Hidden peculiarities in the potential energy time series of a tripeptide highlighted by a recurrence plot analysis: A molecular dynamics simulation

Alessandro Giuliani
Institute for Research on Senescence Sigma Tau, Pomezia, Rome, Italy

Cesare Manetti
Department of Chemistry, University ''La Sapienza,'' Piazzale Aldo Moro, 5-00185 Rome, Italy
(Received 16 January 1996)

A molecular dynamics simulation performed with the GROMOS molecular dynamics program of the tripeptide L-tryptophan, *N*-[N-(5-oxo-L-prolyl)-L-leucyl]-methyl ester (Glp-Leu-Trp-OCH_3) was fully analyzed. The computation of dynamical nonlinear techniques to describe a potential energy time series, based on recurrence plot methodology, highlighted otherwise hidden features of the molecular dynamics in the equilibration phase. [S1063-651X(96)09906-0]

PACS number(s): 87.15.He, 02.70.Ns

The development of nonlinear dynamics has enlarged the range of possible means to describe time series with respect to the traditional statistical ones such as the mean and standard deviation or classical spectral techniques such as Fourier analysis [1–6]. These methods are all based on an embedding procedure performed on the time series; in the case of molecular dynamics (MD) simulations the potentially most relevant techniques are those not requiring the stationarity of the series [1,4,5].

Recurrence plot analysis (RPA) is a technique that was originally developed by Eckmann, Kamphorst, and Ruelle [1] as a purely graphical descriptive tool to analyze dynamic processes without any stationarity constraint. Recently, Webber and Zbilut demonstrated the high sensitivity of this technique in the detection of subtle state changes of the studied dynamics [4,7].

RPA is based on the computation of the distance matrix between the rows of the embedding matrix corresponding to the studied series. The distance matrix gives a general picture of the autocorrelation structure of the series and permits one to derive useful information about the dynamics. The local character of the distance function, upon which RPA is based, makes the technique particularly suited to detect fast transients in the dynamics.

In the present study, we use this technique in the analysis of the time series generated by MD. The automatic characterization of the MD experiments becomes crucial, especially when the simulation is performed on large molecules and/or for long simulation times: in all these cases, the amount of data makes a global ''classical'' analysis of data very difficult.

In order to check the usefulness of this technique in MD studies, use was made of a model both sufficiently simple to be able to interpret the obtained results and sufficiently rich to justify the use of a sophisticated data analysis tool.

For these reasons we chose as a model system a simple biologically active tripeptide (Glp-Leu-Trp-OCH_3) [8] [Fig. 1(a)]. This molecule probably does not undergo a continuous conformational rearrangement of the peptide backbone when interacting with biological systems and discrete structures can be detected.

The presence of an intramolecular hydrogen bond can be hypothesized as the main structural characteristic of this molecule that causes the stabilization of a γ turn. A turn is defined as a site where the peptide chain reverses its overall direction [Fig. 1(b)]. The term γ indicates a turn three residues long [9]. The hypothesized γ turn is stabilized by a hydrogen bond between the C═O residue of the Glp and the NH of the Trp [Figs. 1(a) and 1(b)].

Turn structures have been proposed often for other small peptides with a similar structure in solution. This particular conformation has been reported already in many cases of proline containing peptides such as Pro-Leu-Gly-NH_2 [10,11], Val-Pro-Gly-Gly [12], and Gly-Pro-Val-Gly [12] (Glp: 5-oxoproline; Leu: leucine; Trp: tryptophan; Val: valine; Pro: proline; Gly: glycine). We analyzed the final part of equilibration following a simulated annealing of the tripeptide.

Following the hypothesis of the intramolecular hydrogen bond, we performed a restrained MD simulation where the distance between the aromatic moiety of tryptophan and the ring of Glp was limited. We chose the potential energy value to describe our dynamical simulation. The advantage of using potential energy is twofold: from a technical point of view, working with a single variable permits one to considerably reduce the computational burden with respect to the classical use of the subsequent positions in time of all the atoms in the system. From an interpretation point of view, the use of a meaningful physical parameter, as potential energy is, provides a complementary view with respect to the positional information. Energy is both a global descriptor of the system and an easily interpretable ''physical observable.''

