ELSEVIER

TRAJAN: A tool for analyzing trajectories from molecular simulations[1]

Graham A. Worth,[2] Christophe Lecuyer and Rebecca C. Wade*

European Molecular Biology Laboratory, Heidelberg, Germany

Molecular dynamics simulations of biological systems are notoriously difficult to analyze because of the complexity of the information that they contain. We describe a new method for analyzing trajectories from simulations in order to extract important features of the motion. The trajectory of each atom is partitioned into conformation wells, in each of which its motion is assumed to be predominantly harmonic oscillation around an average position. Thus each atom moves anharmonically through a sequence of wells during the trajectory. The movement of atoms between wells, their ellipsoids of motion within each well, and correlations in the motion of atoms are quantified and can be visualized with molecular graphics. The TRAJAN analysis procedure is applicable to trajectories from equilibrium and nonequilibrium simulations and is not restricted to molecular dynamics simulations. Its application is demonstrated for a range of model systems.

Keywords: molecular dynamics analysis, trajectory visualization, harmonic motion, low-frequency motion, protein flexible loop motion, dichloroethane, peptide, neutral protease

INTRODUCTION

It is often difficult to extract the important information from simulations of molecular systems, even those as small as a tetrapeptide. One reason is that simulations generate huge amounts of data. For example, molecular dynamics (MD) simulations of biological systems are typically 0.1–1 nsec long and coordinate sets are usually recorded every 0.1–5 psec. Another is that biomolecules such as proteins are large heterogeneous molecules with very complex energy landscapes with many local energy minima. Detecting the significant larger scale motions among the many smaller fluctuations can be very difficult under these circumstances.

A number of techniques have been proposed to facilitate the analysis of molecular dynamics trajectories:

Projection of an MD trajectory onto normal mode axes[1] in order to extract low-frequency motions. Comparison of amplitudes for an ideal harmonic case with those from the MD simulation enable anharmonic motion to be analyzed. This approach is, however, restricted to equilibrium simulations.

Analysis of the covariance matrix of the positional deviations[2] in order to extract the "essential dynamics" of proteins.[3,4] This differs from the normal mode approach in that it does not make use of a harmonic description and does not use atomic masses for the analysis. Typically, only a few degrees of freedom are found to account for most of the (anharmonic) positional motion.

Use of digital signal-processing techniques to filter out high-frequency motions.[5–7] In the FOCUS program,[8] data is Fourier transformed, cut, and backtransformed so that the high-frequency motions are removed.

Use of clustering techniques to define the conformational space visited in a simulation and how much time is spent in each distinct part of the space is visited. Clustering has been performed by fuzzy cluster analysis based on the root mean square deviation (rmsd) of Cartesian coordinates[9] and by self-organizing neural nets[10] based on the rmsd of dihedral coordinates.

Special visualization techniques, e.g., dials and windows.[11]

Our approach differs from previous techniques in that it explicitly recognizes that biomolecular motions are typically a combination of harmonic and anharmonic motions. We thus abstract the simulation trajectory to a model in

Color Plates for this article are on pages 146–147.

Address reprint requests to: Rebecca C. Wade, European Molecular Biology Laboratory, Meyerhofstr. 1, 69012 Heidelberg, Germany.
Received 13 February 1996; revised 4 June 1996; accepted 11 June 1996.

[1] TRAJAN is an acronym for TRAJectory ANalysis. Note also that Trajan was a Roman Emperor (AD 98–117). Trajan's column is still standing in Rome. In a somewhat analogous way to the TRAJAN program, it shows the important events during his two campaigns (trajectories) into Dacia.

[2] Present address: Theoretische Chemie, Universität Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany.

655 Avenue of the Americas, New York, NY 10010

S0263-7855/96/$15.00
PII S0263-7855(96)00059-8

which each atom moves (anharmonically) through a series of conformation wells, in each of which it oscillates (approximately harmonically) about an average position. Each atom will move in a different number of wells with different fluctuations, and transitions between wells will take place at different times. Conformation wells are defined on the basis of Cartesian coordinate rmsd (in a similar way to a cluster analysis) and time. Atomic motion within each well is quantified with a harmonic model. Correlations in the motion vectors of different wells and with the interwell vectors are computed. Motion between wells and thermal ellipsoids of motion within wells are easily visualized using a standard molecular graphics package. The complexity of an MD trajectory is thus reduced to a series of low-frequency transitions between coordinates about which high-frequency oscillatory motion is performed. This enables simultaneous or coupled motions of atoms to be detected easily.

We demonstrate this method for three very different examples. The first is dichloroethane (DCE). The energy profile for rotation around its principal dihedral angle is known[12] and thus DCE provides a good example system for testing the analysis procedure. The second is tetraalanine peptide (AAAA). We expect the method to be very useful for analyzing the dynamics of peptides and AAAA provides a simple example. The third example is provided by a set of neutral proteases, and shows how the motion of flexible peptide loops in proteins can be analyzed. Neutral proteases vary in their thermostability and we have studied two of them: thermolysin with a T50 (incubation temperature at which neutral protease activity drops to 50% in 30 min) = 82°C and the *Bacillus subtilis* neutral protease (B. Sub. Npr) with a T50 = 55°C at pH 7.[13] It is thought that local unfolding processes are responsible for making the enzymes susceptible to autolytic cleavage at high temperatures.[13,14] In an engineered B. Sub. Npr, nicking of the inserted 247–258 residue β-hairpin loop (which is also present in thermolysin) is observed, suggesting that as temperature increases its flexibility increases, and it swings away from the protein to a position where it can be attacked by another protease molecule. Mutagenesis studies of B. Sub. Npr[15] indicate that the residues important in stabilizing the β hairpin are those at its tip that interact with another surface loop (residues 210–228) and the C terminus. The aim of the molecular dynamics simulations described here is to understand the factors influencing the mobility and structural stability of the β-hairpin loop and its susceptibility to cleavage.

In the next section, we detail the theory employed in the trajectory analysis, its implementation in the TRAJAN program, and its application to simulations of example systems. The results are given and discussed in the subsequent section. We conclude with a summary of the capabilities of this method of trajectory analysis.

