

Recurrence quantification analysis as a tool for characterization of molecular dynamics simulations

Cesare Manetti and Marc-Antoine Ceruso

Department of Chemistry, University of Rome "La Sapienza," Piazzale Aldo Moro, 5-00185 Rome, Italy

Alessandro Giuliani

TCE Laboratory, Istituto Superiore di Sanità, Rome 00161, Italy

Charles L. Webber, Jr.

Department of Physiology, Loyola University Medical Center, 2160 South First Avenue, Maywood, Illinois 60153

Joseph P. Zbilut

Department of Molecular Biophysics and Physiology, Rush University, 1653 West Congress, Chicago, Illinois 60612

(Received 5 June 1998)

A molecular dynamics simulation of a Lennard-Jones fluid and a trajectory of the B1 immunoglobulin G-binding domain of streptococcal protein G (B1-IgG) simulated in water are analyzed by recurrence quantification, which is noteworthy for its independence from stationarity constraints, as well as its ability to detect transients and both linear and nonlinear state changes. The results demonstrate the sensitivity of the technique for the discrimination of phase sensitive dynamics. The physical interpretation of the recurrence measures is also discussed. [S1063-651X(99)11201-7]

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

I. INTRODUCTION

The quantitative analysis of molecular dynamics (MD) trajectories implies the need for the individuation of salient phenomena embodied in time series data: Unique patterns in the dynamics require taxonomies. This need has engendered, as a consequence, the use of classical multivariate data analysis techniques such as principal components analysis (PCA) [1,2] and cluster analysis (CA) [3].

An important criterion for the choice of analysis method for MD trajectories is the method's dependence on dynamical components of the data set and its relative independence from purely statistical characteristics. This requirement is expressible in terms of "phase information" sensitivity and is broadly defined as having properties that are destroyed by random shuffling of the series itself (shuffling sensitive information). From this perspective, the usual statistical descriptors (e.g., mean and rms) do not carry any phase information (shuffling resistant), while both PCA and CA retain some information about the dynamics of the system (shuffling sensitive).

In a previous paper [4] we introduced recurrence quantification analysis (RQA) as an alternative analysis technique to obtain phase information about the energy landscape of simulated polypeptide systems. This technique, proposed by Eckmann, Kamphorst, and Ruelle as a purely graphical tool and made quantitative by Zbilut and Webber [5-7], in contradistinction to PCA, gives a local view of the studied series, based on single distance pairs and not on the distribution of distances. This means that while autocorrelation functions can only show general trends, RQA is particularly suited for the detection of fast transients [8] and the consequent localization in time of the salient features of the dynamics such as changes of state. Furthermore, RQA's inde-

pendence from stationarity constraints, its ability to detect both linear and nonlinear dynamics [7,8,1], and its ability to discriminate between signals and random noise [9] make RQA well suited for a detailed characterization of MD trajectories (e.g., identification of microstates).

In this work, RQA was applied to the potential energy time series of conformational space explored during MD simulations. The main goal was to demonstrate the ability of RQA to discriminate the dynamics of a simple system [a Lennard-Jones (LJ) fluid, which does not carry any phase information, since it is a purely statistical, shuffling-resistant system] from a complex system, such as the MD trajectory of B1-IgG simulated in water (a protein that demonstrates shuffling-sensitive phase information, due to the existence of structured paths between its microstates) [10].

The testing strategy is straightforward: While in the case of a LJ fluid the RQA measures must remain invariant after shuffling, they should change significantly in the case of the protein. We will try to sketch a physical characterization of RQA measures relative to MD.

II. MATERIALS AND METHODS

A. LJ system MD simulations

For the LJ simulation, we considered a system of 125 particles, enclosed in a cube of side L , with periodic boundary conditions interacting through a two-body potential of the LJ type:

$$V(r) = 4[(\sigma/r)^{12} - (\sigma/r)^6], \quad (1)$$

with the parameter $\sigma = 3.405 \text{ \AA}$ corresponding to argon and the energies expressed in units of ϵ ($\epsilon = 119.8 \text{ K}$) [11]. The simulations were performed at different temperatures and

varying L at different density. We used the same protocol for both simulations: As a first step, the initial velocities were taken from a Maxwellian distribution to perform 200 ps of simulation with only the last 100 ps being used for analysis. The potential energy time series was sampled at 0.05 ps.

B. Protein MD simulation

All simulations of the protein were performed with the GROMACS simulation package [12]. A modification [13] of the GROMOS87 [14] force field was used with additional terms for aromatic hydrogens [15] and improved carbon-oxygen interaction parameters [13]. SHAKE [16] was used to constrain bond lengths, allowing a time step of 2 fs.