The analysis of average potential energy in the chosen range of equilibration highlights a slightly decreasing trend in the time series, not provoking substantial conformational changes, thus pointing to the existence of only one preferred conformation: as expected the hydrogen bond stabilized the conformation in an inverse γ turn.

Nevertheless, in the same time interval, the RPA parameters highlighted the existence of considerable oscillations

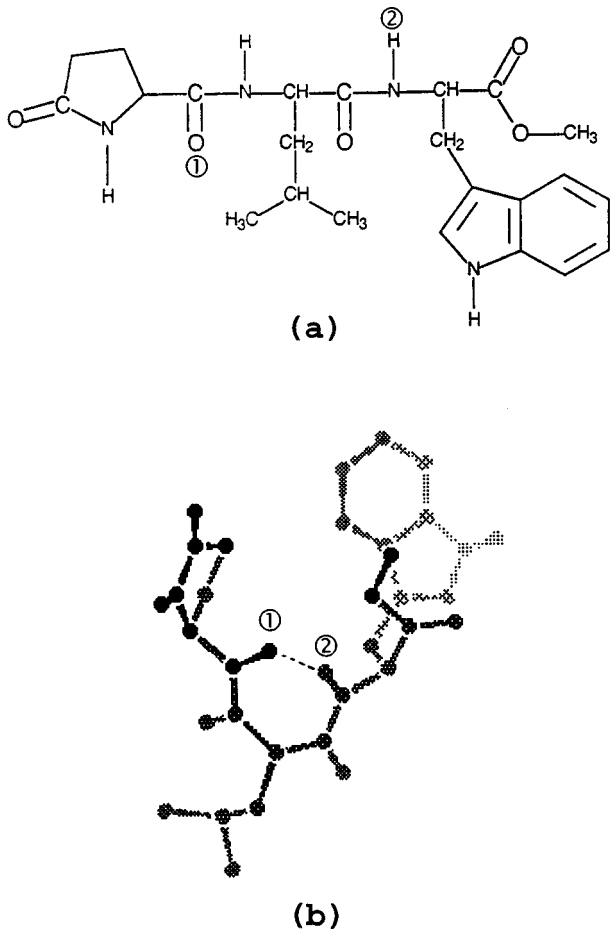

FIG. 1. (a) Schematic linear representation of the peptide Glp-Leu-Trp-OCH_3. The points 1 and 2 are the sites involved in hydrogen bonding. (b) Bent structure (γ turn) as obtained by restrained molecular dynamics simulation of the peptide. The broken line represents the hydrogen bond. Only the hydrogen atoms attached to nitrogen and oxygen are shown.

other than the slight drift of the energy. These oscillations are visible at a window size in which the average values of potential energy only register the drift and most likely correspond to the substates [13,14].

MOLECULAR DYNAMICS SIMULATION

Molecular dynamics simulations *in vacuo* were performed with the GROMOS molecular dynamics program [15]. The applied empirical potential energy function contains terms representing bond angle bending, harmonic dihedral angle bending (out-of-plane, out-of-tetrahedral configuration), sinusoidal dihedral angle torsion, van der Waals, and electrostatic interactions [16]. A dielectric constant, $\varepsilon=1$, was used and the cutoff radius for the nonbonded interactions was chosen in order to include all interactions. Between the protons attached to C-3 and C-4 of the ring of Glp and C-4, C-5, and C-6 of the Trp an attractive half-harmonic restraining potential was applied to force the molecule in a turn:

$$V_{DR}(d_{kl})=\begin{cases}\frac{1}{2}K_{DR}(d_{kl}-d_{kl}^{*})^2 & \text{if } d_{kl}>d_{kl}^{*}\\ 0 \quad \text{if } d_{kl}\leqslant d_{kl}^{*},\end{cases}$$

where K_{DR} is the force constant (10 kJ mol^{-1} Å^{-2}), d_{kl} is the distance between protons, and d_{kl}^{*} is the reference distance (5 Å). To hinder the *cis-trans* peptide isomerization at high temperature [17] a dihedral angle restraining potential was included in the force field:

$$V_{da}=(\phi_l,\delta_l)=K_{da}[1+\cos(n_l\phi_l-\delta_l)],$$

with K_{da}, the force constant, $=10$ kJ mol^{-1}.

