

Identifying and correcting non-Markov states in peptide conformational dynamics

Dmitry Nerukh,^{a)} Christian H. Jensen, and Robert C. Glen

Department of Chemistry, Unilever Centre for Molecular Sciences Informatics, Cambridge University, CB2 1EW Cambridge, United Kingdom

(Received 3 November 2009; accepted 2 February 2010; published online 23 February 2010)

Conformational transitions in proteins define their biological activity and can be investigated in detail using the Markov state model. The fundamental assumption on the transitions between the states, their Markov property, is critical in this framework. We test this assumption by analyzing the transitions obtained directly from the dynamics of a molecular dynamics simulated peptide valine-proline-alanine-leucine and states defined phenomenologically using clustering in dihedral space. We find that the transitions are Markovian at the time scale of ≈ 50 ps and longer. However, at the time scale of 30–40 ps the dynamics loses its Markov property. Our methodology reveals the mechanism that leads to non-Markov behavior. It also provides a way of regrouping the conformations into new states that now possess the required Markov property of their dynamics.

© 2010 American Institute of Physics. [doi:10.1063/1.3328781]

I. INTRODUCTION

Conformational dynamics of biomolecules is defined by the free energy surface of the molecular system, which, in turn, defines experimentally measurable molecular structures and their kinetics. Thus, an effective approach to the analysis of the conformational dynamics can provide an insight into the mechanisms of molecular processes. One such approach, the Markov state model (MSM), has recently been a topic of very active development.^{1–7} The theory statistically describes the conformational states and the transitions between them. It is intuitive and mathematically well developed.

One of the main difficulties in using MSM is an optimal choice of molecular coordinates and the definition of discrete states in these coordinates. For the theory to be applicable, the states have to possess a Markov property, that is, the probability of the appearance of the next state should depend on the current state only. At the same time, the states should be physically meaningful or at least should allow the construction of physical molecular conformations. For this reason, the states are often defined by clustering geometrically similar biomolecular structures. To ensure the Markov property of the states, either the time step is chosen “long enough” for the system to forget the history or the states are constructed from a large number of initial “microstates.”⁸ The Markov property is fundamental in the framework of protein conformational analysis where developing “a reliable and unsupervised test for Markovianity” is listed among the major open problems in Ref. 9.

The geometric definition of the states is not optimal, both with respect to Markov property and for the resulting parameters of MSM. For example, we have found^{10,11} that the rates of the transitions between the conformations of a peptide are extremely sensitive to geometrically defined state

boundaries, especially if the time scale is below the time threshold for Markovian behavior. The errors are amplified in the case of many states, which is often the case for realistic systems.

The time step has to be optimal as well. Too long time intervals between the analyzed conformations risk to miss important short lived intermediate states. They are also computationally inefficient; not only longer simulations are needed, but also the statistic on rarely occurring states accumulates very slowly. Thus, the time step has to be as short as possible but long enough to ensure the Markov property of the states.

A more rigorous approach to the choice of time step is based on the analysis of transition probabilities. Since for a Markov process all transitions are independent, the product of probabilities of n consecutive transitions should be equal to the probability of a single transition from the first to the n th time step. When the whole transition matrix for a set of states is considered, this corresponds to the relation $T_{n\Delta t} = T_{\Delta t}^n$, where T_t is a matrix of the probabilities of the transitions between all pairs of states over time t and Δt is the time step for a single transition. We have shown that this criterion can be conveniently expressed in terms of the eigenvalues of T .¹⁰ This method provides the low bound of Markovianity, however, it does not give any explanation of the mechanism that leads to non-Markovian behavior nor does it suggest ways to improve the analysis for time scales below the Markovian threshold.

In this work, we present a method that rigorously finds the Markovian time scale. Moreover, it explains in detail the mechanism that leads to non-Markov behavior of the biomolecule. Using this approach, new states can be defined from the initial, largely arbitrary phenomenological states. The new states possess Markov property and give an indication for the physical origin of non-Markovianity.

