting points from the ϕ/ψ map of alanine dipeptide. Each run was equilibrated for 0.5 ps and run for an additional 50 ps. A step size of 0.5 fs was used.

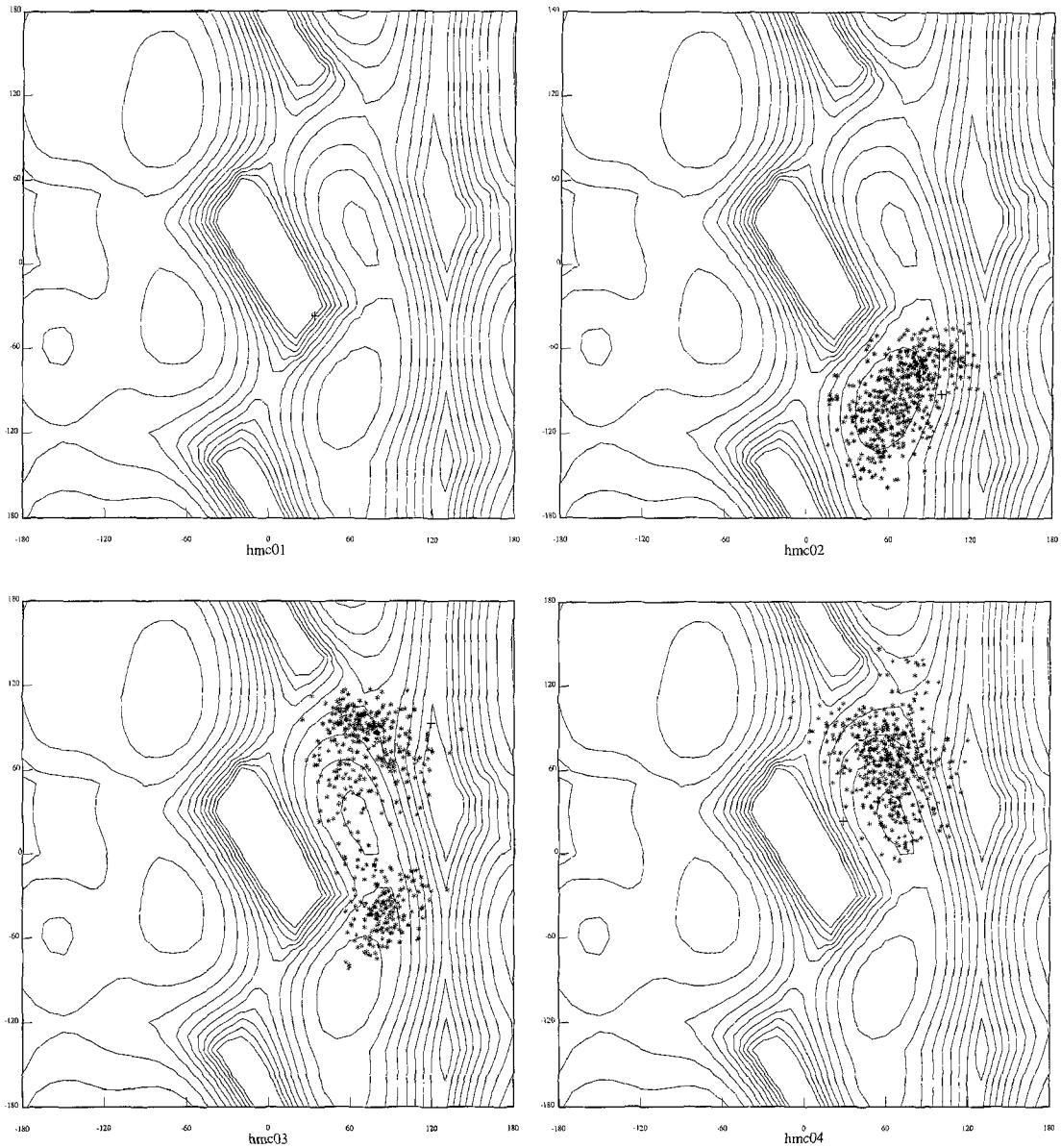


Figs. 2.5–2.8. RUSH dynamics (points 5–8).

The simulations were performed in vacuo, using the potentials of Robson and Platt [9] with a dielectric constant of 3.5. Every 200th point was saved.

