Identification of metastable states in peptide's dynamics

Svitlana Ruzhytska,[1,a)] Martin Nilsson Jacobi,[1] Christian H. Jensen,[2] and Dmitry Nerukh[2]
[1]*Department of Energy and Environment, Chalmers University of Technology,*
Gothenburg 412 96, Sweden
[2]*Department of Chemistry, Unilever Centre for Molecular Sciences Informatics, Cambridge University,*
Cambridge CB2 1EW, United Kingdom

A recently developed spectral method for identifying metastable states in Markov chains is used to analyze the conformational dynamics of a four-residue peptide valine-proline-alanine-leucine. We compare our results to empirically defined conformational states and show that the found metastable states correctly reproduce the conformational dynamics of the system. © *2010 American Institute of Physics.* [doi:10.1063/1.3493333]

I. INTRODUCTION

Conformational dynamics is fundamental for the majority of biomolecular transformations, including such well known biochemical processes as protein folding and ligand docking. The biomolecule passes through many intermediate configurations on the way from the initial to the final state. The latter two are normally well defined and can be experimentally identified. This is not the case, however, for the intermediate states that are often only possible to elucidate in computer simulation. The information on the correct intermediates, their probabilities, and the transitions between them is necessary for the estimation of correct, experimentally verifiable rates of biochemical reactions and understanding their mechanisms. For this reason, identifying the conformational states and investigating their dynamics have recently been a topic of very active development (see Refs. 1–8 for a few recent examples). It should be noted that it is currently impossible to predict the conformational states of a biomolecule from its chemical structure. The entire dynamical evolution of the biomolecular system (often including water molecules explicitly) is required. However, even all first principle information on the molecular motion on the atomistic scale does not provide the conformational states directly: the information has to be properly processed. Moreover, the resulting physicochemical parameters of the whole biomolecular process strongly depend on how the states are defined, and subtle variations in the states definition can lead to orders of magnitude differences in the molecular characteristics, especially the kinetics of the process.[9,10]

Significant efforts are devoted to developing ways of grouping elementary conformations (the "microstates") into the intermediate metastable states (the "macrostates"); see, for example, the recent overview in Ref. 8. However, almost all of them use the structural similarity in either configurational or free energy space for initial clustering of the instantaneous molecular conformations into the microstates. This is done *before* the application of the kinetic approaches to forming the macrostates. The procedure is necessary for ac-cumulating the data on the transitions between the microstates; however, an unbiased approach based on purely dynamic (not structural) considerations is highly desirable.

We have recently investigated a four-residue peptide valine-proline-alanine-leucine (VPAL) (Fig. 1).[9–11] The peptide is a convenient model for studying conformational changes as one of the smallest biomolecules having dynamics that resembles the "folding" process. The system is small enough to sample the conformational space exhaustively. This is important for statistically correct conclusions. In our investigations,[9–11] we assigned the molecular configurations to different states empirically by simply drawing straight boundaries between the clusters on the Ramachandran plots (Fig. 2). It is therefore important to investigate how this definition of the states relates to those rigorously defined by the time spent in each conformation and using formal analysis of the Markov chain formed by the system's conformations.

The most efficient techniques for identification of metastable states in large Markov chains are based on the level structure of the eigenvectors of the transition matrix whose corresponding eigenvalues are clustered close to 1. The basic idea of using eigenvectors to identify tightly connected sets of states was first introduced by Fidler[12] in the context of graph partitioning. The same idea was later applied to identify metastable states in reversible Markov chains.[13] In this paper, we use a recent spectral method that can be applied to nonreversible processes.[14]

We here show that the dynamics of the molecule can be described as having several metastable conformations with quick transitions between them. We also demonstrate that these conformations largely coincide with our previous, empirically defined ones. Moreover, the metastable conformations found in this paper reproduce more subtle features of the dynamics that we found using completely different approach, the Markovian property of the states.[11]

II. IDENTIFYING METASTABLE STATES USING SPECTRAL METHODS

Consider a Markov chain with a state vector x describing the probability distribution over the states and a transition matrix P such that $x(t+1)=Px(t)$. Assume that the number of

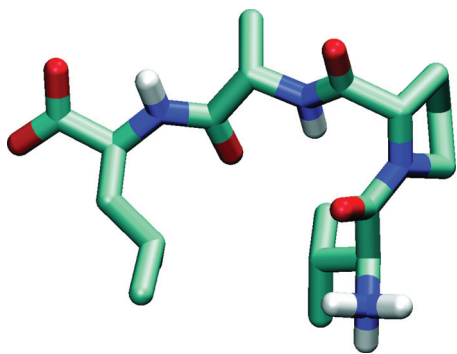a)Electronic mail: svitlana@chalmers.se.