^{a)}Electronic mail: dn232@cam.ac.uk.

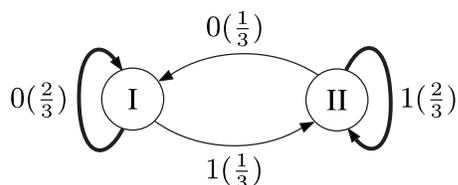


FIG. 1. The ϵ -machine for a two-state process (see text); the number at the arcs shows the symbol emitted during the transition between the causal states and the corresponding probability is in parenthesis.

II. METHOD

The methodology adopted is the “computational mechanics” (CM), developed by Crutchfield *et al.*^{12–14} CM builds a statistic on infinitely long histories (“pasts”) of symbols s_i representing the state of the system at times t_i , $\tilde{s}_i \equiv \{\dots s_{i-2} s_{i-1} s_i\}$, by analyzing the “futures” $\tilde{s}_i \equiv \{s_{i+1} s_{i+2} \dots\}$ following each past. The algorithm groups the pasts into classes called “causal states” ϵ_j . The criterion of grouping is the equivalence of the probabilities of the futures, that is, two pasts \tilde{s}_i and \tilde{s}_j are assigned to the same causal state if the distributions of their futures are the same: $P(\tilde{s}_i | \tilde{s}_i) = P(\tilde{s}_j | \tilde{s}_j)$, where $P(X|Y)$ is the probability of X given Y . Thus, instead of the transitions between the states s_i , the dynamics of the system is described by the probabilistic transitions between the causal states ϵ_j . Importantly, the sequence of ϵ_j is *Markovian by definition* regardless of the properties of the original process s_i . The collection of the causal states together with the transition probabilities between them is called an “ ϵ -machine.” The detailed definition of an ϵ -machine is provided in the Appendix. ϵ -machines can be reconstructed from observed data using the causal state splitting reconstruction (CSSR) algorithm described and implemented in Ref. 15.

As an illustrative example, consider a string of symbols 0 and 1:

...111101001000001011111100001111...

The string is generated using a Markovian process with the transition probabilities from symbol 0: $P_{0 \rightarrow 0} = \frac{2}{3}$, $P_{0 \rightarrow 1} = \frac{1}{3}$ and similarly for the transitions from symbol 1: $P_{1 \rightarrow 1} = \frac{2}{3}$, $P_{1 \rightarrow 0} = \frac{1}{3}$. The ϵ -machine for this process is shown in Fig. 1. For simplicity, consider only pasts of length 2: $\{00, 01, 11, 10\}$. Causal state I contains the pasts 00 and 10 because they lead to 0 and 1 with the same probabilities given above. When in state I, the system with probability $\frac{2}{3}$ emits symbol 0 and remains in the same state I. Symbol 1 is emitted and with probability $\frac{1}{3}$ the system is transferred to state II. Exactly the same situation with 0 and 1 interchanged is seen for state II. Because the process is Markovian, the same ϵ -machine is obtained when considering pasts of any length. In this case, state I (II) would contain all histories of the form $\{^*0\}$ ($\{^*1\}$), where * stands for any history including the null history.

The power of this representation becomes obvious when the process is non-Markovian. In our example, this would be the case if, for example, histories 00 and 10 would lead to different distributions of the next symbol. This would mean that the next state depends not only on the current state (0)