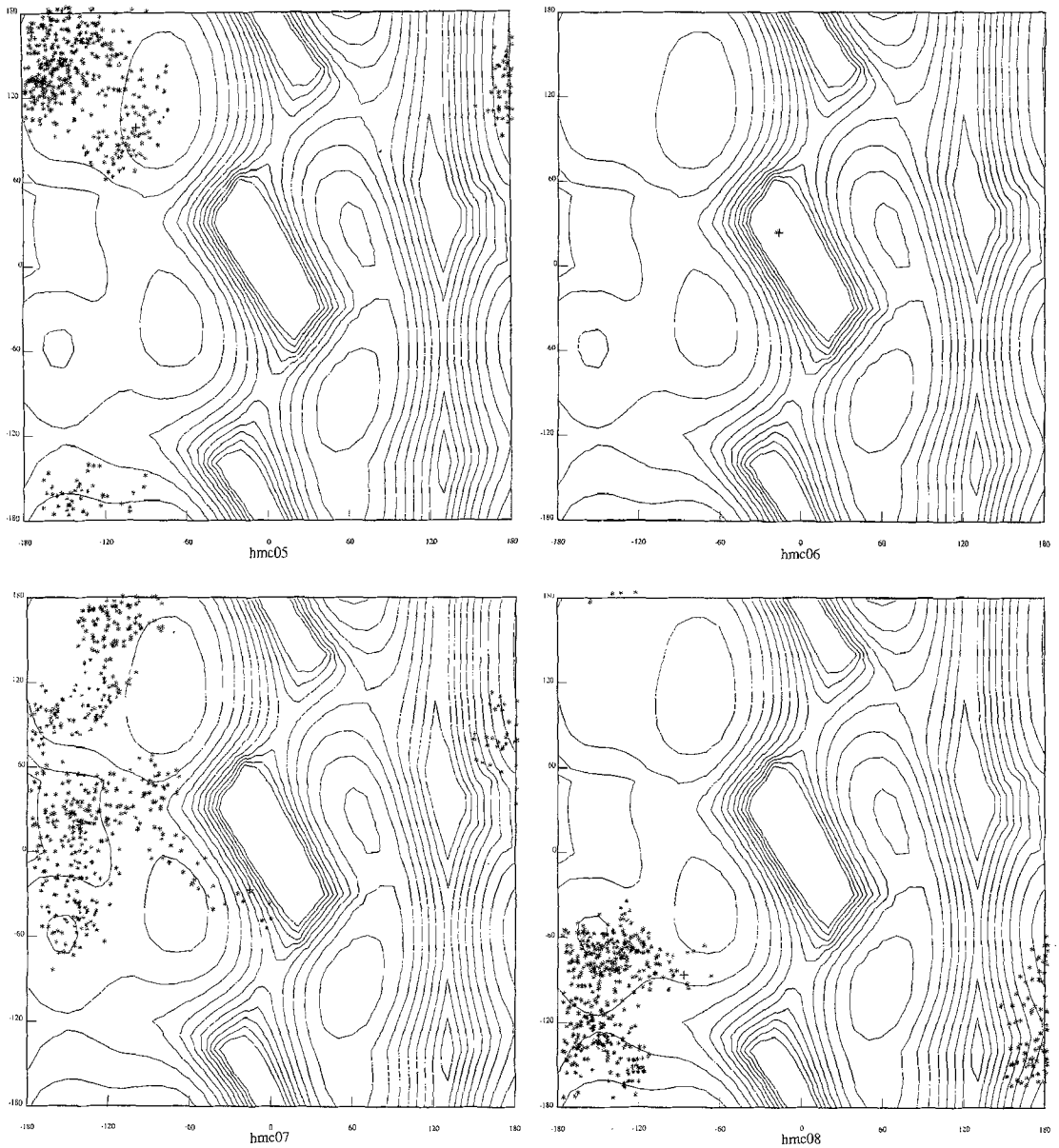
Analysis

The conformational space was divided into four volumes, corresponding to quadrants on the



Figs. 3.1–3.4. Hybrid Monte Carlo (points 1–4).