For the evaluation of this potential all protons were treated explicitly. For all other terms only protons attached to nitrogen or oxygen atoms were treated explicitly. A bond stretching term was not included in the calculation; the SHAKE [18] algorithm was used to constrain bond lengths. The initial conformation of the MD simulation was a random one. All atoms were given an initial velocity obtained from a Maxwellian distribution at the desired initial temperature. The rescaling of the temperature during the run was obtained by a coupling with an external bath [19] with a time constant $\tau=0.1$ ps. A time step of 0.002 ps was used in the simulation.

The search for the structure that accounts for the restraints was performed by the simulated annealing procedure [20], based on a high-temperature run followed by a slow cooling. Several simulated annealings were performed. They differed by the presence of different sets of restraints and by a starting point that occurred at different times of high-temperature simulation. The common procedure was the following: a first 40-ps simulation at $T=800$ K was performed with no inclusion of the restraining potentials. It was followed by a 20-ps simulation at $T=800$ K including the restraining potential. The simulated annealing was performed in 50 ps with the final temperature at 300 K. Finally the system was equilibrated at 300 K for 190 ps. The last 30 ps were used for analysis.

DATA ANALYSIS

The potential energy series relative to the equilibration was submitted to a ten-dimensional embedding. The embedding space was constructed by the method of delays [21]: the space is generated by the construction of a multivariate matrix having as columns the original series shifted by a fixed lag consecutively applied to the series. The embedding dimension equals the number of columns of the matrix. For example, given the series 10, 11, 21, 32, 41, . . . the corresponding three-dimensional embedding space at a lag of 1 is

10	11	21
11	21	32
21	32	41
32	41	
41		

The rows of the embedding matrix correspond to subsequent epochs of length n (embedding dimension). In our case the lag was equal to 1 and correspondent to 0.05 ps.

The choice of ten-dimensional embedding was dictated by independent results obtained by a singular value decomposition (SVD) of the positions of atoms during the final part of