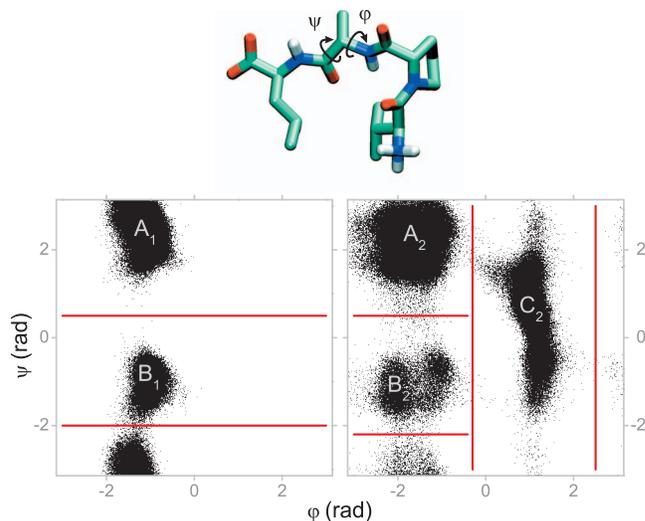


FIG. 2. The four residue peptide VPAL and the Ramachandran plots for the proline (left) and alanine (right). The clustering is marked by the boundaries that define the conformations as pairwise combinations of the indices from the sets $\{A_1, B_1\}$ and $\{A_2, B_2, C_2\}$.

but also on the previous state (0 or 1). An ϵ -machine can quantify such non-Markovian dynamics (of any order) by analyzing the pasts of all lengths. We stress again that the dynamics of the causal states themselves is Markovian, an important property for our analysis of molecular transitions. More specific use of CM for quantifying non-Markovian dynamics is considered in Sec. IV.

We have demonstrated the usefulness of CM in studying local dynamics of peptides in the process of the transitions between conformations.^{16,17} Here, we apply the framework to the global dynamics of the transitions between the peptide’s conformations.

III. SIMULATION DETAILS AND MOLECULAR SYSTEM

In this investigation, we analyze a molecular dynamics trajectory. The simulation was performed using the software package GROMACS 3.2.¹⁸ The system examined was the four residue peptide valine-proline-alanine-leucine (VPAL) solvated in 874 water molecules. The peptide is shown in Fig. 2. The simulation box was $30 \times 30 \times 30 \text{ \AA}^3$. The force field was Gromos 53a6.^{19–21} This is optimized for biomolecular systems interacting with water. Electrostatic interactions were treated by particle mesh Ewald summation technique. Periodic boundary conditions were used. The temperature was kept at 300 K using the Berendsen thermostat.²² Atomic positions were recorded every 0.5 ps. The integration algorithm was a Verlet type and the integration step was 0.002 ps. The system was equilibrated before it was sampled for 500 ns. This produced a total of 10^6 data points.

We start with a common method of defining the states of the system by clustering the spatial configurations of the molecule. The conformational states are defined by clustering the molecular dynamics (MD) simulated trajectories in the Ramachandran space (Fig. 2). Each Ramachandran plot is clustered independently and the molecule’s configurations are defined by the cluster indices from each plot. Not all possible combinations of index values are realized in the

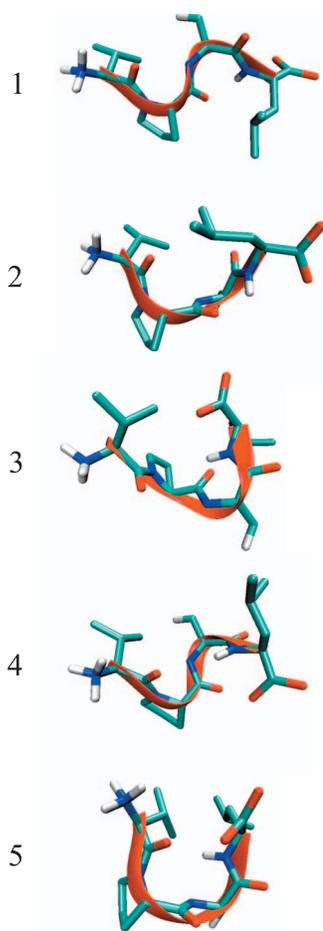


FIG. 3. The average conformations of the VPAL molecule defined by the clusters shown in Fig. 2. 1: A_1A_2 ; 2: B_1A_2 ; 3: $A_1C_2+B_1C_2$; 4: A_1B_2 ; and 5: B_1B_2 .

trajectory. For example, for the chosen peptide (Fig. 2), the conformation B_1C_2 was very sparsely populated and was therefore joined with A_1C_2 into one conformation, thus resulting in five total conformations of the molecule (Fig. 3).