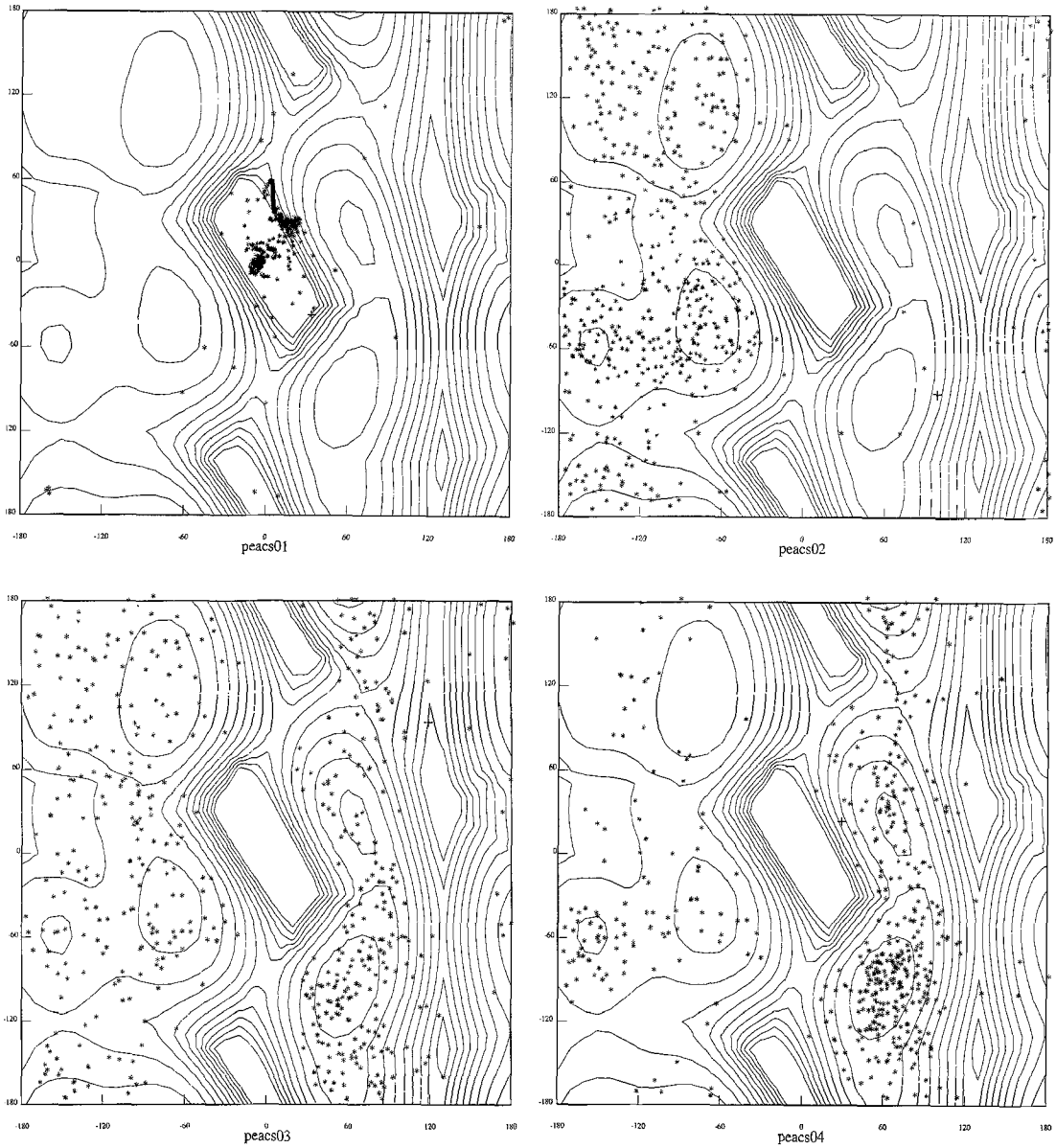
surface of the torus representing dihedral angle space, i.e., the conformational space partitioned as to different values of the dihedral angles ϕ (C'-N-C-C') and ψ (N-C-X'-N), these quadrants being E ($\phi < 0.0$, $\psi > 0.0$), H ($\phi < 0.0$, $\psi < 0.0$), L ($\phi > 0.0$, $\psi > 0.0$) and G ($\phi > 0.0$, $\psi < 0.0$). The EHLG terminology relates to conformations for amino acid residues commonly seen in globular proteins, viz: E = extended chain, H = right-hand α-helix, L = left-hand α-helix, and



Figs. 3.5–3.8. Hybrid Monte Carlo (points 5–8).

G = glycine-like, corresponding to a conformation particularly prevalent to glycine residues. In the tables, L is represented by Q1, E by Q2, H by Q3, and G by Q4.