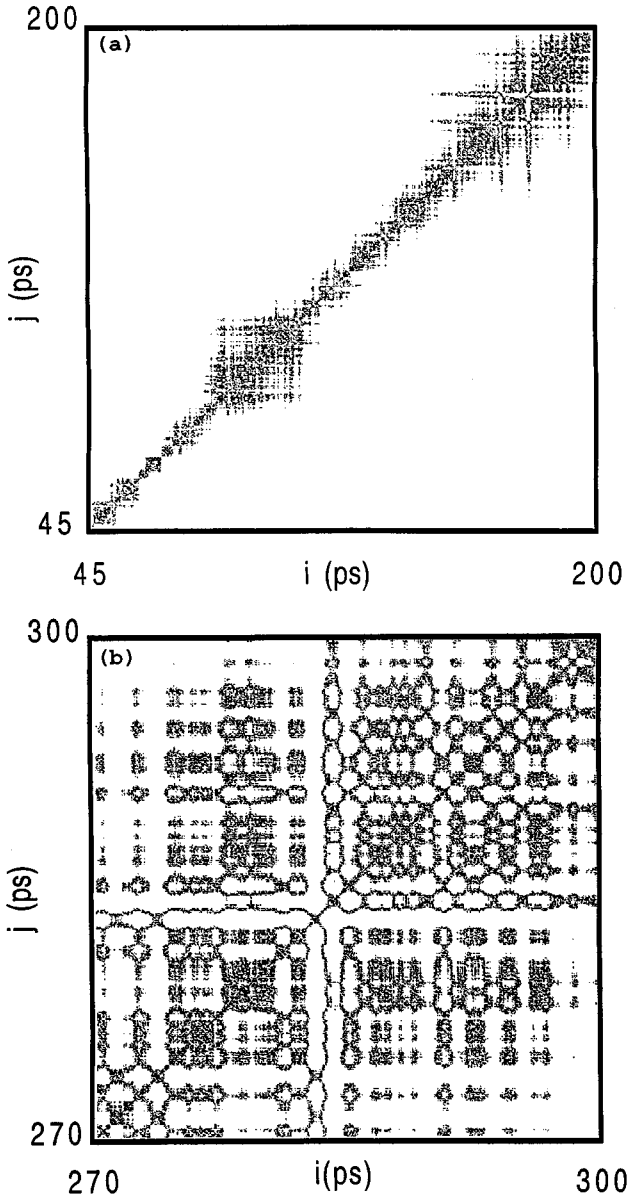

FIG. 2. (a) ''Cooling phase.'' A recurrence plot for potential energy (45–200 ps of simulation). A total of 3100 data points has been used. The embedding dimension was 10. The abscissa and the ordinate of the plot correspond, respectively, to the rows (i) and the columns (j) of the distance matrix based on the embedding matrix, therefore the main diagonal of the plot corresponds to the line of identity of the points. (b) ''Equilibration phase.'' A total of 600 points has been used (270–300 ps of simulation).

the simulation following the method of principal components analysis [21–26]; ten components explained 91% of total variability. In any case, the choice of the dimensionality of the embedding is consistent with the existence of four independent oscillators [21].

On the embedding matrix we applied the RPA. The RPA is based on the computation of a distance matrix between the rows (epochs) of the embedding matrix of a time series. This distance matrix is displayed by darkening the pixel located at specific (i,j) coordinates, which corresponds to a distance value between i and j rows (epochs) lower than a predetermined cutoff. The features of the distance function make the