The clustering is performed by choosing dihedral angles as cutoff angles between the different regions of conformational space. We use the two central pairs of dihedrals as the terminal residues are very flexible and do not define the long lived conformational states of the molecule. The resulting clustering is delineated by the straight lines in Fig. 2 and the corresponding averaged molecular conformations are shown in Fig. 3.

IV. RESULTS AND DISCUSSION

CM is formulated using the assumption of infinitely long pasts and futures. In calculations, a finite history length l has to be chosen and this is one of the adjustable parameters of the CSSR algorithm.¹⁵ The second parameter is the significance level σ used in comparing the distributions $P(\vec{s}|\vec{s}_i)$ (the Kolmogorov–Smirnov test is used). Therefore, for robust results, the analysis of the ϵ -machine as a function of these two parameters should be performed. Too long a history length or too large a σ value (more strict threshold for two distributions to be considered equivalent) lead artificially to too many causal states. This is equivalent to undersampling the

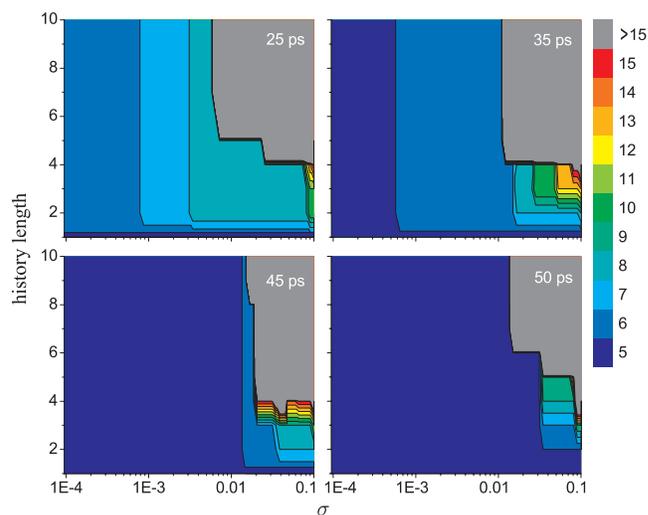


FIG. 4. The number of causal states in the ϵ -machines of VPAL dynamics as a function of the history length l and the tolerance σ for four different time steps.

histories. The number of possible histories grows exponentially with the history length, therefore, for long histories, an exponential increase in the number of data points is also needed. Thus, as the authors of CSSR recommend, the value of σ should be chosen such that there is a “plateau” in the number of causal states as a function of l . If there are more than one such value of l , then the lowest has to be chosen (according to the minimality principle of CM). This constant value of l is the “true” value of the history length for a stable ϵ -machine architecture.

The data analysis consisted of the VPAL trajectory symbolized using the five conformational states described in Sec. III. The number of causal states of the corresponding ϵ -machines as a function of σ and l are shown in Fig. 4. We have found that the CSSR algorithm converges producing stable ϵ -machine architectures for all cases studied (four examples are represented in Fig. 4). The causal states and corresponding history lengths are those that produce large areas of the same color on the plots and the larger ones correspond to regions that have more stable size of ϵ -machine.

For $l=1$, the machine cannot have more causal states than the number of symbols in the alphabet (the number of conformational states); in our case, five. If the process is Markovian, the same five states are found for longer histories. The plots show that the dynamics of the peptide remains Markovian down to a time step of ≈ 40 – 50 ps. The structure of the ϵ -machine corresponding to this dynamic regime is shown in Fig. 5. Here, each causal state has the most probable transition to itself. This corresponds to a commonly known transition matrix (the matrix of the probabilities of the transitions between the conformations) with the diagonal elements having the highest values describing the metastable conformational states of the molecule. The transitions between the causal states (the arcs on Fig. 5) describe the configurational transitions that have the probabilities given by the off-diagonal elements of the transition matrix