The initial protein configuration was taken from the protein databank (1pga) [17]. The protein was immersed in a pre-equilibrated box of simple point charge (SPC) water [18], while 4 water molecules with the highest electrostatic potential were replaced by sodium ions, resulting in an electrically neutral cubic box ($a \approx 4.1$ nm) containing 1790 water molecules and 4 sodium counter ions for a total of 5936 atoms. Care was taken that all crystallographic water molecules be conserved.

In order to prepare the solvated system for molecular dynamics, a three-step procedure was followed. Using a restraining harmonic potential, all heavy atoms of the protein and the crystallographic water oxygens were constrained to their initial positions while surrounding SPC water molecules were first minimized and then submitted to 5 ps of constant volume MD at 300 K. The resulting system was then minimized, without any constraints, before starting constant temperature and constant volume MD. A nonbonded cutoff of 1.2 nm was used for both LJ and Coulomb potentials. The pair lists were updated every ten steps. A constant temperature of 300 K was maintained by coupling to an external bath [19] using a coupling constant ($\tau = 0.002$) equal to the integration time step (2 fs). A total of 1.9 ns of simulation were produced in this manner. The potential energy of the protein was sampled every 0.1 ps.

III. RECURRENCE QUANTIFICATION ANALYSIS

RQA was introduced to the physical sciences by Eckmann, Kamphorst, and Ruelle in 1987 [5] as a purely graphical technique. Five years later Zbilut and Webber [6] enhanced the technique by defining five nonlinear descriptors of the recurrence plot that were found to be diagnostically useful in the quantitative assessment of time series structure in fields ranging from molecular dynamics to physiology [4,20,21]. This technique has been demonstrated to be particularly useful in quantifying transient behavior far from equilibrium in relatively short time series [8]. This feature is particularly important in the detection of unique patterns in complex data sets [22].

RQA is based on the computation of a Euclidean distance matrix between the rows (epoch) of an embedded matrix of the scalar time series (in this case the MD simulations of potential energy) at a fixed lag (time delay method), as originally suggested by Ruelle and recorded by Packard *et al.* [23]. Takens's [24] theorem states a mathematical relation between the embedding dimension of a suitable scalar observable and the real dimension of the attractor of the corresponding dynamical system (d): $n = 2d + 1$, where n is the

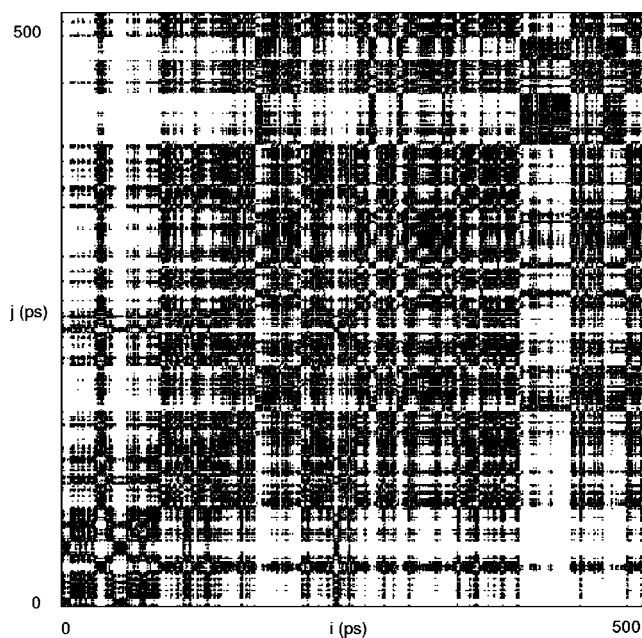


FIG. 1. Protein recurrence plot.

minimum embedding dimension to ensure the reconstructibility of the underlying attractor. This relation holds true for deterministic, noiseless, systems. In the presence of noise from whatever source (e.g., system, quantization), however, it has been shown that higher embeddings are necessary [25]. From this it is obvious that the embedding be sufficiently high so as to capture all the essential dynamics: The error of choosing a too high an n is not as great as choosing too low a value of n . This is particularly true in MD simulations, where it is crucial to have sufficient dimensionality to estimate higher-order moments giving information on correlations between the movements of two or more atoms in a protein [26]. In the present case, 10 was dictated by our previous work on a tetrapeptide dynamics [4].