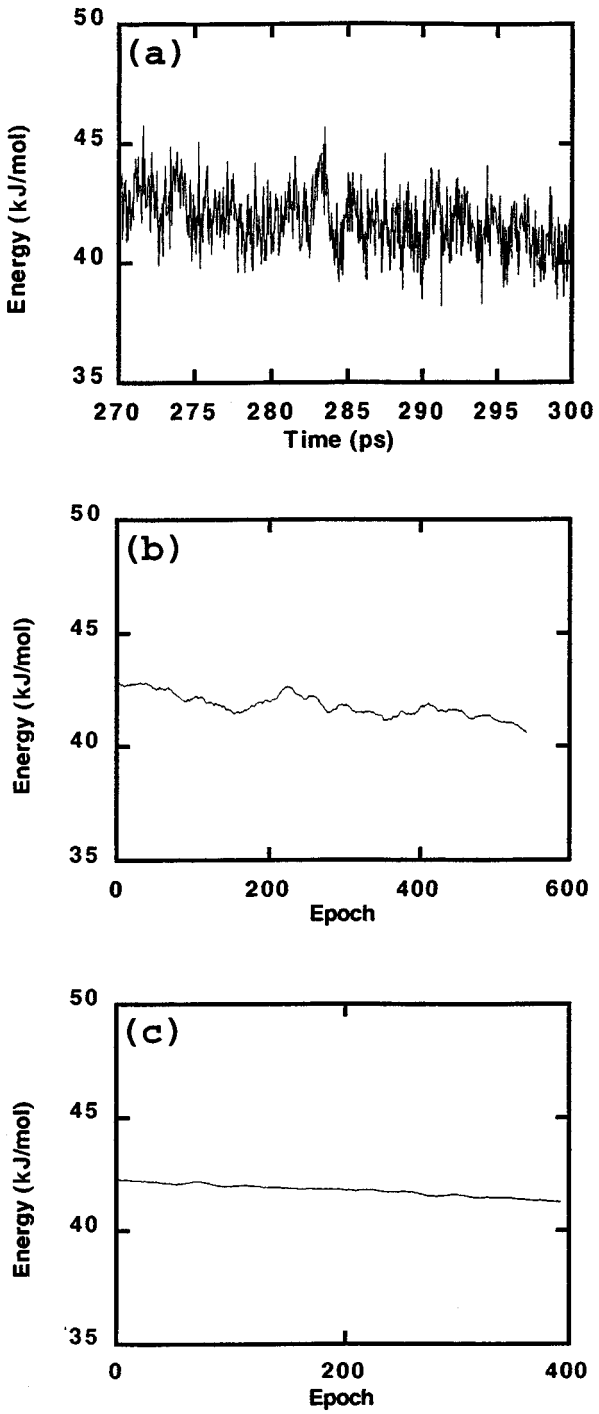

FIG. 3. (a) Graph of the potential energy measured in the range 270–300 ps, corresponding to the equilibration phase. (b) Moving average representation of the potential energy (50-point window) in the equilibration phase. The figures is consistent with (a) in terms of time. The horizontal axis reports the epochs corresponding to a 50-point window beginning at each selected time point. (c) Moving average representation of the potential energy (200-point window). The moving average procedure substitutes to the real value of potential energy at each time point a mean value relative to a predetermined window (epoch). This makes not directly appreciable the correspondence between the time points (horizontal axis) and the vertical axis. For this reason we prefer not to give an explicit dimension unit to the horizontal axis.

plot symmetric ($D_{i,j}=D_{j,i}$) and with a darkened main diagonal correspondent to the identity line ($D_{i,j}=0,\ i=j$). The darkened points individuate the recurrences (recurrent points) of the dynamical process and the plot can be considered as a global picture of the autocorrelation structure of the system under study [1,4].

Besides the global impression given by the graphic appearance of the plot, Webber and Zbilut [4] developed five quantitative techniques to describe the recurrence plot, which were demonstrated to be very useful in quantifying the evolution of the studied dynamical process.

The technique we took into account was percent recurrence (REC), which quantifies the percentage of the plot occupied by recurrent points. The changes of REC along the simulation were displayed by a continuous averaging along the time series. In order to study the evolution of the potential energy during the simulation we performed two separate analyses with two different moving average window sizes of 200 and 50 points, respectively. The consecutive windows were largely overlapped, the frame being shifted one point at a time; i.e., in the case of a 200-point frame, the first window started at point number 1 and ended at point number 200, the second window started at point number 2 and ended at point number 201 and so on; in the case of the 50-point window, the first frame went from 1 to 50, the second frame from 2 to 51, and so forth.

This use of moving average is common in dynamical studies and has some advantages with respect to the choice of contiguous windows. The main advantages are the possibility of direct comparisons of graphs relative to different choices of window size and the elimination of edge effects [3,27,28].

Averaging at different window sizes corresponds to filtering the series for different rhythms [28]. The smaller the window size, the more rugged the relative graph, for the obvious reason that each new window differs from the previous one for a major relative proportion of information (in our case 1/50 versus 1/200). This fact, in analogy to a renormalization group approach, makes the large window graphs more linked to the long range order parameters with respect to the small window ones, which are more dominated by fast noisy oscillations [29].

The global dynamic characteristics of the cooling phase and the final part of the equilibration phase are made evident by the recurrence plots [Figs. 2(a) and 2(b)]. The presence of a very strong main order parameter (temperature drift) in the cooling phase constrains all the recurrences along the main diagonal of the plot. This behavior corresponds to the existence of strong similarities in potential energy only between points very close along the time axis [Fig. 2(a)]. During the equilibration phase the drift is much weaker and the temperature is almost constant, thus allowing the presence of similarities in potential energy time series between points distant in time, generating the intermingled ''correlation texture'' of Fig. 2(b).

The time series relative to the equilibration phase was dynamically shuffled randomly changing the relative position of the values inside each row of the embedding matrix [4]. This procedure keeps the statistical characteristics of the time series invariant while altering the phase information exquisitely linked to the dynamics. We performed 30 reshuffings of the series; this procedure gave rise to dramatic changes of recurrence, which are reported in terms of mean $\pm$ standard deviation (unshuffled recurrence is 13.1%; shuffled recurrence is 24.3 ± 1.6). These results point to the existence of a meaningful dynamics even in the equilibration phase according to the Frauenfelder hypothesis [13,14].