METHODS

Detection of conformation wells

A simple model of atomic motion in a protein is that an atom moves in a potential well, defined by its environment, from time $t = 0$ to $t = t_1$. After this time, the environment has changed enough to cause a transition to a different energy well, which is then occupied until time $t = t_2$, and so on. Each time an atom undergoes such a transition, sometimes as a concerted movement with other atoms, sometimes alone, the protein moves to a new conformational substate.

Assuming that in each well, which we shall call a "conformation well," the atom moves under the influence of an approximately harmonic energy potential, the average coordinate from its trajectory defines the position of the energy minimum. Any significant sample of the trajectory during this time should also have the same average coordinate, i.e., for the *n*th minimum of the *i*th atom,

$$\langle q_i \rangle_{t_{n,\text{initial}}}^{t_{n,\text{final}}} = \langle q_i \rangle_{t_a}^{t_b} \qquad (1)$$

where $\langle q_i \rangle$ is the average x, y, or z coordinate of atom i between the given time limits, and the times t_a and t_b lie between the initial and final times of the *n*th minimum ($t_{n,\text{initial}}$ and $t_{n,\text{final}}$).

The procedure used to define the different wells is shown schematically in Figure 1. The trajectory for one atom is initially split up into short consecutive windows of equal duration that may overlap, and the average coordinate for each window is calculated. The whole trajectory is then assigned as the first conformation well and its average coordinate calculated. This is compared to the average coordinates of all the windows. If any window has a significantly different average value, the trajectory assigned to the first conformation well is shortened in length by one window, and the comparison of the average coordinate of each window with that of this new, shorter conformation well is repeated. This is repeated until all the average coordinates in

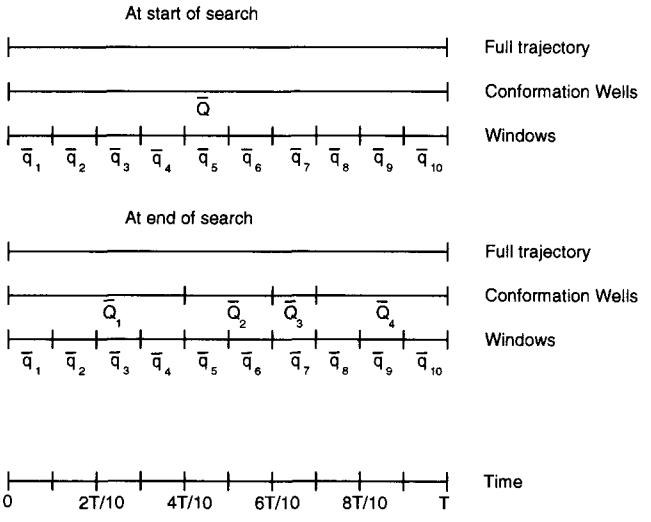


Figure 1. Schematic representation of how the trajectory of an atom is partitioned into conformation wells. At the start of the analysis, all the coordinates for an atom are ordered in time in the "full trajectory," which is of length T. The trajectory is divided into equal-length "windows," in this example 10 windows of length T/10. The average coordinate of the ith *window is* $\bar{q}_i$ *and the average coordinate of the full trajectory is* $\bar{Q}$. *After the search, the full trajectory has been divided into "conformation wells," with the average coordinate of the* jth *well being* $\bar{Q}_j$; *and the distance between the* $\bar{Q}_j$ *and the* $\bar{q}_i$ *coordinates for the windows in this well being less than the specified distance* d_{cur}

all the windows within the first conformation well agree within a user-defined cutoff distance with that of this conformation well. The analysis then moves on to the remaining part of the trajectory, assigning it to the second conformation well, before comparing its average coordinates to the corresponding windows and adjusting the well length as before. This is repeated for subsequent wells until the full trajectory has been assigned to conformation wells. The algorithm is shown in detail in Figure 2.

The cutoff distance, d_{cut}, used for the comparisons is defined by

$$(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \leq (d_{cut})^2 \quad (2)$$

where Δx, Δy, and Δz are the differences between the coordinates for the whole conformation well and the window average. The user can control the fineness and accuracy of the search by making an appropriate choice of cutoff distance, window length, and window overlap.

Characterization of conformation wells

The second part of the analysis is to construct covariance matrices for the coordinates of each atom in each conformation well:

$$\begin{bmatrix} \Delta x^2 & \Delta x\Delta y & \Delta x\Delta z \\ \Delta y\Delta x & \Delta y^2 & \Delta y\Delta z \\ \Delta z\Delta x & \Delta z\Delta y & \Delta z^2 \end{bmatrix} \quad (3)$$

where Δx, Δy, and Δz are the deviations from the average coordinate. The trace of this matrix is related to the experimental thermal fluctuation or B-factor of the atom by[16]

$$B = \frac{8\pi^2}{3}(\Delta x^2 + \Delta y^2 + \Delta z^2) \quad (4)$$

The eigenvectors of the covariance matrix (3) define axes along which, in the harmonic approximation, the coordinates will be normally distributed with respect to their average coordinates. The square roots of the covariance matrix eigenvalues, the standard deviations, are then a measure of the percentage of the coordinates within the motion or thermal ellipsoid defined by these axes. Thermal ellipsoids can be used to show the relative sizes of the conformation wells.

As the motion axes are vectors, correlations between the motion of different atoms can be computed from their scalar products. Correlations between the motion axes and the vectors joining consecutive minima can be computed similarly in order to see whether the atoms tend to make transitions governed by the axes of their motion or by other influences.

The B-factors and average deviation from the starting positions are also averaged over specified regions of interest. For example, for the neutral proteases, these regions were defined as the whole protein, the residue 247–258 β hairpin, and the residue 210–228 exposed loop in order to compare the mobilities of these regions across a set of proteins.

After analysis, the coordinates of all the conformational substates detected are output to allow the motion of the protein to be viewed with the high-frequency atomic fluctuations removed. (We define a conformation well for a single atom and a conformational substate by the conformation wells for all the atoms in the molecule that are selected for analysis.) A new conformational substate is defined whenever an atom moves to a new conformation well. However, in order to filter out conformational substates that are occupied only for a very short time, a new conformational substate is defined only if the occupation time of the current conformational substate is at least 0.02 times the occupation time of the new conformation well. If this is not the case, the current conformational substate is redefined to include the new atomic conformation well. The conformational substates of selected atoms, e.g., a loop region or the C_α atoms of a protein, may be output, along with their respective motion axes in each of the conformation wells. This permits easy visualization of the relative magnitudes of motion of the different parts of the molecule.