Thus the sequence of vectors (\mathbf{x}_i), embedded in \mathbb{R}^n , define a function on an $n \times n$ array according to the rule: Darken the (i, j) th element of the array if $\mathbf{x}_j \in B(\mathbf{x}_i, r)$, where $B(\mathbf{x}_i, r)$ is the ball or radius r , centered at \mathbf{x}_i (see [4,7] for details). The features of the distance function make the plot symmetric $D_{i,j} = D_{j,i}$ with a darkened main diagonal corresponding to the identity line ($D_{i,j} = 0, j = i$). 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. Consequently, a recurrence plot visualizes the distance matrix, which in turn represents the autocorrelation present in the series at all possible time scales. In fact, it is important to note that the distance is computed for all the possible pairs of epochs, the elements near the principal diagonal of the plot corresponding to short-range correlations (the diagonal marks the identity in time), and the long-range correlations corresponding to points distant from the main diagonal. Besides the global impression given by the graphic appearance of the plot (see Figs. 1 and 2 for the RQA plot of the protein and the fluid), the measures developed by Webber and Zbilut [7,8] allow for a quantitative description of the recurrence structure of the plot. This is an important consideration since visual inspection of plots can at times lend themselves to misinterpre-

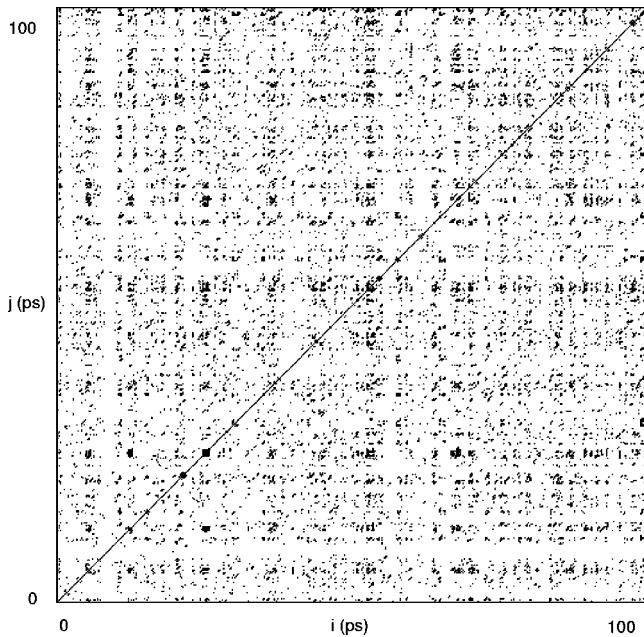


FIG. 2. LJ fluid recurrence plot.

tation due to the vagaries of human perception. Additionally, quantification then allows for hypothesis testing.

The RQA descriptors are as follows. REC denotes recurrence, which quantifies the percentage of the plot occupied by recurrent points. It corresponds to the proportion of recurrent pairs over all the possible pairs of epochs or, equivalently, the proportion of pairwise distances below the chosen radius among all the computed distances. DET denotes determinism and is the percentage of recurrent points that appear in sequence, forming diagonal line structures in the distance matrix. DET corresponds to the amount of patches of recurrent behavior in the studied series, i.e., to portions of the state space in which the system resides for a time longer than expected by chance alone (see [27,28]). This is a crucial point: A recurrence can, in principle, be observed by chance whenever the system explores two nearby points of its state space. On the contrary, the observation of recurrent points consecutive in time (and then forming lines parallel to the main diagonal) is an important signature of deterministic structuring [5,29]. The superposition between determinism

and Lyapunov exponents is a proof of this point [5]. ENT denotes entropy, which is defined in terms of the Shannon-Weaver formula for information entropy [7,30] computed over the distribution of length of the lines of recurrent points and measures the richness of deterministic structuring of the series. LYAP is simply the length (in terms of consecutive points) of the longest recurrent line in the plot. LYAP was found to accurately predict ($r=0.93$) the value of the maximum Lyapunov exponent in a logistic map going from a regular to chaotic regime [8]. Finally, TREND is the regression coefficient of the relation between time (in terms of distance from the main diagonal) and the amount of recurrence. TREND quantifies the fading away of recurrence going forward in time and represents a measure of stationarity [7]. Additionally, a time series of any of the RQA descriptors can be produced by windowing the originally embedded scalar series and overlapping in a manner similar to time-varying spectral plots (see, e.g., [8] and Figs. 3, 4, and 7).

IV. RESULTS

All RQA descriptors were computed for both the LJ fluid and the protein MD potential energy time series. In particular, for the fluid, data from a simulation at $T=0.8$ (usual reduced unit) were used. To test for the null hypothesis that the MD series are stochastic, the original trajectories were randomly shuffled to obtain 30 copies of each series (Table I). The 95% confidence intervals for the RQA descriptors were computed and the position of the original series relative to the confidence intervals checked. Except for REC (and here it is noted that the value for the MD simulation falls within the range of obtained shufflings), the null hypothesis for LJ fluid could not be rejected, pointing to the stochastic character of the fluid simulation. For the protein MD, however, the RQA values were well beyond the confidence limits of the shuffled series, thus demonstrating the presence of strong “phase information” for the protein dynamics. These features are qualitatively evident when looking at the recurrence plots of the protein (Fig. 1) and the LJ fluid (Fig. 2): While protein shows a very rich and intermingled texture, the LJ plot is much more homogeneous.