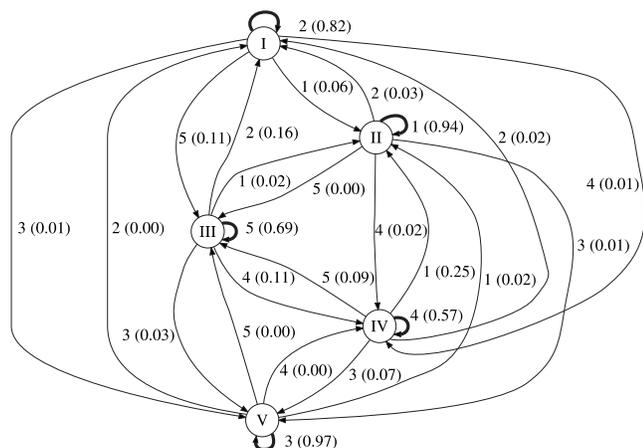


FIG. 5. ϵ -machine for the Markovian regime of the peptide's dynamics; the time step of the data is 50 ps; the causal states consist of the following histories: I: $\{^*2\}$, II: $\{^*1\}$, III: $\{^*5\}$, IV: $\{^*4\}$, and V: $\{^*3\}$.

$$T^{50 \text{ ps}} = \begin{bmatrix} 0.9391 & 0.0622 & 0.0189 & 0.2401 & 0.0207 \\ 0.0251 & 0.8209 & 0.0027 & 0.0143 & 0.1550 \\ 0.0132 & 0.0080 & 0.9717 & 0.0804 & 0.0250 \\ 0.0195 & 0.0050 & 0.0059 & 0.5456 & 0.1298 \\ 0.0031 & 0.1039 & 0.0008 & 0.1195 & 0.6695 \end{bmatrix}.$$

In this Markovian regime, the two descriptions (ϵ -machine and transition matrix between the conformational states) are equivalent and this is confirmed by the same transition probabilities.

However, at the time scale of ≈ 30 – 40 ps, a fundamentally different situation is observed. For histories $l \geq 2$, an additional causal state is required to adequately describe the dynamics. This indicates that the dynamics becomes non-Markovian. The ϵ -machine for this regime is shown in Fig. 6. The causal states corresponding to the Markovian situation are labeled with the same numbers I–V, while the additional

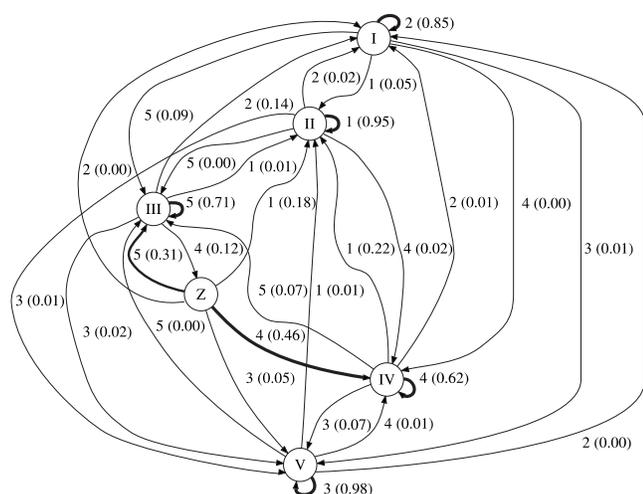


FIG. 6. An additional causal state Z compared to the Markovian ϵ -machine (Fig. 5) signifies the non-Markovian property of the transition from the configurational state 5 to 4; the time step of the data is 40 ps; the causal states consist of the following histories: I: $\{^*2\}$, II: $\{^*1\}$, III: $\{^*5\}$, IV: $\{^*14, ^*24, ^*34, ^*44\}$, V: $\{^*3\}$, and Z: $\{^*54\}$.