A molecule may visit the same conformational substate more than once during a trajectory. If desired, this may be detected after all the conformation wells and correlations have been computed. The molecule is defined as being in the same conformational substate as another visited earlier when the following criterion is satisfied:

$$\sum_{i=1}^{N_{atom}} (\Delta x_i)^2 + (\Delta y_i)^2 + (\Delta z_i)^2 \leq (d_{cut})^2 \, N_{atom} C_{av} \quad (5)$$

where Δx_i, Δy_i, and Δz_i are the differences in conformation well coordinates of atom i in the two conformational substates, N_{atom} is the number of mobile atoms, and C_{av} is a user-defined averaging criterion between 0 and 1. The smaller C_{av} is, the stricter the definition of a conformational substate. C_{av} should be chosen to suit the system to be analyzed, but we have found that a value of 0.4 usually gives reasonable results. When the criterion in expression (5) is satisfied, the atomic coordinates are averaged. This procedure is repeated for all the conformational substates in the trajectory. The average coordinates of each conformational substate are written in the output coordinates file in the order in which they first appear. Thermal ellipsoids and B-factors for the conformational substates are calculated by averaging those of the conformation wells in each substate.

Data pretreatment

Depending on the system simulated, superposition of all coordinate frames may be necessary before analyzing the atomic motions. This can either be done by a least-squares fit of all or selected atoms, or by the exact superposition of three covalently connected atoms by application of one translation and two rotation matrices. The superimposed coordinates are stored for analysis and graphical display.

The user should select the atoms for which the analysis will be done. For small systems, all atoms can be analyzed, but for larger systems it is usually better to select a subset of atoms for analysis, e.g., C_α atoms in a peptide, or the residues in a particularly flexible loop of a protein.

The user must specify the initial window length, window overlap, and d_{cut} and the atoms for which d_{cut} is computed. There is a playoff between these parameters. In the limit of a window length of 1, i.e., each coordinate set is a separate window and there is no window overlap, d_{cut} must be large enough to catch all the fluctuations around the mean position in a conformation well. In the other limit, when the

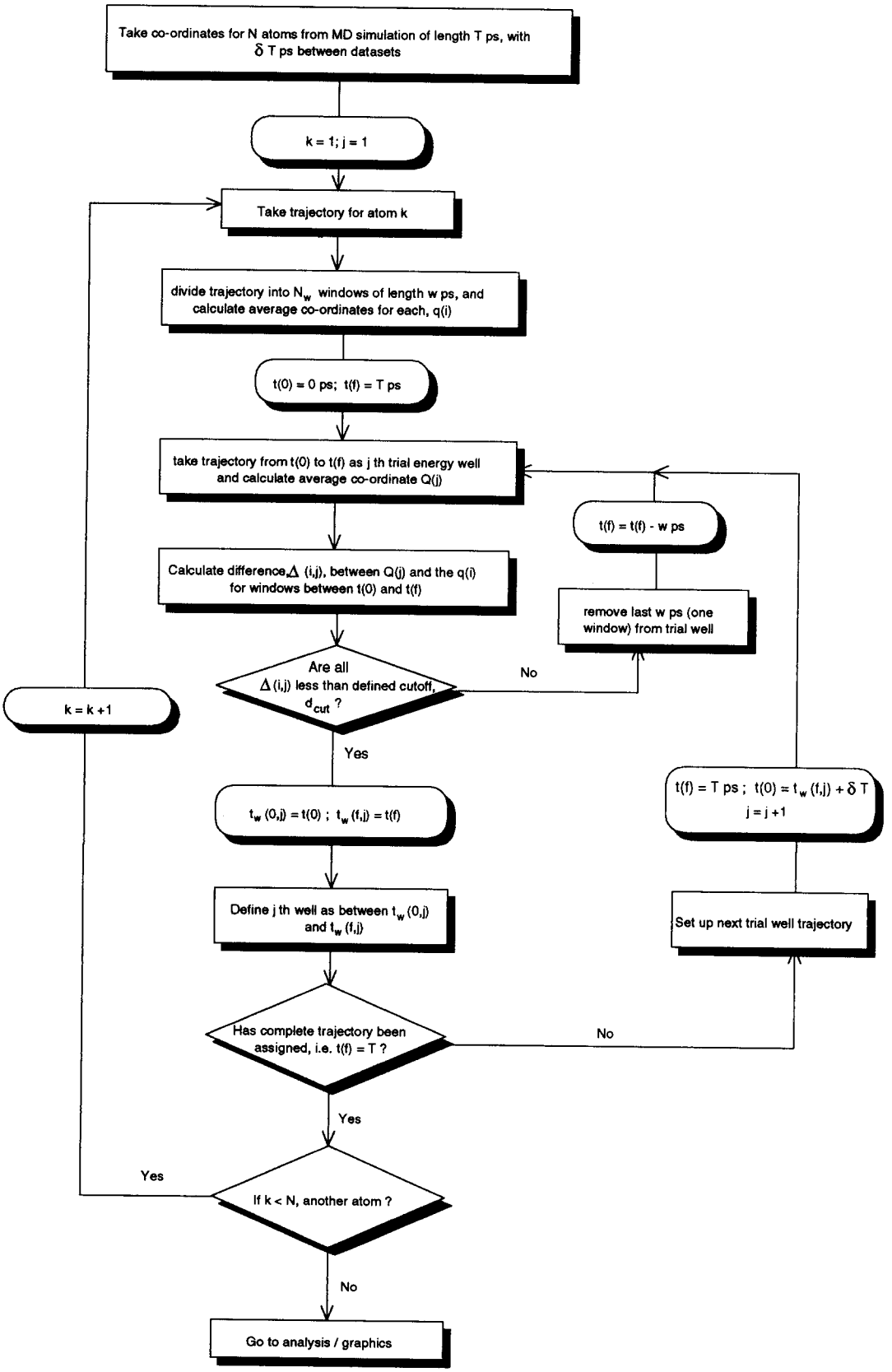


Figure 2. Flow chart of the algorithm to assign conformation wells in the TRAJAN program. The program also carries out data pretreatment (atomic superposition) and posttreatment (assignment of unique conformational substates, regional mobility computation).