FIG. 1. The valine-proline-alanine-leucine (VPAL) peptide molecule. Carbon atoms are light blue, nitrogens are dark blue, oxygens are red, and hydrogens are gray.

possible states is large such that they adequately represent the dynamics of the system. Metastable states of a Markov chain are defined as sets of the states such that the transitions between the states within the same metastable state are much more likely than the transitions between the states belonging to different metastable states. Identifying the metastable states given a large transition matrix can be achieved using the analysis of the eigenvalues and the eigenvectors of P.[13,14] The basic idea can be understood by considering the extreme case of a completely decoupled Markov chain with a transition matrix that can be permuted, i.e., changing the ordering of the states, into a block diagonal form,

$$P = SP_0 S^{-1} = \begin{pmatrix} P^{(1)} & 0 & \cdots \\ 0 & P^{(2)} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \quad (1)$$

where P_0 is the original transition matrix where the rows and columns are not ordered in such a way that the block diagonal structure is obvious. Each of the blocks defines a metastable [in Eq. (1), completely stable] aggregated state. The block matrices on the diagonal, $P^{(n)}$, describe the transition probabilities within the corresponding stable state, i.e., they reflect the fact that the stable states define independent Markov processes. The problem is now, given P_0, to find the permutation S that reveals the stable states. The set of complex eigenvalues λ_i (also called the "spectrum" of the matrix) and (right) eigenvectors of P, u_i, such that $Pu_i = \lambda_i u_i$, can be constructed directly from the different blocks $P^{(n)}$. To see this, let $u^{(n)}$ be an eigenvector of $P^{(n)}$ with an eigenvalue λ.
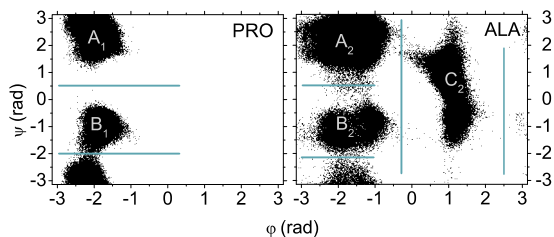

FIG. 2. The Ramachandran plots for the proline (left) and alanine (right) residues of the VPAL molecule. These plots are partitioned into the areas marked A_1, B_1, A_2, B_2, and C_2, where index 1 stands for proline and index 2 for alanine. Each combination of two areas with different indices corresponds to one conformation of the peptide. This partitioning was used in Refs. 9 and 10 for defining the metastable conformations.

Clearly, the vector $u = (0, \ldots, 0, u^{(n)}, 0, \ldots, 0)^T$ is then an eigenvector of P with eigenvalue λ.

A general property of a Markov transition matrix is that its largest eigenvalue, called the Perron–Frobenius eigenvalue, is 1. The corresponding right eigenvector defines the stationary distribution, and the elements of the left eigenvector are all identical (see Ref. 14 for details). The trivial form of the left eigenvector with the eigenvalue 1 follows from the conservation of probability; $\Sigma_k P_{kl} = 1 \ \forall l$ implies $v^T P = v^T$, if all the elements of v are identical. Since each block $P^{(n)}$ is a Markov chain, each with its own Perron–Frobenius eigenvalue, the total transition matrix P has N eigenvalues at unity, if it contains N metastable states. The vector

$$v^{(n)} = (0, \ldots, 0, \underbrace{a, a, \ldots, a, a}_{\text{block } n}, 0, \ldots, 0), \quad (2)$$

with a constant a given by the normalization of the eigenvector, is a left eigenvector with eigenvalue 1. The idea of how the eigenvectors can be used to identify the metastable states now becomes clear, since the states in a block can be identified by the identical nonzero elements in one of the left Perron–Frobenius eigenvectors. One possible complication remains. The eigenvectors corresponding to a degenerate eigenvalue λ are not unique since any linear combination is also an eigenvector: $(\alpha v^{(n_1)} + \beta v^{(n_2)})^T P = \lambda (\alpha v^{(n_1)} + \beta v^{(n_2)})^T$. A linear combination of vectors in the form in Eq. (2) has the form

$$v = (\underbrace{a_1, \ldots, a_1}_{\text{block } 1}, \underbrace{a_2, \ldots, a_2}_{\text{block } 2}, \ldots). \quad (3)$$