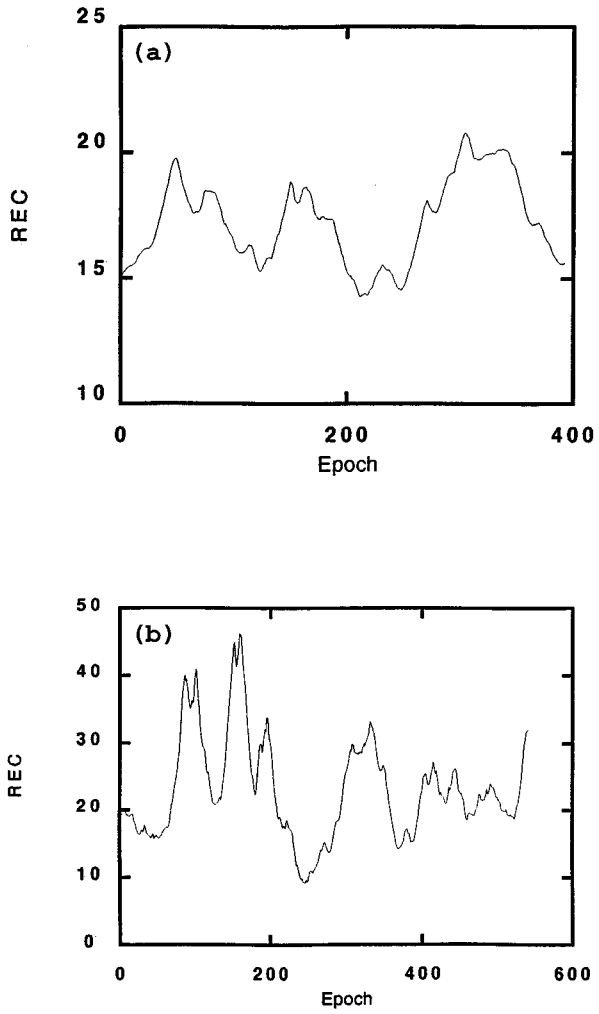

FIG. 4. (a) Percent recurrence (REC) relative to the large window size (200 points). (b) Percent recurrence (REC) relative to the small window size (50 points).

The average potential energy dynamics of the equilibration phase was monitored by means of the moving average approach described previously using as variables a statistical parameter (mean of the potential energy) and a dynamical technique (percent recurrence). The comparison between the raw potential energy data and the moving average ones for 50 and 200 windows, respectively, highlights the effect of window size [Figs. 3(a)–3(c)].

The moving average procedure makes the drift apparent even at the small window size. The large window, averaging out a bigger amount of rapid fluctuation than the small one, enhances the general trend of the data [29]. The SVD of the potential energy time series computed on the embedding space [21] highlighted a first nonoscillating component (drift), which explains 37% of the total variability, confirming the graphic impressions given by Fig. 3. The percent recurrence (REC) relative to the large window size [Fig.

4(a)], shows very large fluctuations in the considered part of the time series. In the same scale the mean value of the potential energy is approximately constant with only the very slight drift made evident [Fig. 3(c)].

The ''small window'' plot [Fig. 4(b)] is more rugged than the ''large window'' one; consequently the state transitions in correspondence with the REC fluctuations are related to ''faster'' effects. In natural settings, different from the artificial character of molecular dynamics, these fluctuations are more dominated by noise than the large window ones [29].

The possible structural implications of the REC oscillations are to be explored on a case-by-case basis. More generally it should be noted that the oscillations in REC point to state transitions in the dynamics that are potentially interesting [4,7].

In this work we show that RPA detects changes of states (rhythms) that are potentially informative in analyzing a MD simulation. The movements of the discrete parts of the molecule can be compared to the various musical instruments playing in an orchestra. In this view the conformation can be defined as a melody that is characterized by a general rhythm (first component of singular value decomposition) and substates to be seen as variations on this unique theme, recognized also for little changes in the instrument part (features detected by RPA).