window length is equal to the conformation well length or at least large enough to be completely representative of the conformation well and there is no window overlap, d_{cut} can be zero. In practice, finite sampling and the unknown nature of the atomic fluctuations mean that these parameters should lie between these limits. Nonoverlapping windows with a length of 10 coordinate sets for coordinates recorded at 0.2-psec intervals is reasonable for most MD trajectories. d_{cut} must be specified according to the motions of the system studied. For a peptide simulated by MD, a d_{cut} of 1.5–2.0 Å for C_α atoms is usually suitable considering the amplitude of typical thermal fluctuations. We also expect the TRAJAN program to be useful for the analysis of Brownian dynamics (BD) simulations of peptides, protein loops, and even enzyme–substrate encounter, in which a reduced representation is used in order to simulate longer trajectories than are possible with MD. The motion of particles representing groups of atoms rather than single atoms would be analyzed, and the TRAJAN program would be used simply as a cluster analysis tool. In this case, a larger d_{cut}, e.g., ~4 Å for particles representing residues, would be used.

Implementation

Analysis is performed by the TRAJAN program, which was written specifically for this purpose in Fortran 77. There are three input files:

1. A file specifying all input parameters (.cai file)
2. A Brookhaven PDB format file containing the atom names and types in the order in which they are to be analyzed
3. A binary trajectory file containing atom coordinates at equally spaced times. Trajectory files can be read in .coo format from the ARGOS MD program (versions 5.8 and 6.0)[17] and in .dcd format from the CHARMm[18] and UHBD[19] programs. The user must specify the timestep when analyzing a .dcd format file

The program produces a log file and files detailing the conformation wells and conformational substates detected (.anl file), atomic correlations (.cor file), and mobility parameters of specified regions (.reg file). In addition, it produces trajectory files of the conformation wells and conformational substates (in the .coo format of the ARGOS program). To display the motions through conformation wells and the thermal ellipsoids with the QUANTA molecular graphics program suite, the pdb and .coo files should be converted to msf and dcd format files using the auxiliary programs pdb2msf and gendcd, respectively.

Simulation and analysis protocols for example systems

All simulations were performed with the ARGOS molecular dynamics program[17] on Silicon Graphics workstations.

Dichloroethane 1,2-Dichloroethane (DCE) was simulated *in vacuo* using a united atom representation and GROMOS force field parameters.[20] With this model, there is only one dihedral angle, Θ, in DCE and the energy profile for rotation around it is given by an analytical expression as the sum of three cosines. The energy profile has three potential wells with minima at Θ = 60, 180, and 300°. The deepest well at Θ = 180° is symmetric whereas the other two are asymmetric. Two simulations were carried out at 300 K for 12 psec with a timestep of 2 fsec and bonds constrained with the SHAKE algorithm with a tolerance of 0.00001 nm^2[1]: one starting with Θ = 180° in an NVE[3] ensemble (with no velocity reassignment) and the other starting with Θ = 0° in an NVT[3] ensemble (with a temperature relaxation time of 0.01 psec). Six thousand coordinate sets were collected (every step) and superimposed by overlaying one of the chlorine atoms and the two carbon atoms. Analysis with the TRAJAN program was carried out with nonoverlapping windows of 10 coordinates sets and d_{cut} = 1.68 Å (for the remaining chlorine atom) to determine the conformation wells, and C_{av} = 0.4 to average conformation wells to compute conformational substates.

Tetraalanine peptide Tetraalanine (AAAA) was simulated *in vacuo* with explicit polar hydrogen atoms using the GROMOS force field.[22] The N and C termini were blocked with acetyl and methylamide groups, respectively. AAAA was built in an extended conformation, heated to 300 K in 10 psec, and then simulated for 100 psec in an NVE ensemble with velocity reassignment every 100 steps at 1 000 K. A timestep of 2 fsec was used. Five hundred coordinate sets were collected (every 0.2 psec) and superimposed by performing a least-squares fit of C_α atoms. Analysis with the TRAJAN program was carried out with nonoverlapping windows of 10 coordinate sets and d_{cut} = 1.5 Å.

Neutral protease proteins The GROMOS force field[22] was used with explicit polar hydrogens and the SPC/E water model.[23] All bonds were constrained with the SHAKE algorithm with a tolerance of 0.001 nm^2[1] and a timestep of 2 fsec was used. Although it is known that the presence of solvent affects the motion of residues on the surface of a protein, for reasons of simplicity the simulations were done with selected crystal waters only and a relative dielectric constant of 5 to approximate the screening effect of the dielectric environment provided by water.

The starting structure of thermolysin was taken from the crystal structure in the Brookhaven Protein Databank (reference 3TLN).[24] All titratable residues were assigned their usual protonation state at pH 7 on the basis of visual inspection of the structure except for His-105, which was treated as double protonated. The positions of other histidine protons was chosen to maximize hydrogen bonding with neighboring residues. Histidines 142, 146, and 231 were protonated on the δ nitrogen while histidines 74, 88, 216, and 250 were protonated on the ϵ nitrogen. Protons were added to the protein with the ARGOS program and to the crystal waters with the graphics program QUANTA3.2.[25]

Thermolysin contains four calcium ions and one zinc ion. Energy minimization with full $+2e$ charges on these ions caused distortion from the crystal structure, with the ligands moving too close to the ions. Reduced charges were therefore assigned as follows. Shen et al.[26] calculated charges for simulation of aspartate, glutamate, and histidine ligands of zinc and calcium in superoxide dismutase. Taking these values to model the polarization produced by the ion, his-

[3] NVE and NVT denote quantities kept constant during an MD simulation: N, number of particles; V, volume; E, energy; and T, temperature.

tidine has an overall charge of +0.32*e* while aspartate and glutamate have overall charges of –0.45*e*. The charge on an ion, q_{ion}, is then given by

$$q_{ion} = +2 - 0.32N_{HIS} - 0.55N_{ASP} - 0.55N_{GLU} \quad (6)$$

where N_X is the number of ligands of type *X* around the ion. The charge on the zinc ion was therefore 0.81*e*, while the calcium ions had charges of 0.63, 1.18, 0.9, and 1.45*e* (given in the order found in the 3TLN Brookhaven file). With these charges, the crystal geometry was retained on energy minimization.