V. DISCUSSION

A relevant portion of the theoretical work on MD was based upon LJ fluid simulations, performed by Rahman [31].

TABLE I. RQA results for shuffling.

	Shuffled mean	95% confidence interval	Range	MD simulation
LJ Fluid				
REC	0.76	0.74–0.78	0.66–0.87	0.688
DET	39.47	38.85–40.08	36.57–42.63	39.03
ENT	2.33	2.31–2.35	2.25–2.50	2.347
LYAP	17	16–18	14–21	16
TREND	0.009	–0.03–0.05	–0.23–0.24	–0.005
Protein				
REC	0.72	0.69–0.75	0.58–0.86	5.12
DET	39.20	38.08–40.31	33.20–44.88	69.48
ENT	2.31	2.27–2.36	2.06–2.62	3.25
LYAP	14	13–15	10–21	31
TREND	0.033	–0.08–0.15	–0.45–0.81	–2.27

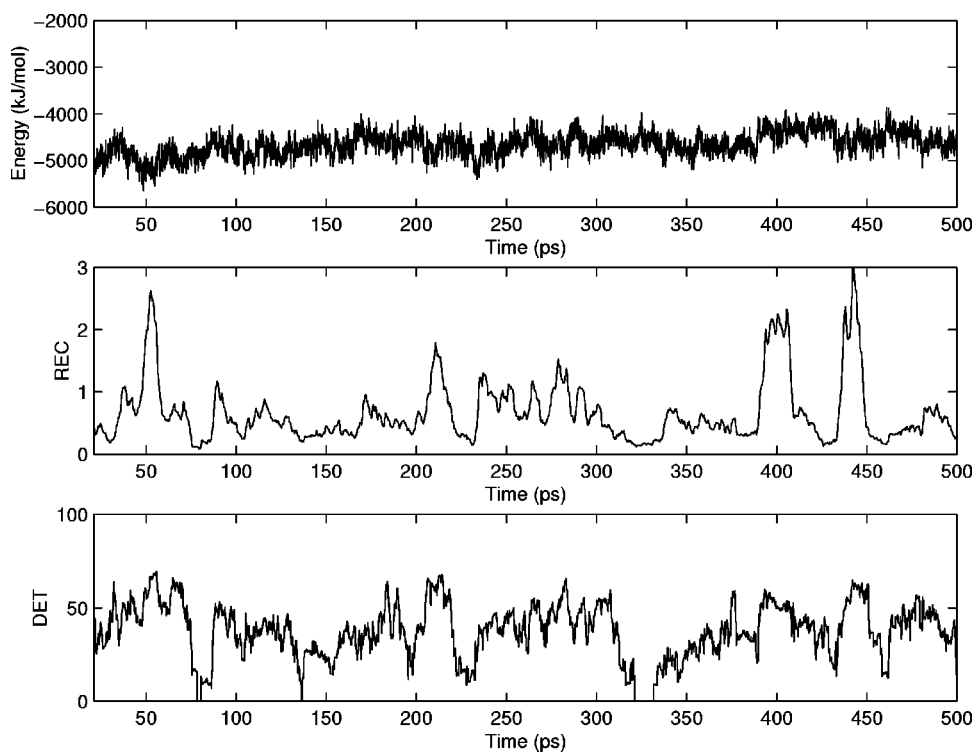


FIG. 3. Time series of protein MD simulation (top), with the respective REC and DET plots, calculated on a 200 point windowed series overlapping one point at a time.

These trajectories can be defined as recurrent, Hamiltonian, mixing, and K flow, or following some authors, “Lyapunov unstable” [32]. The simulated LJ system evolves toward an equilibrium state and the constant energy surface defined by the initial conditions is accessible to the system itself. The motion of such a system is at least mixing so as to sample all the explored surface.

In the case of the LJ fluid, the result obtained, in Eckmann, Kamphorst, and Ruelle’s terms [5], can be defined as “autonomous,” i.e., typical of a system evolving following

time-independent equations: This corresponds to our operational definition of a “shuffling resistant” potential energy time series. In fact, the RQA measures of the shuffled series are not statistically different from the original series. This behavior corresponds to a randomlike sampling of the phase space of the system, even if the sampling is driven by a deterministic “engine” such as MD. These kinds of “experiments” were used by Verlet and co-workers [11,33] to compute thermodynamic properties of fluids following a formalization introduced by Birkhoff [34,35] and based on