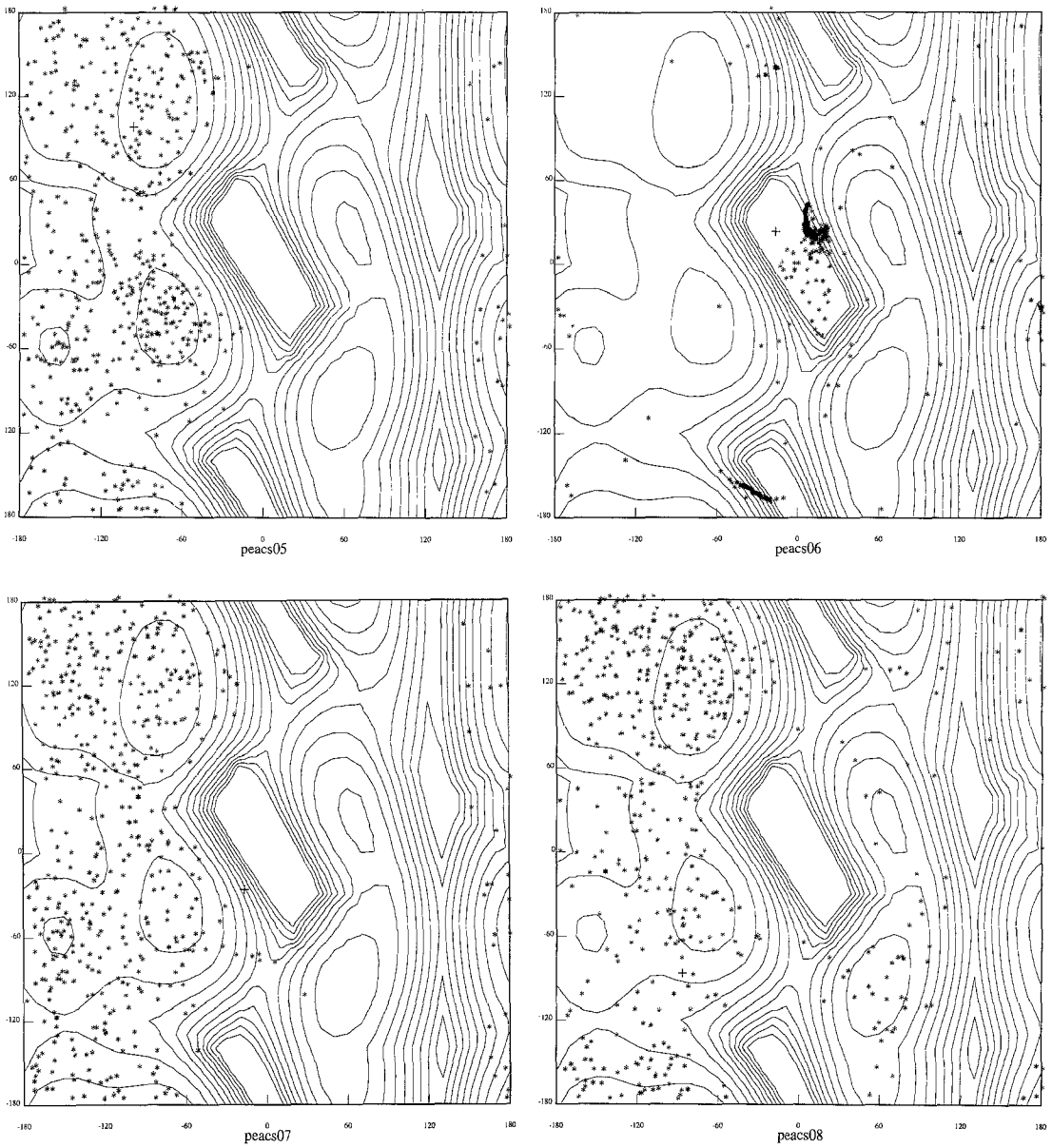
The performance was assessed in a very simple way by the number of crossings between the quadrants of the novel algorithms, compared with the standard MD (run at 310 K with a 0.5 fs time step). A check was made every 200 steps. The resulting ratio gives the 'Probing Index'.



Figs. 4.1–4.4. PEACS (points 1–4).