One further change was made to the crystal structure. The N terminus is pointing directly at the amide nitrogen of glutamine 31, while the isoleucine 1 backbone oxygen is pointing at the amide oxygen of the same residue. This arrangement leads to distortion of these residues on energy minimization, and so the amide group of glutamine 31 was rotated 180° to form two hydrogen bonds with the isoleucine 1 backbone before energy minimization.

The added protons were energy minimized and then relaxed with high-temperature MD (10 psec in an NVE ensemble at 500 K followed by 10 psec in an NVT ensemble at 350 K with a temperature relaxation time of 0.01 psec) with the nonhydrogen atoms kept fixed. All water molecules further than 4 Å from the ions were then discarded and all remaining atoms were energy minimized.

The initial B. Sub. Npr structure was a model built by G. Vriend[14,27] on the basis of homology to the structure of thermolysin. The β hairpin corresponding to residues 248–257 in thermolysin was inserted in this structure. The mutations T249V and V256T were made to change this hairpin to the *Bacillus stearothermophilus* sequence that was used in the mutation experiments. Protons were added and the same protonation states assigned as for thermolysin. (Histidines 142, 146, and 231 were again assigned protons on the δ nitrogen, while histidines 74 and 105 were assigned protons on the ϵ nitrogen and histidine 105 was doubly protonated.) The charge on the zinc ion was 0.26*e* and on the two calcium ions in this structure was 0.63*e*.

The sequence LPNT before aspartate 223 was absent from the homology model as it has no counterpart in thermolysin. As these residues were in one of the loops to be studied they were added. A search of the structural databases showed no obvious structure for these residues and so they were inserted in random conformations. In an analogous manner to thermolysin, the added parts of the protein were relaxed, i.e., the protons, residues 245 to 260 (the β hairpin), and residues 210 to 230 (encompassing the added residues LPNT). After energy minimization and 10 psec of NVE at 500 K, 160 psec of NVT simulation at 350 K was performed. After this time, however, the added loop had not adopted a reasonable fold and appeared to be "knotted up." To cure this, a short simulation using the ARGOS program's "dynamic docking" option was made on this region. This is in effect a free energy perturbation simulation without collecting the energetic data. Starting with very small intramolecular energy terms on the residues in the loop, a 40-psec simulation of 20 windows was run. At the end of each window, the intramolecular energy terms on the loop were increased until, for the last 2 psec, the full values were used. At the end of this simulation, the loop had a more reasonable fold and the potential energy had decreased.

At the end of this relaxation, the protein was fully energy minimized. To reduce computing time, constraints were added to reduce the size of the system. The neutral proteases have two domains. As both the loops of interest are in the C-terminal domain, the N-terminal domain was fixed. The atoms in the hinge region between the domains were restrained with a harmonic potential of 250 kJ/mol/nm^2 to act as a buffer and the remaining atoms were left free. Both thermolysin and B. Sub. Npr were then heated to 350 K and 300 psec of MD was run under NVT conditions, collecting coordinates every 0.2 psec. At the end of the B. Sub. Npr simulation, short (50 psec) slow growth perturbations were made to produce a further four mutants: L315A, L315G, N243L, and the combination N243L–N216H–K305S. After an equilibration time of 20 psec, 300 psec of MD was run for these structures.

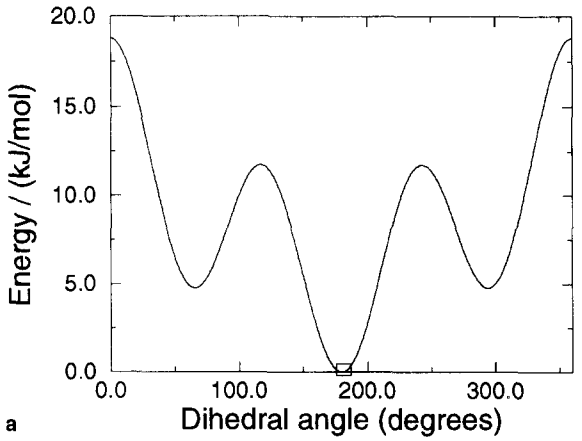
The trajectories from the final 200 psec of each neutral protease simulation were analyzed with the TRAJAN program. A d_{cut} value of 1.8 Å and nonoverlapping windows of 10 coordinate sets were employed. With these parameters, the buffer zone atoms, which were restrained by harmonic potentials, moved in one conformation well only.

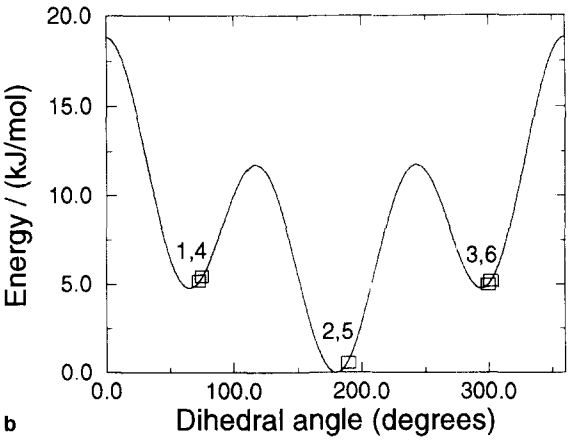
RESULTS AND DISCUSSION

Dichloroethane

Dichloroethane is a simple molecule that has only one dihedral angle when modeled with a united atom representation. The energy profile for rotation around this dihedral angle is known analytically and thus DCE provides a good test case for the TRAJAN analysis. The results for DCE are displayed in Figure 3. When simulation was started at the energetically most favorable position with $\Theta = 180°$ and with the total energy of the system kept constant, the molecule only sampled one conformation well (or conformational substate as only one atom moves in this molecule when translational and rotational motion are removed) and lacked sufficient kinetic energy to cross the barrier to move to another (see Figure 3a). It performed harmonic motion about the lowest energy position and thus the average position of the conformation well is at 180°.