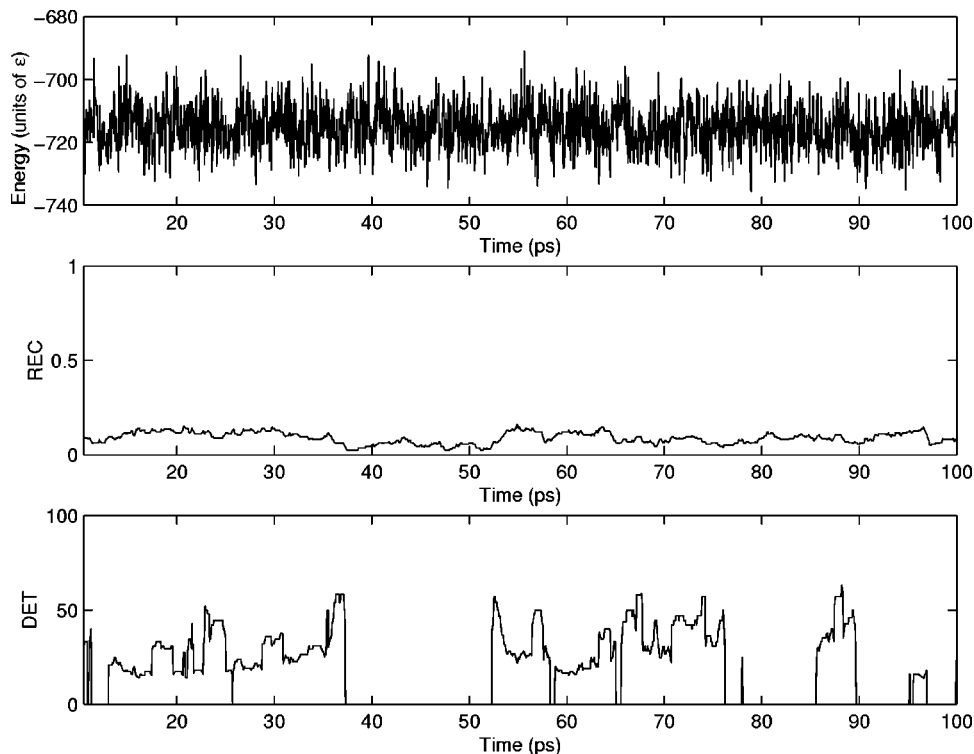


FIG. 4. Time series of LJ fluid simulation (top), with the respective REC and DET plots, calculated as in Fig. 3.

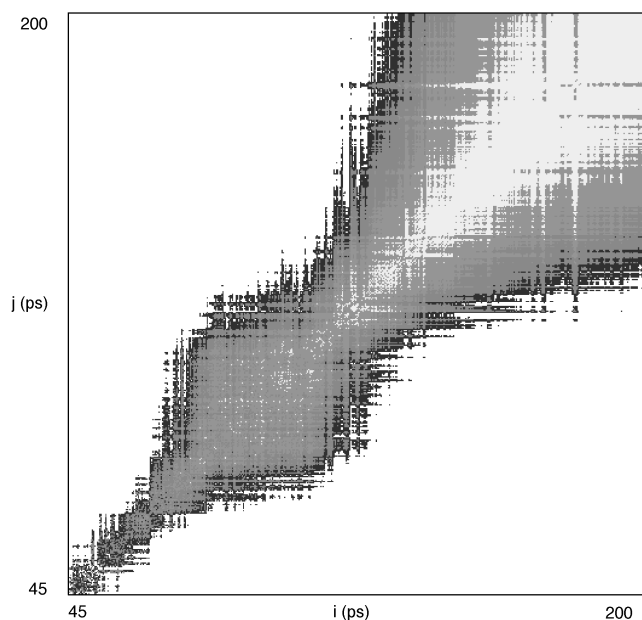


FIG. 5. Simulated annealing tripeptide recurrence plot.

Boltzmann's view of ergodicity assumptions.

The recurrence plot of the protein simulation (Fig. 1) immediately shows the impossibility of direct averaging of the data. The simple visual inspection of the plot highlights abrupt changes in the texture pointing to multiple minima in the trajectory (rugged landscape as opposed to flat surface). This point is underscored in Fig. 3, which depicts REC and DET in a windowed series. REC presents as a rugged landscape, while DET displays several discontinuities. More importantly, the shuffling procedure significantly alters the numerical values of the RQA descriptors, thus demonstrating the “shuffling sensitivity” of the underlying trajectory. The