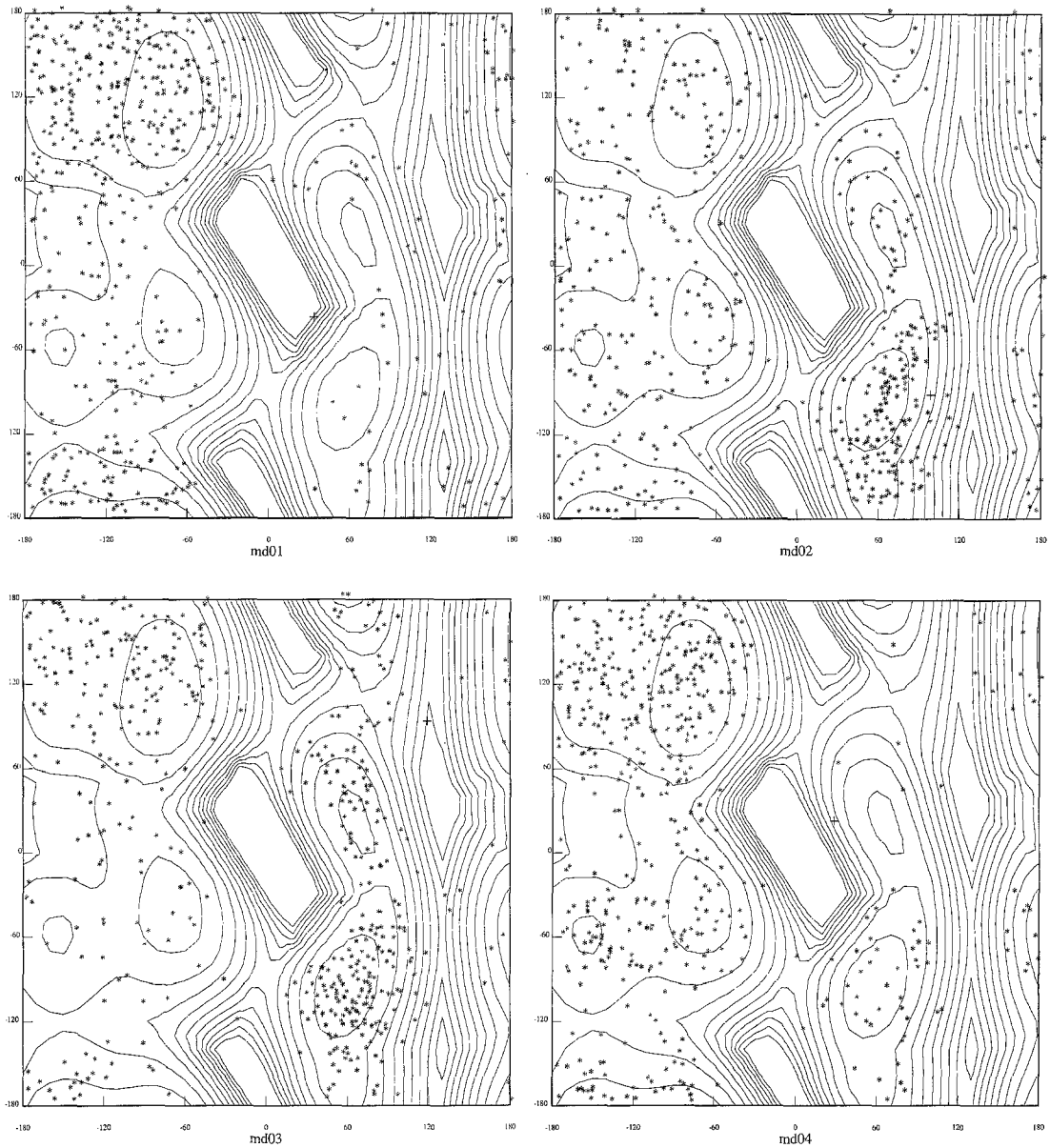
RESULTS AND DISCUSSION

Figures 2.1–2.8, 3.1–3.8, 4.1–4.8 and 5.1–5.8 show ϕ/ψ contour maps (with contours 0.8 kcal/mol apart) of alanine dipeptide with all other degrees of freedom relaxed (a flexible map) for the RUSH, HMC, PEACS and normal MD methods, respectively. The vertical axis represents ψ and



Figs. 4.5–4.8. PEACS (points 5–8).