Figure 3b and c displays results when simulation was started at the energetically most unfavorable position with $\Theta = 0°$ and performed under NVT conditions. Figure 3b shows the conformation wells visited during the first 3.2 psec of the simulation. In this simulation, DCE had sufficient kinetic energy to sample more than one conformation well and in this time Θ rotated through 360° twice. The molecule moved through six conformation wells corresponding to three unique conformational substates. As sampling about the energy minima was not completely symmetric within each energy well, the computed positions of the conformation wells differ slightly from the positions of the energy minima. The conformation well position for the symmetric well deviates from the energy minimum position because the position of this conformation well is dependent on the positions of the other two wells. d_{cut} was chosen as exactly the distance (1.68 Å) between conformations at energy barriers between wells for rotation around the C–C





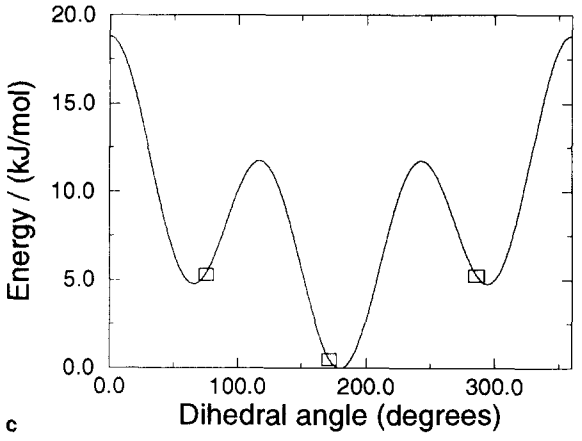


Fig. 3. Analysis of 12-psec molecular dynamics trajectories of 1,2-dichloroethane in vacuo *at 300 K. The line shows the energy profile for rotation of angle Θ around the C–C bond for the force field used. Only one conformation well or conformational substate is visited during the simulation started from Θ = 180° under NVE conditions with no velocity reassignment (a). All three conformational substates are visited when simulation is started from Θ = 0° under NVT conditions (b and c). (b) shows the order in which the conformation wells are visited during the first 3.2 psec (wells 2 and 5 overlap) while (c) shows the positions of the conformational substates obtained from averaging conformational well positions during the full 12-psec simulation.*

bond. Thus, if the conformation well position for an unsymmetric energy well is shifted, so will the conformation well position for a symmetric energy well be. Figure 3c shows the conformational substates visited during 12 psec obtained by averaging the conformation wells visited during the simulations. Again, it can be seen that the conformation well positions are computed near, but not exactly at, the potential energy minima.

These results show that the TRAJAN program is able to detect conformational substates at conformations approximately corresponding to energy minima.

Tetraalanine peptide

Forty-three sequentially visited conformational substates were located for the trajectory of AAAA when all atoms were considered and 11 when only C_α atoms were analyzed. Structures for the latter are displayed in Figure 4 with the thermal ellipsoids of motion of the C_α atoms. It is immediately apparent that the C_α atom of residue 4 performs oscillatory motion within the same conformation well throughout the simulation (its computed B-factor is 29.8 $Å^2$). In the first 50 psec, only the C_α atom of residue 2 explores more than one conformation well. In the second 50 psec, the C_α atoms of residues 1 and 3 also start to explore more conformation wells. Correlations in motion can be observed. For example, in conformational substates 7–8 (64–84 psec), where residue 2 is undergoing the largest amplitude motions in the whole trajectory (its computed B-factor for this well is 82.2 $Å^2$), the directions of the major axes of motion of the C_α atoms of residues 1 and 2 are very similar (scalar product, 0.994). Correlation in the direction of motion of these atoms persists in conformational substate 9 before being broken in substates 10–11. Clearly this example is somewhat artificial given the high temperature and lack of explicit solvent in the simulation. However, it can be expected that similar sorts of features would be detected in analysis with the TRAJAN program of much longer room-temperature simulations of peptides in explicit solvent.

Neutral proteases

Analysis of the motion of the C_α atoms in the thermolysin trajectory showed 93 sequentially visited conformational substates. The C_α traces for six of them spanning the space represented by all the conformational substates are shown in Color Plate 1a. The complete absence of motion in the N-terminal domain is because this region was fixed during the simulations. The major motion is clearly between residues 210 and 228 (at the top of Color Plate 1a) and in the β hairpin between residues 247 and 258 (to the right in Color Plate 1a). In particular, the latter is seen to unfold away