state causing non-Markovianity is designated with letter Z. The transition matrix between the configurational states for this time scale is

$$T^{40 \text{ ps}} = \begin{bmatrix} 0.9485 & 0.0542 & 0.0156 & 0.2076 & 0.0141 \\ 0.0217 & 0.8444 & 0.0022 & 0.0100 & 0.1382 \\ 0.0109 & 0.0067 & 0.9764 & 0.0690 & 0.0198 \\ 0.0168 & 0.0036 & 0.0052 & 0.6005 & 0.1190 \\ 0.0022 & 0.0912 & 0.0006 & 0.1129 & 0.7090 \end{bmatrix}.$$

It is very instructive to compare the causal states describing the Markovian transitions with the corresponding transition elements of the $T^{40 \text{ ps}}$ matrix. For this, we calculated the relative difference as a percentage for $\Delta T_{i,j} = (T_{i,j}^\epsilon - T_{i,j}^{40 \text{ ps}}) / T_{i,j}^{40 \text{ ps}}$, where the elements of T^ϵ are the transition probabilities between the causal states I–V of Fig. 6

$$\Delta T(\%) = \begin{bmatrix} 0.0 & 0.8 & -7.6 & 6.9 & -17.4 \\ 0.2 & 0.1 & -5.0 & -2.4 & 1.1 \\ -6.5 & -16.4 & 0.1 & 2.5 & 2.9 \\ 2.6 & -61.1 & 0.5 & 4.0 & 2.7 \\ -10.9 & 2.2 & 4.5 & -35.2 & -0.4 \end{bmatrix}.$$

The diagonal elements are not significantly different, with the exception of conformational state 4. This is not surprising since the additional state Z modifies only the transition $5 \rightarrow 4$. In contrast, the transitions between the conformations are modified significantly, especially those that involve conformational states 4 and 5. The transitions between other conformations (not involving 4 and 5) are also changed substantially, especially the $2 \rightarrow 3$ transition (-16.4%).

In order to elucidate the mechanism that leads to non-Markovian behavior, we analyze the content and connections of the causal state Z. This causal state contains only one history 54 (machines built from longer histories have the history *54 that carries the same information since * corresponds to any history). The only causal state that leads to state Z is state III which contains the histories ending with 5. This is a consequence of the fact that causal state Z has only one history with the next to the last symbol 5. The two main transitions out of state Z are those emitting symbols 4 and 5 (Fig. 6). The former means that the molecule changes the conformation to configuration 4, while the latter indicates that the molecule returns to state 5, where it came, from two steps before.

Thus, some trajectories passing through conformational state 4 behave differently and for this reason they are selected into a separate causal state Z. It is therefore logical to separate them into an additional conformational state (Fig. 7). This will make both types of trajectories visiting conformation 4 Markovian (but different). Now the conformational states coincide with the causal states. This process can be generalized to situations where more non-Markovian trajectories are found through the application of CM. The obvious advantage of this approach to building the correct conformational states from the initial empirical states is that the Markov property is fulfilled automatically: the causal states are Markov by definition. The physical mechanism of such non-Markov behavior of some trajectories is that they are, in

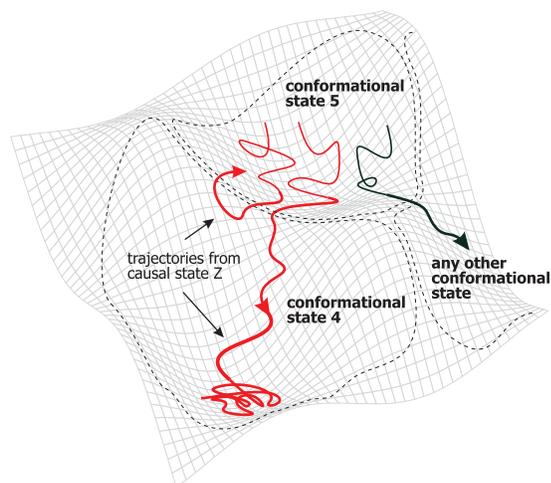


FIG. 7. Schematic illustration of the criteria for defining new conformational states.

reality, several true free energy minima lumped together by the *ad hoc* procedure of defining the conformational states. How this is related to ϵ -machines requires additional investigation.