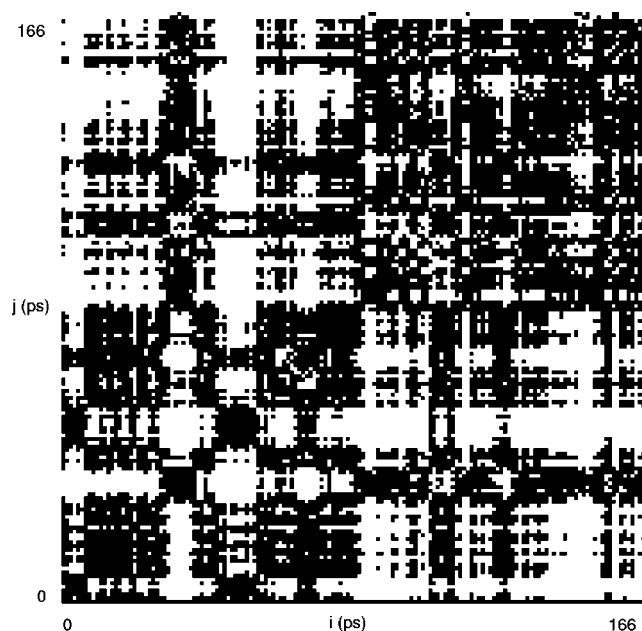


FIG. 6. Detail of Fig. 1 protein with two sections (squares, upper right and lower left) clearly separated, representing two different microstates.

“ergodic” constraints of complete accessibility and mixing are not sufficient to make the system evolve to an equilibrium situation given the finite time of the simulation and as a result the trajectory is trapped in a limited portion of the energy surface. In such situations we can speak of metastable states [36,37], which obviate the possibility of computing direct averages. In order to compute physical measures on such simulations, the local minima of the phase space must be revealed and their relative depth estimated.

The thermalization algorithms used in MD are not guaranteed to preserve the microcanonical properties of the system

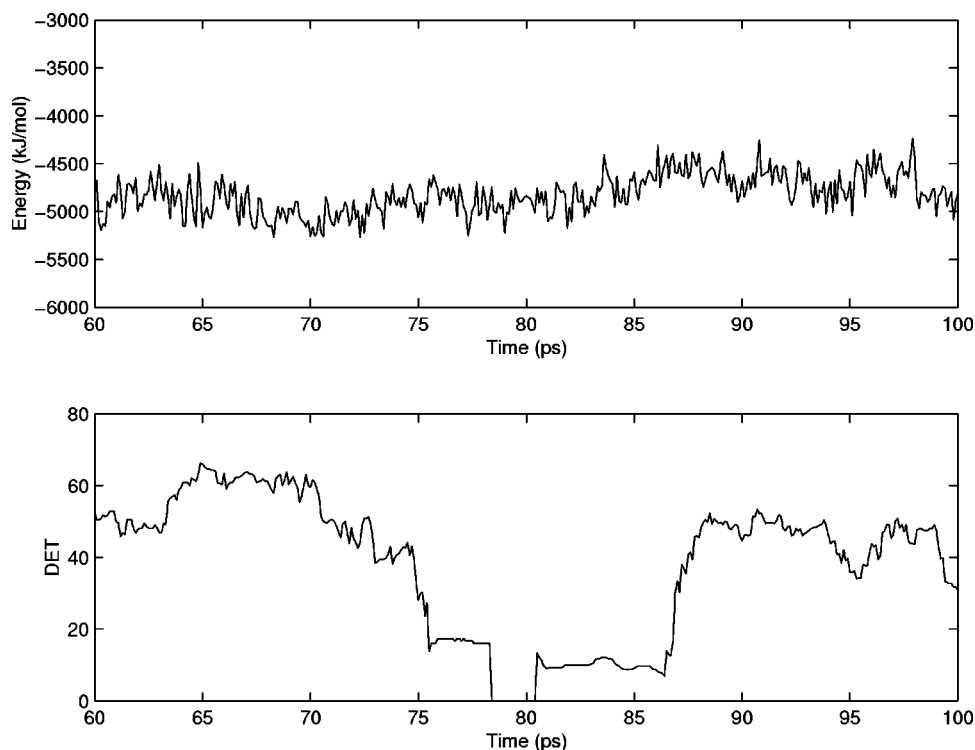


FIG. 7. Section of protein MD simulation between microstates of Fig. 6 (top), with the respective DET plot. Note the clear divergence between the two microstates at 78–81 ps.

[19]; nevertheless, we think that the quantitative RQA measures can be correlated to the thermodynamic properties of the system under investigation. In any case, these measures allow us to derive some useful information about the shape of the energy landscape of the simulation. As a matter of fact, the basic algorithm of recurrence plots was developed by Eckmann *et al.* [38] with the aims of reconstructing the dynamics relative to a time series in a finite-dimensional space and of generating a tangent map of the reconstructed dynamics in order to calculate Lyapunov exponents.