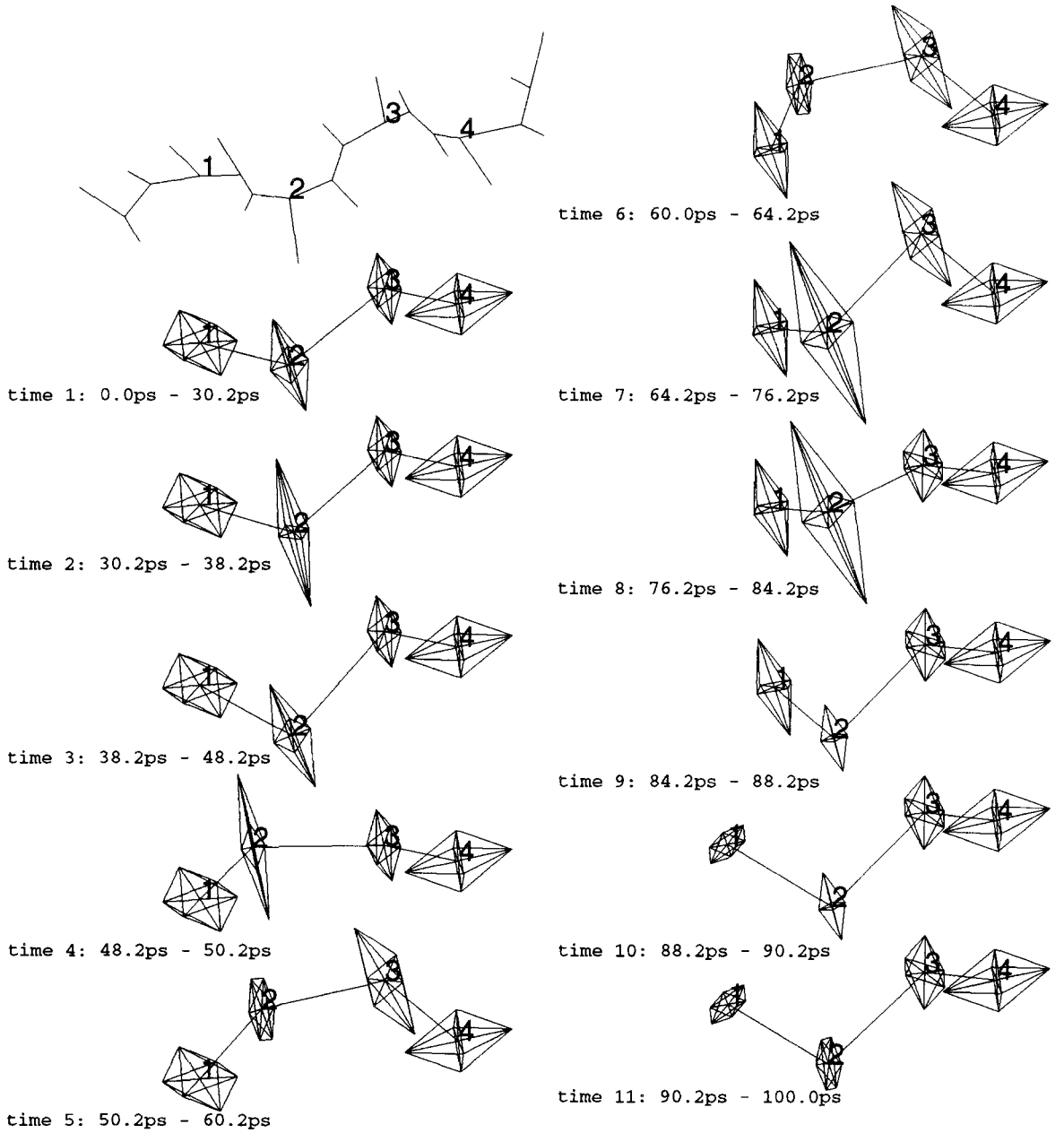


Figure 4. The 11 sequentially visited conformational substates identified from analysis of C_α atom motions during 100-psec trajectory of tetraalanine (full atomic structure shown top left) simulated in vacuo *at high temperature. Thermal ellipsoids showing motion of the C_α atoms in their conformation wells are displayed.*

from the protein as the simulation progresses. The remaining residues in the C-terminal domain either stay in a single conformation well, or move between a few wells that are close together.

Color Plate 1b and c shows snapshots from the original trajectory (used as input for the TRAJAN program) and demonstrates how the analysis has clarified the protein dynamics. In Color Plate 1b the snapshots are taken at the start of each conformation well, while in Color Plate 1c they are taken at regular 40-psec intervals. The high-frequency motion of the atoms, particularly in the core of the protein, makes the underlying motion much harder to see. In addition, the selection of snapshots at regular intervals in Color Plate 1c means that part of the motion of the loop is missed. Color Plate 1a shows that the β hairpin moves down half way, then up toward its initial position, then down fully and then starts to return upward. From Color Plate 1c, it appears that the loop moves continually downward until, toward the end of the trajectory, it starts to move up—thus, some of the large-scale oscillatory features are not detected.

Color Plate 2 shows an example of how the motion axes, obtained from diagonalization of the covariance matrices for each atom, show the relative degree of movement of the atoms. In particular, it is noticeable from the long axes of the motion ellipsoids that the β-hairpin residues not only move between very different conformation wells, but also move much further in a well than the other residues. This suggests that the potential energy surface for these residues is a set of wide and rather flat energy minima. Atoms that show significant correlations in their motion axes are highlighted in blue (residues 249–252 and 256) and are on one side of the β hairpin. The thermal ellipsoids of the $C_{\gamma 2}$ atom of residue 256 and the hydroxyl proton of residue 251 are highly correlated, indicating that concerted motion is also present between side chains in close proximity.

In contrast to thermolysin, neither B. Sub. Npr nor its mutants are very mobile. Analysis of C_α coordinates in the B. Sub. Npr trajectory resulted in only 22 sequentially visited conformational substates. These were all close together, and unfolding on the scale of the β hairpin shown in Color Plate 1 was not observed. There are closer contacts between the β hairpin and the 210–228 loop and the C terminus and coupled transitions between conformation wells can be observed for C_α atoms in these different regions of the sequence. Trajectories for the mutants displayed similar dynamics. To quantify the relative mobility of the five B. Sub. Npr-related proteins and thermolysin, the B-values were calculated as explained in Methods, and averaged over the backbone atoms of various regions of interest. There were the whole protein (ignoring fixed and restrained atoms), the 210–228 loop, and the 248–257 β hairpin. These values are listed in Table 1[28] with the proteins listed in order of decreasing thermostability.

Although caution should be exercised in interpreting the simulations owing to the lack of an explicit representation of solvent; it can be noted that Table 1 shows a clear inverse correlation between mobility in the two surface loops and thermal stability. This is in contrast to the accepted hypothesis that thermal instability results from mobility. While this unexpected correlation may be an artifact of the simulation conditions (inaccuracies in the force field or insufficient sampling), it suggests that the free energy barrier for unfolding of the β-hairpin loop in B. Sub. Npr and its mutants is much greater than in thermolysin, in part because of interactions with a more rigid 210–228 surface loop. One possibility to account for the relative stabilities is that, once the loop has swung out from the protein, it is more likely to stay out in the case of B. Sub. Npr than in thermolysin, allowing time for it to be attacked by another protease molecule. Another possibility is that different degradation mechanisms operate for thermolysin compared to the B. Sub. Npr. These possible mechanisms could be tested in further simulations with the application of special sampling techniques to force the hairpin to swing away from the protein in B. Sub. Npr. or techniques such as Brownian dynamics, with which longer time scales can be explored. Analysis of trajectories from such simulations could also be performed with the TRAJAN program.

Table 1. Average B-factors (in Å^2) calculated for different regions of the neutral proteases simulated[a]

	All residues	Residue 247–258 loop (β hairpin)	Residue 210–228 loop
Thermolysin	15.1	38.3	24.0
B. Sub. Npr mutant 1[b]	16.1	34.0	20.8
B. Sub. Npr mutant 2[c]	15.4	32.4	20.8
B. Sub. Npr mutant 3[d]	14.1	32.6	17.6
B. Sub. Npr mutant 4[e]	13.1	28.7	15.0
B. Sub. Npr mutant 5[f]	14.1	29.6	17.6

[a]The proteins are listed in order of decreasing thermostability.[15,28]
[b]B. Sub. Npr + β hairpin + N243L + N216H + K305S.
[c]B. Sub. Npr + β hairpin + N243L.
[d]B. Sub. Npr + β hairpin.
[e]B. Sub. Npr + β hairpin + L315A.
[f]B. Sub. Npr + β hairpin + L315G.