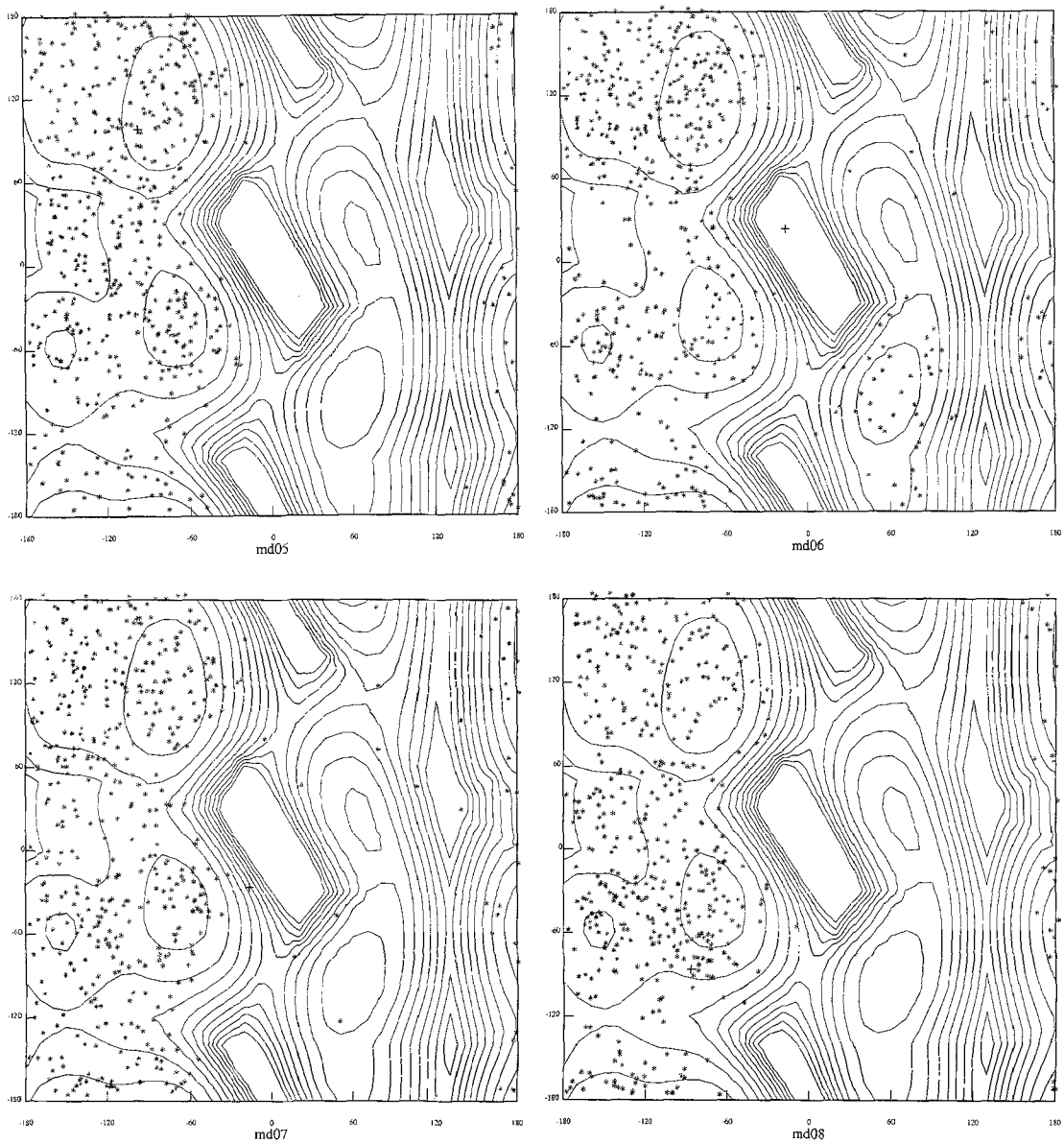
the horizontal axis represents ϕ. The starting point for each simulation is represented by a '+'. The specific dihedral angle starting values and the number of points in each quadrant, as well as the value for the Probing Index for each method are given in Table 1. Two of the HMC runs did not converge (Points 1 and 6) because of the very high energy starting points. Since the trends for the HMC runs are clear from the other runs, it was not thought important to modify the



Figs. 5.1–5.4. Normal Newtonian dynamics (points 1–4).

input parameters to achieve convergence for these runs.