In the recurrence plot, a recurrence is scored (and the respective point darkened) whenever $|x_j - x_i| < d$. We can think of this inequality as the numerator of the incremental ratio, dy/dx , where dx corresponds to the time interval between two sampled points in phase space. Thus the recurrent points are corresponding segments of the trajectory going through valleys of the multidimensional space on which the potential energy is projected by the embedding procedure. These valleys have, by definition, a low slope (d) (and thus they are recurrent in phase space). On a more general note, the global texture of the plot is linked to the ruggedness of the explored landscape. A dense texture is linked to smooth slopes and a flat landscape, while a coarse texture points to steeper energetic barriers with the lack of texture (no recurrent points at all) revealing transitions. This qualitative picture is consistent with the analysis of the logistic map by RQA [8] where the phase transitions of the system (changes in dynamical regime) were registered by the RQA measures.

Looking at the recurrence plot of the LJ fluid (Fig. 2), the loss of any preferential directionality of the system (quantitatively proved by the shuffling invariance) is clear. REC (Fig. 4) is considerably smoother than Fig. 3, while DET is erratic, reflective of the very low recurrence values. The directionality of Fig. 2 is highlighted when it is compared to the plot relative to the annealing phase of a tripeptide (Fig. 5) (see [4]) where the directionality was imposed by a strong order parameter, i.e., decrease of temperature. The protein recurrence plot (Fig. 1) has a preferential directionality in time (shuffling sensitivity) that allows us to appreciate the effective dimensionality of the explored conformational space. With an adequate sampling time we can resolve both the Frauenfelder substrates [10,26,39] in terms of large-scale typology of the plot (Figs. 6 and 7) and the features of the single substrate in terms of texture (Fig. 8).

It is important to note that, in practice, several different variables could have been chosen for the MD simulations, such as Coulomb energy or van der Waals energy. Potential energy was chosen since it is a global descriptor of the physical motions that can be adequately sampled [26] relative to long relaxation times as compared to anharmonic motions

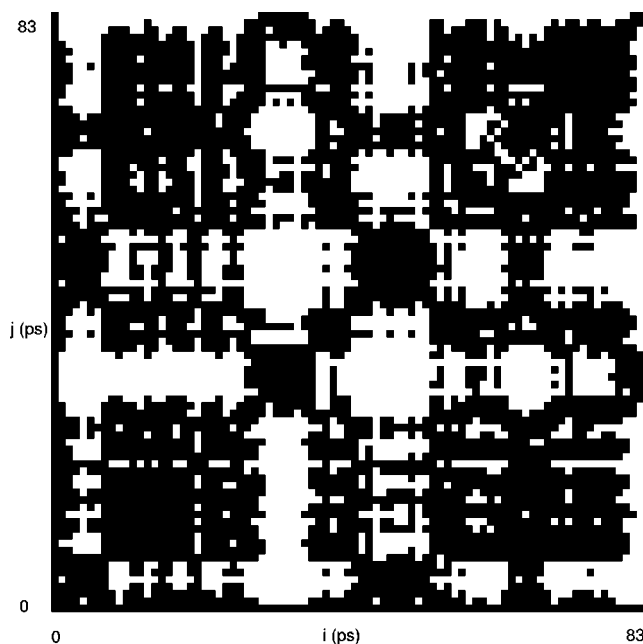


FIG. 8. Detail of lower left microstate of Fig. 6.

[40]. Thus potential energy is suited for the studied time scale. In a related matter, while the generalized ergodic measure (GEM) of Straub and Thirumalai [41] can compute a distribution of energy barriers between substrates, currently this is not possible with RQA. The advantage of RQA over GEM is its sensitivity to local fast transients, thus being much more suited to dynamically individuate phase transitions. In this sense, the two methods are complementary: The GEM is better for energy characterization and RQA is better for time resolution.

Finally, it should be remarked that RQA allows for the identification of putatively important events along the studied dynamics using only one variable (potential energy) instead of many (e.g., single dihedral angles). The meaning of these events may in fact require specific analyses using variables such as vibrational times, ring flips of amino acids, folding times, protonation times, or diffusion times of water and require a case by case approach. Our analysis is particularly useful in the calculation of free energy difference between reactant and product in which various MD simulations can reduce uncertainties in free energy calculations [42], although, as with MD studies in general, it is still limited by the general time scale of physiologically important substrates that are on the order of tens of microseconds to milliseconds. In summary, RQA seems to constitute a very promising tool for the characterization of conformational substrates in MD simulations.

[1] A. E. Garcia, Phys. Rev. Lett. **68**, 2696 (1992).
 [2] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins: Struct., Funct., Genet. **17**, 412 (1993).
 [3] M. E. Karpen, D. J. Tobias, and C. L. Brooks III, Biochemistry **32**, 412 (1993).
 [4] A. Giuliani and C. Manetti, Phys. Rev. E **53**, 6336 (1996).