CONCLUSIONS

The trajectory analysis procedure described here simplifies the analysis of simulated trajectories and assists the identification of important features. This is done by abstracting atomic motion to an easily conceptualised model of anharmonic transitions between wells in which atoms move approximately harmonically. The procedure is suitable for the analysis of equilibrium and nonequilibrium simulations. Both harmonic and anharmonic motion can be analyzed and quantified, and correlations between the motions of individual atoms during a trajectory can be detected.

The TRAJAN program is available on request to R.C.W. (e-mail: wade@embl-heidelberg.de).

ACKNOWLEDGMENTS

We thank Dr. T.P. Straatsma for provision of the ARGOS program, Dr. G. Vriend for bringing neutral proteases to our attention and providing us with an initial homology model of B. Sub Npr, and F. Nardi and Dr. V. Lounnas for critical reading of the manuscript.

REFERENCES

1 Horiuchi, T. and Go, N. Projection of Monte Carlo and molecular dynamics trajectories onto the normal mode axes: Human lysozyme. *Proteins* 1991, **10,** 106–116
2 Ichiye, T. and Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991, **11,** 205–217
3 Amadei, A., Linssen, A.B.M., and Berendsen, H.J.C. Essential dynamics of proteins. *Proteins* 1993, **17,** 412–425
4 van Aalten, D.M., Amadei, A., Linssen, B., Eijsink, V.G., Vriend, G., and Berendsen, H.J. The essential dynamics of thermolysin: Confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins* 1995, **22,** 45–54
5 Sessions, R.B., Dauber-Osguthorpe, P., and Osguthorpe, D.J. Filtering molecular dynamics trajectories to reveal

low-frequency collective motions: Phospholipase A2. *J. Mol. Biol.* 1989, **210,** 617–633
6 Osguthorpe, D.J. and Dauber-Osguthorpe, P. Extraction of the energetics of selected types of motion from molecular dynamics trajectories by filtering. *Biochemistry* 1990, **29,** 8223–8228
7 Dauber-Osguthorpe, P. and Osguthorpe, D.J. Partitioning the motion in molecular dynamics simulations into characteristic modes of motion. *J. Comput. Chem.* 1993, **14,** 1259–1271
8 Osguthorpe, D.J. and Dauber-Osguthorpe, P. FOCUS: A program for analyzing molecular dynamics simulations, featuring digital signal-processing techniques. *J. Mol. Graphics* 1992, **10,** 178–184
9 Gordon, H.L. and Somorjai, R.L. Fuzzy cluster analysis of molecular dynamics trajectories. *Proteins* 1992, **14,** 249–264
10 Karpen, M.E., Tobias, D.J., and Brooks, C.L. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: Analysis of 2.2-ns trajectories of YPGDV. *Biochemistry* 1993, **32,** 412–420
11 Swaminathan, S., Ravishankar, G., Beveridge, D.L., Lavery, R., Etchebest, C., and Sklenar, H. Conformational and helicoidal analysis of the molecular dynamics of proteins: "Curves," dials and windows for a 50 psec dynamic trajectory of BPTI. *Proteins* 1990, **8,** 179–193
12 Wiberg, K.B., Keith, T.A., Frisen, M.J., and Murcko, M. *J. Phys. Chem.* 1995, **99,** 9072
13 Eijsink, V.G.A., van der Burg, B., Vriend, G., Berendsen, H.J.C., and Venema, G. Thermostability of *Bacillus subtilis* neutral protease. *Biochem. Int.* 1991, **24,** 517–525
14 Vriend, G. and Eijsink, V. Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases. *J. Comput. Aided Mol. Design* 1993, **7,** 367–396
15 Eijsink, V.G.A., Vriend, G. van der Burg, B., van der Zee, J.R., Veltman, O.R., Stulp, B.K., and Venema, G. Introduction of a stabilizing 10 residue beta-hairpin in *Bacillus subtilis* neutral protease. *Protein Eng.* 1992, **5,** 157–163
16 Willis, B.T.M. and Pryor, A.W. *Thermal Vibrations in Crystallography.* Cambridge University Press, Cambridge, 1975
17 Straatsma, T.P. and McCammon, J.A. ARGOS, a vectorized general molecular dynamics program. *J. Comput. Chem.* 1990, **8,** 943–951
18 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 1983, **4,** 187–217
19 Madura, J.D., Briggs, J.M., Wade, R.C., Davis, M.E., Luty, B.A., Ilin, A., Antosiewicz, J., Gilson, M.K., Bagheri, B., Scott, L.R., and McCammon, J.A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 1995, **91,** 57–95
20 Beutler, T.C. and van Gunsteren, W.F. The computation of a potential of mean force: Choice of the biasing potential in the umbrella sampling technique. *J. Chem. Phys.* 1994, **100,** 1492–1497
21 Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion in a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 1977, **23,** 327–341
22 van Gunsteren, W.F. and Berendsen, H.J.C. *Groningen Molecular Simulation (GROMOS) Library Manual.* Biomos, Groningen, The Netherlands, 1987
23 Berendsen, H.J.C., Grigera, J.R., and Straatsma, T.P. The missing term in effective pair potentials. *J. Phys. Chem.* 1987, **91,** 6269–6271
24 Holmes, M.A., Tronrud, D.E., and Matthews, B.W. Structural analysis of the inhibition of thermolysin by an active-site-directed irreversible inhibitor. *Biochemistry* 1983, **22,** 236–240
25 *QUANTA release 3.2.* Molecular Simulations, Inc., Waltham, Massachusetts 1992
26 Shen J., Wong, C.F., Subramaniam, S., Albright, T.A., and McCammon, J.A. Partial electrostatic charges for the active center of Cu, Zn superoxide dismutase. *J. Comput. Chem.* 1990, **11,** 345–350
27 Eijsink, V.G.A., Vriend, G., van der Burg, B., Venema, G., and Stulp, B.K. Contribution of the C-terminal amino acid to the stability of *Bacillus subtilis* neutral protease. *Protein Eng.* 1990, **4,** 99–104
28 Eijsink, V.G.A., Vriend, G., van der Vinne, B., Hazes, B., van der Burg, B., and Venema, G. Effect of changing the interaction between domains on the thermostability of *Bacillus* neutral proteases. *Proteins* 1992, **14,** 165–170