As can be qualitatively seen from Figs. 2.1–2.8, RUSH dynamics most consistently samples more of the ϕ/ψ conformational space. This ability to rapidly move between different regions of local minima is reflected in the large values of the Probing Index. In fact, for all runs except one, the Probing Index was the largest for the RUSH simulations. It is also very interesting that these



Figs. 5.5–5.8. Normal Newtonian dynamics (points 5–8).

results are almost independent of the starting point, regardless of whether it is of high or low potential energy.

The normal MD algorithm had the next best set of Probing Indices (Figs. 5.1–5.8). It was able to sample conformational space relatively well. However, the amount of sampling was more dependent on the starting point than in the RUSH algorithm.

TABLE 1
PROBING INDEX AND QUADRANT OCCUPANCY NUMBER FOR STARTING CONFORMATIONS 1–8

Method	Probing index	Q1	Q2	Q3	Q4
Starting conformation 1[a]					
MD	118	43	282	146	29
HMC	–	–	–	–	–
RUSH	115	202	96	34	168
PEACS	70	364	45	75	16
Starting conformation 2[b]					
MD	95	49	127	125	199
HMC	0	0	0	0	500
RUSH	139	36	212	208	44
PEACS	95	13	192	277	18
Starting conformation 3[c]					
MD	75	85	171	58	186
HMC	5	318	0	0	182
RUSH	127	100	132	148	120
PEACS	70	80	125	120	175
Starting conformation 4[d]					
MD	83	15	271	162	52
HMC	10	494	2	0	4
RUSH	125	56	195	187	62
PEACS	47	109	32	84	275
Starting conformation 5[e]					
MD	95	14	267	203	16
HMC	54	46	411	43	0
RUSH	134	46	222	166	66
PEACS	86	7	223	255	15
Starting conformation 6[f]					
MD	83	76	218	133	73
HMC	–	–	–	–	–
RUSH	129	68	122	150	60
PEACS	66	237	92	124	47
Starting conformation 7[g]					
MD	120	22	242	220	16
HMC	43	32	340	128	0
RUSH	142	60	193	169	78
PEACS	100	10	251	222	17
Starting conformation 8[h]					
MD	99	12	221	254	13
HMC	42	0	4	415	81
RUSH	127	100	132	148	120
PEACS	88	25	284	147	44

[a] $\psi = -40.0$, $\phi = 30.0$
[b] $\psi = -95.0$, $\phi = 95.0$
[c] $\psi = 90.0$, $\phi = 115.0$
[d] $\psi = 20.0$, $\phi = 25.0$
[e] $\psi = 95.0$, $\phi = -100.0$
[f] $\psi = 20.0$, $\phi = -20.0$
[g] $\psi = -30.0$, $\phi = -20.0$
[h] $\psi = -90.0$, $\phi = -90.0$