[5] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, Europhys. Lett. **4**, 973 (1987).
 [6] J. P. Zbilut and C. L. Webber, Jr., Phys. Lett. A **171**, 199 (1992).
 [7] C. L. Webber, Jr. and J. P. Zbilut, J. Appl. Physiol. **76**, 965 (1994).

- [8] L. Lligona-Trulla, A. Giuliani, J. P. Zbilut, and C. L. Webber, Jr., *Phys. Lett. A* **223**, 225 (1996).
- [9] J. P. Zbilut, A. Giuliani, and C. L. Webber, Jr., *Phys. Lett. A* **237**, 131 (1998).
- [10] G. U. Nienhaus, J. D. Müller, B. H. McMahon, and H. Frauenfelder, *Physica D* **107**, 297 (1997).
- [11] L. Verlet, *Phys. Rev.* **159**, 98 (1967).
- [12] D. van der Spoel, H. J. C. Berendsen, A. R. van Buuren, E. Apol, P. J. Meulenhoff, A. L. T. M. Sijbers, and R. van Drunen, *Gromacs User Manual* (Nijenborgh 4, 9747 AG Groningen, 1995). Internet:<http://rugmd0.chem.rug.nl/gmx>.
- [13] A. R. van Buuren, S. J. Marrink, and H. J. C. Berendsen, *J. Phys. Chem.* **97**, 9206 (1993).
- [14] W. F. van Gunsteren and H. J. C. Berendsen, *Gromos Manual (BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry)* (University of Groningen, Groningen, 1987).
- [15] W. F. van Gunsteren, S. Billeter, A. Eising, P. Hünenberger, P. Krüger, A. Mark, W. Scott, and I. Tironi, *Biomolecular Simulation: The Gromos96 Manual and User Guide* (Biomos bv, Zürich, 1996).
- [16] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- [17] P. T. Gallagher, P. B. Alexander, and G. L. Gilliland, *Biochemistry* **33**, 4721 (1994).
- [18] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331–342.
- [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] D. Mestivier, N. P. Chau, X. Chanudet, P. Baudeceau, and P. Larroque, *Am. J. Physiol.* **41**, H1094 (1997).
- [21] P. Faure and H. Korn, *Proc. Natl. Acad. Sci. USA* **94**, 6505 (1997).
- [22] J. P. Zbilut, A. Giuliani, C. L. Webber, Jr., and A. Colosimo, *Protein Eng.* **11**, 87 (1998).
- [23] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
- [24] F. Takens, *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young (Springer-Verlag, Berlin, 1980), pp. 366–381.
- [25] M. Ding, C. Grebogi, E. Ott, T. Sauer, and J. A. Yorke, *Physica D* **69**, 404 (1993).
- [26] J. B. Clarage, T. Romo, B. K. Andrews, B. M. Pettitt, and G. N. Phillips Jr., *Proc. Natl. Acad. Sci. USA* **92**, 3288 (1995).
- [27] M. Zak, J. P. Zbilut, and R. E. Meyers, *From Instability to Intelligence*, Lecture Notes in Physics Vol. M49 (Springer, Heidelberg, 1997).
- [28] J. P. Zbilut, M. Zak, and R. E. Meyers, *Biol. Cybern.* **75**, 277 (1996).
- [29] J. P. Zbilut, A. Giuliani, and C. L. Webber, Jr., *Phys. Lett. A* **237**, 131 (1998).
- [30] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [31] A. Rahman, *Phys. Rev.* **136**, A405 (1964).
- [32] J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods* (Wiley, New York, 1992), p. 53.
- [33] J. L. Lebowitz, J. K. Percus, and L. Verlet, *Phys. Rev.* **153**, 250 (1967).
- [34] G. D. Birkhoff, *Proc. Natl. Acad. Sci. USA* **17**, 656 (1931).
- [35] J. v. Neumann, *Proc. Natl. Acad. Sci. USA* **18**, 263 (1932).
- [36] J. Ford, *Adv. Chem. Phys.* **24**, 155 (1973).
- [37] J. D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **87**, 3526 (1990).
- [38] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, *Phys. Rev. A* **34**, 4971 (1986).
- [39] A. Ansari, J. Berendzen, D. Braunstein, B. R. Cowen, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young, *Proc. Natl. Acad. Sci. USA* **82**, 5000 (1985).
- [40] P. J. Steinbach and B. R. Brooks, *Proc. Natl. Acad. Sci. USA* **93**, 55 (1996).
- [41] J. E. Straub and D. Thirumalai, *Proteins: Struct., Funct., Genet.* **15**, 360 (1993).
- [42] A. Hodel, T. Simonson, and R. O. Fox, *J. Phys. Chem.* **97**, 3409 (1993).