The new conformational state (the molecular configurations that are in state 4 and have been in state 5 at the previous time step) can be visualized and compared to the complementary state (all other configurations of state 4, Fig. 8). It is clear that the two states are mixed uniformly and cannot be separated using geometric clustering. Probably, the states can be distinguished in the space including other degrees of freedom not considered in the present analysis. For example, they could be the conformations of the side chains, or the terminal residues dihedrals, or even some water molecules that play important roles in the conformational dynamics of the peptide. The new state constitutes a significant fraction of states visited by the biomolecule. The fraction of the number of occurrences of the new state is $\approx 20\%$.

The new additional states are conceptually similar to expanding the state space of dynamical system with the help of Takens embedding theorem.²³ By using the time delayed coordinates, it is possible to include the memory of the system into the state space. However, CM is more suitable in our case since it automatically selects only those time moments when the trajectory is non-Markovian and needs to be expanded into a separate state.

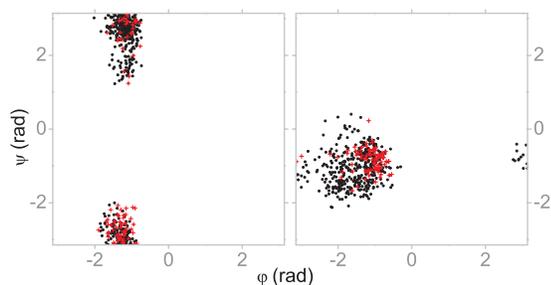


FIG. 8. Conformation 4 divided into two classes as described in the text; red crosses: the conformations corresponding to the histories *54 ; black dots: all other conformations.

TABLE I. The average residence time for each of the conformational states from Fig. 3.

State	Residence time (ps)
1	83
2	25
3	191
4	10
5	13

Finally, an important issue is the relevance of the time scales described above. Is it justified to investigate the data at a 40 ps time step (non-Markovian) or a 50 ps time step (when everything behaves Markovian)? An obvious answer to this question involves the time which the trajectory spends in the initially defined conformational states. This time should be long enough for the system to “forget” the history of the previous states. Having the simulated trajectory, it is trivial to calculate this and the result is given in Table I. Since the trajectory passes states 2, 3, and 5 faster than the non-Markovian time scale investigated above, our results define a relevant description of the dynamics. It should be noted that Altis *et al.*²⁴ have found the same match of time scales of non-Markovian behavior and residence times for heptaalanine. It could make sense to analyze the dynamics at even shorter time scales, but in this case the ϵ -machine becomes complex and requires more tedious analysis, which is the subject of our current research.

V. CONCLUSIONS

The Markovian character of transitions between molecular conformations crucial for correct transition rate calculations is analyzed for a four residue peptide simulated using MD in explicit water for 0.5 μ s. It is found that the dynamics can be considered Markovian at the time scale of 50 ps and longer. However, at the time scale of 30–40 ps the dynamics is clearly non-Markovian. The methodology that we use reveals the mechanism of non-Markovianity and shows that the trajectories passing through one of the conformational states are split into two dynamically different groups. The residence times in each of the conformational states are comparable or even smaller than 30–40 ps, thus the described non-Markovian effect is essential for the dynamics of the peptide. Finally, the methodology allows constructing the conformational states that are Markovian and can be deduced from the initial, largely arbitrary conformational states.

ACKNOWLEDGMENTS

The authors would like to thank Frank Noe and Christof Schuette for stimulating comments on the manuscript. The work is supported by Unilever and the European Commission (EC) (Contract No. 012835-EMBIO).