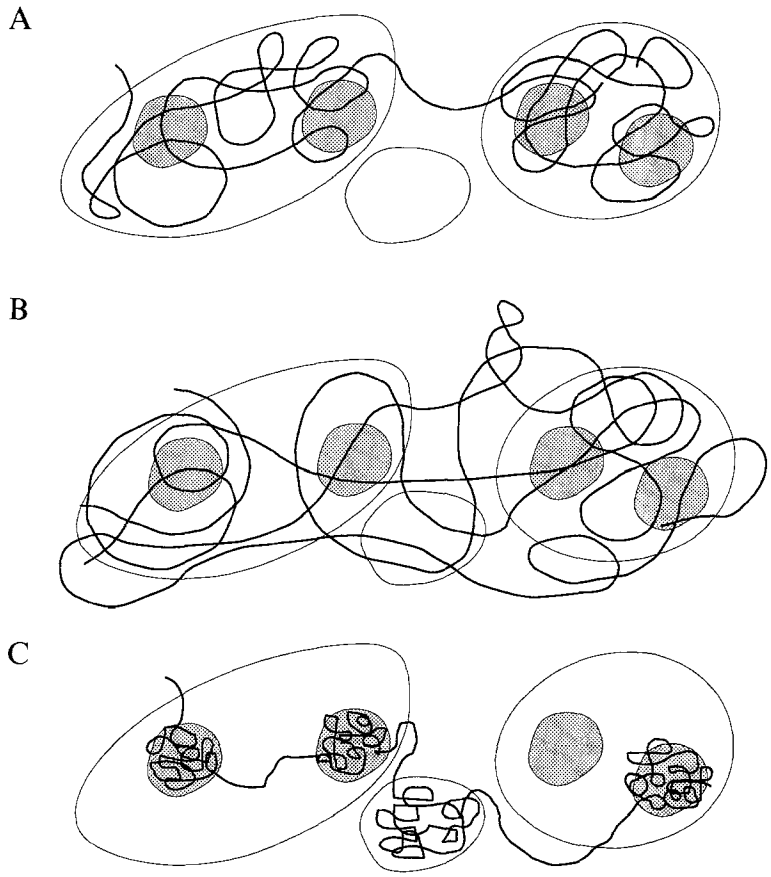


Fig. 6. Current conception of the differences between the trajectories of (A) normal molecular dynamics; (B) RUSH dynamics; and (C) Hybrid Monte Carlo. Assessing Hybrid Monte Carlo is particularly difficult, as it seems to perform quite differently on different time scales, and may perform differently again over very long time scales for which runs are not yet complete.

The PEACS algorithm performed nearly as well as the standard MD method for most points, but got trapped for a long period using the high-energy points 1 and 6 as starting points. It should be noted that much better performance could probably be achieved by PEACS for this run, if different choices for the two parameters (the strength of coupling of the system to the external potential bath and the value for the potential energy relaxation time) had been made.

Finally the HMC method had the worst performance at sampling conformational space globally, but the best at sampling the space locally about the starting point. Earlier results showed that simulations using HMC equilibrated about an order of magnitude faster than those obtained using temperature-rescaled MD [7].

CONCLUSIONS

In many senses it is somewhat unfair to compare the above methods, because they can vary drastically in performance in different contexts and in any event have somewhat different aims.

There is most generally a case for choosing MD over HMC, because studies of time sequences over short time scales and its relation to infrared spectroscopy, for example, make MD preferable as the 'purer' method of choice. Nonetheless, comparison between HMC and MD is somewhat unfair on the grounds of their essential close relationship in that MD remains the essential part of HMC, and HMC transforms smoothly into MD as the lengths of the MD trajectory component increase relative to the number of MC steps. The PEACS algorithm clearly has a dependency on the annealing schedule used. Fast cooling causes entrapment while slow cooling explores more conformational space.

The MD methods will also vary in performance according to the temperature at which the simulation is run, the cooling schedule (not tested here), and the values of any parameters present. The comparisons made here vary the starting points on the potential energy surface and study the relative ability of the methods to sample ϕ/ψ conformational space.

Nevertheless, as is shown in Fig. 6, RUSH dynamics appears to be a very promising method for quickly sampling conformational space without getting as easily trapped in local regions as the other methods. Calculations are currently underway to study how this method performs on a much larger molecule.

REFERENCES

1 Brooks III, C.L., Karplus, M. and Pettitt, B.M., Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Wiley, New York, NY, 1988.

2 a. Van Gunsteren, W.F. and Weiner, P.K. (Eds.) Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, ESCOM, Leiden, 1989.

b. Van Gunsteren, W.F., Weiner, P.K. and Wilkinson, A.J. (Eds.) Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Vol. 2, ESCOM, Leiden, 1993.

3 Cotterill, R.M.J., Platt, E. and Robson, B., In Soumpasis, M. and Jovin, T. (Eds.) Computation on Biomolecular Structures: Achievements, Problems and Perspectives, Schloss Ringberg Symp. Proc., Springer, Berlin, 1991, p. 157.

4 Li, J., Platt, E., Waskowycz, B., Cotterill, R.M.J. and Robson, B., Biophys. Chem., 43 (1992) 221.

5 Cotterill, R.M.J. and Madsen, J.U., In Bohr, H. (Ed.) Characterizing Complexity, World Scientific, Singapore, 1990, pp. 177–191.

6 Duane, S., Kennedy, A.D., Pendleton, B.J. and Rowat, D., Phys. Lett., 195 (1987) 216.

7 Chen, Y., Brass, A., Pendleton, B.J. and Robson, B., Biopolymers, 33 (1993) 1307.

8 Van Schaik, R.C., Van Gunsteren, W.F. and Berendsen, H.J.C., J. Comput.-Aided Mol. Design, 6 (1992) 97.

9 Robson, B. and Platt, E., J. Mol. Biol., 188 (1986) 259.

10 Robson, B. and Garnier, J., Introduction to Proteins and Protein Engineering, Elsevier, Amsterdam, 1986.