We point out that only in cases of very small molecules is it possible to assign the variation in the conformations to a particular and unique element in the molecular structure, while, in general, variations are due to collective motions. This, however, is not relevant if the scope of inquiry is to detect dynamic variations of molecular structure simulation that are not associated with large fluctuations in potential energy, i.e., concerted movements for interaction, or to detect substates [14].

The use of potential energy for this approach is, in our opinion, a crucial point. Until now the potential energy was viewed, in MD studies, only as an indicator of the relative proximity to a minimum: only the ''drift'' of this parameter was taken into consideration. The potential energy drift was used to check the possibility to perform ''average'' measures as thermodynamical ones.

With the increase in sensitivity due to RPA analysis we can observe the ''shape'' of the potential energy trajectory in terms of temporal ''texture.'' This texture can be identified as the image of the oscillation of the system between different substates.

ACKNOWLEDGMENTS

Thanks are due to Professor J.P. Zbilut and Professor Alexander Tenenbaum for helpful discussion. We wish to thank Professor Hans Frauenfelder for his encouragement.

[1] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, Europhys. Lett. **4**, 973 (1987).
[2] T. Elbert, W. J. Ray, Z. J. Kowalik, J. E. Skinner, K. E. Graf, and N. Birbaumer, Physiol. Rev. **74**, 1 (1994).
[3] N. Pradhan and D. Narayana Dutt, Comput. Biol. Med. **23**, 281 (1993).
[4] C. L. Webber and J. P. Zbilut, J. Appl. Physiol. **76**, 965 (1994).
[5] L. Glass, M. R. Guevara, and A. Shrier, Ann. NY Acad. Sci. **584**, 168 (1987).
[6] A. L. Goldberger, D. R. Rigney, and B. J. West, Sci. Am. **262**, 42 (1990).
[7] C. L. Webber, M. A. Schmidt, and J. M. Walsh, J. Appl. Physiol. **78**, 814 (1995).
[8] M. S. Gilardini Montani, L. Tuosto, M. Delfini, D. Guerritore, G. Starace, V. Politi, and E. Piccolella, Immunopharma. **25**, 51 (1993).
[9] G. D. Rose, L. M. Gierasch, and J. A. Smith, Adv. Prot. Chem. **37**, 1 (1985).
[10] T. Higashijima, M. Tasumi, T. Miyazawa, and M. Miyoshi, Eur. J. Biochem. **89**, 543 (1978).
[11] R. Deslauriers, R. Walter, and I. C. P. Smith, FEBS Lett. **37**, 27 (1973).
[12] K. D. Kopple and A. Go, Biopolymers **15**, 1701 (1976).
[13] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[14] H. Frauenfelder and P. G. Wolynes, Phys. Today **47** (2), 58 (1994).
[15] W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual* (Biomos, Groningen, The Netherlands, 1987).
[16] J. Åqvist, W. F. van Gunsteren, M. Leijonmarck, and O. Tapia, J. Mol. Biol. **183**, 461 (1985).
[17] R. E. Bruccoleri and M. Karplus, Biopolymers **29**, 1847 (1990).
[18] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[19] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[20] S. Kirkpatrick, C. D. Jr. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[21] D. S. Broomhead and G. P. King, Physica D **20**, 217 (1986).
[22] A. E. García, Phys. Rev. Lett. **68**, 2696 (1992).
[23] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins **17**, 412 (1993).
[24] S. Hayward, A. Kitao, F. Hirata, and N. Go, J. Mol. Biol. **234**, 1207 (1993).
[25] M. H. Hao and H. A. Scheraga, Biopolymers **34**, 321 (1994).
[26] T. D. Romo, J. B. Clarage, D. C. Sorensen, and G. N. Phillips, Jr., Proteins **22**, 311 (1995).
[27] P. R. Killeen, Behav. Brain Sciences **17**, 105 (1994).
[28] G. Tiao and M. R. Grupe, Biometrika **67**, 365 (1980).
[29] D. A. Rand and H. B. Wilson, Proc. R. Soc. London B **259**, 111 (1995).