APPENDIX: COMPUTATIONAL MECHANICS

All past \tilde{s}_i and future \tilde{s}_i halves of bi-infinite symbolic sequences centered at times i are considered. Two pasts \tilde{s}_1 and \tilde{s}_2 are defined equivalent if the conditional distributions

over their futures $P(\tilde{s}^1|\tilde{s}_1)$ and $P(\tilde{s}^2|\tilde{s}_2)$ are equal. A *causal state* $\epsilon(\tilde{s}_i)$ is a set of all pasts equivalent to \tilde{s}_i : $\epsilon_i \equiv \epsilon(\tilde{s}_i) = \{\lambda : P(\tilde{s}|\lambda) = P(\tilde{s}|\tilde{s}_i)\}$. At a given moment, the system is at one of the causal states, and moves to the next one with the probability given by the transition matrix $T_{ij} \equiv P(\epsilon_j|\epsilon_i)$. The transition matrix determines the asymptotic causal state probabilities as its left eigenvector $P(\epsilon_i)T = P(\epsilon_i)$, where $\sum_i P(\epsilon_i) = 1$. The collection of the causal states together with the transition probabilities define an ϵ -machine.

It is proven²⁵ that the ϵ -machine is

- (1) a *sufficient* statistic, that is it contains the complete statistical information about the data;
- (2) a *minimal sufficient* statistic, therefore, the causal states cannot be subdivided into smaller states; and
- (3) a *unique minimal sufficient* statistic, any other one simply relabels the same states.

¹C. Schuette, A. Fischer, W. Huisinga, and P. Deuffhard, *J. Comput. Phys.* **151**, 146 (1999).

²W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).

³W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, T. J. C. Ward, Y. Zhestkov, and R. Zhou *J. Phys. Chem. B* **108**, 6582 (2004).

⁴F. Noe, I. Horenko, C. Schuette, and J. C. Smith, *J. Chem. Phys.* **126**, 155102 (2007).

⁵J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *J. Chem. Phys.* **126**, 155101 (2007).

⁶S. Park and V. S. Pande, *J. Chem. Phys.* **124**, 054118 (2006).

⁷J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, *Multiscale Model. Simul.* **5**, 1214 (2006).

⁸G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, *J. Chem. Phys.* **131**, 124101 (2009).

⁹F. Noe and S. Fischer, *Curr. Opin. Struct. Biol.* **18**, 154 (2008).

¹⁰C. H. Jensen, D. Nerukh, and R. C. Glen, *J. Chem. Phys.* **128**, 115107 (2008).

¹¹C. H. Jensen, D. Nerukh, and R. C. Glen, *J. Chem. Phys.* **129**, 225102 (2008).

¹²J. P. Crutchfield and K. Young, *Phys. Rev. Lett.* **63**, 105 (1989).

¹³J. P. Crutchfield and K. Young, in *Entropy, Complexity, and Physics of Information*, SFI Studies in the Sciences of Complexity Vol. VIII, edited by W. Zurek (Addison-Wesley, Reading, 1990).

¹⁴J. P. Crutchfield, *Physica D* **75**, 11 (1994).

¹⁵C. R. Shalizi and K. L. Shalizi, in *Uncertainty in Artificial Intelligence: Proceedings of the Twentieth Conference*, edited by M. Chickering and J. Halpern (AUAI, Arlington, 2004), pp. 504–511.

¹⁶D. Nerukh, G. Karvounis, and R. C. Glen, *Lecture Notes in Bioinformatics*, Lecture Notes in Computer Science Vol. 4216 (Springer, Berlin/Heidelberg, 2006), pp. 129–140.

¹⁷D. Nerukh, G. Karvounis, and R. Glen, *Complexity* **10**, 40 (2004).

¹⁸D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).

¹⁹B. Hess and N. F. A. van der Vegt, *J. Phys. Chem. B* **110**, 17616 (2006).

²⁰C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, and W. F. van Gunsteren, *Eur. Biophys. J.* **34**, 273 (2005).

²¹C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, *J. Comput. Chem.* **25**, 1656 (2004).

²²H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).

²³F. Takens, *Detecting Strange Attractors in Turbulence*, Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, 1981), pp. 366–381.

²⁴A. Altis, M. Otten, P. H. Nguyen, R. Hegger, and G. Stock, *J. Chem. Phys.* **128**, 245102 (2008).

²⁵C. Shalizi, K. Shalizi, and R. Haslinger, *Phys. Rev. Lett.* **93**, 118701 (2004).