Therefore, if a hidden block diagonal structure exists in P_0, the dominant left eigenvectors of P_0 are of the form of Eq. (3) but with the elements in some random order. The states belonging to the same metastable state can therefore be identified by identical elements in the N left Perron–Frobenius eigenvectors (N being the number of metastable states). This can also be expressed as a clustering problem, where each state is associated with a vector consisting of its elements in the N dominant left eigenvectors of P_0. For states belonging to the same metastable state, these vectors of length N should be the same.

The discussion so far has actually not been on metastable, but completely stable, states: in Eq. (1), there are no transitions between the blocks. Metastability occurs when the transition matrix can be rearranged into a block dominant rather than block diagonal form. This means that the elements outside the blocks in P are either few, small, or both small and few. In this case, we expect the conclusions from the spectral analysis to hold approximately: the Perron–Frobenius eigenvalue is almost degenerate with a cluster of N eigenvalues close to 1, and the corresponding left eigenvectors show an appropriate level structure. This picture is correct for Markov chains with symmetric or reversible transition matrices.[13] However, in more general situations, the spectral analysis is not necessarily stable to perturbations. This follows from the fundamental fact that eigenvectors of nonsymmetric matrices can be very sensitive to perturbations, i.e., the eigenvector problem can be ill-conditioned. This problem can however be addressed by constructing a

symmetric matrix with eigenvectors sharing the same level structure as the original transition matrix (see Ref. 15 for details).

III. NUMERICAL SIMULATIONS

A. Obtaining molecular dynamics trajectory

We analyze the molecular dynamics trajectory of the four-residue peptide, valine-proline-alanine-leucine (VPAL), shown in Fig. 1. The trajectories of the system are found using a molecular dynamics simulation package GROMACS.[16] The peptide molecule was placed in a $3 \times 3 \times 3$ nm box with periodic boundary conditions. The solvent was modeled by 874 simple point charge (SPC) water molecules.[17] The pressure and temperature were kept constant during the simulation at 1 atm and 300 K using the Berensen thermostat.[18] The interactions were modeled by the forcefield $53a6$ optimized for biomolecular systems interacting with water.[19–21] After initial equilibration, the trajectory was sampled every 0.5 ps, and the atomic positions were recorded for 220 ns, resulting in a total of 440 000 data points. However, we tested our methods and showed that using only 220 000 data points obtained from 110 ns simulation did not affect the results but appreciably speed up the calculations. Therefore, further discussion will be held for 220 000 data points.

B. Defining Markov microstates of the peptide

The configurations of the VPAL peptide are obtained in terms of the atomic positions in Cartesian coordinates. The peptide consists of 64 atoms in total, which produce $64 \times 3 = 192$ coordinate values per time step and our 220 000 time frames result in 42.24×10^6 trajectory data values. Since we are only interested in the changes of the internal configuration of the biomolecule, it is sufficient to work with the residues' dihedral angles. We used the angles of the two central residues, proline and alanine, because the outermost residues are too flexible and do not define the general structure of the peptide. Each residue is described by the two torsional angles; thus, four angle values describe the molecular configuration, which in total gives $220 000 \times 4 = 880 000$ trajectory data values. Changes of these angles with time are shown in Fig. 2 (the Ramachandran plots).

Converting the continuous trajectory into a set of discrete microstates has been done by partitioning the four-dimensional hypercube describing the angular coordinates ($[0, 2\pi[\times [0, 2\pi[\times [0, 2\pi[\times [0, 2\pi[$) into cells by dividing each dimension into ten equal intervals. For our trajectory, only 1306 of the total 10^4 possible microstates were nonempty. On average, each found microstate corresponded to 170 time configurations, and the occupancy fluctuated between 1 and almost 12 000 points per microstate (Fig. 3).

The transition probability matrix was constructed from the trajectory by counting the number of transitions between the microstates and normalizing the total probability to one (Fig. 4).

However, care must be taken at this stage since the dynamics of the resulting states can be non-Markovian, if the time step is chosen too short. It was shown in Ref. 9 that the minimum time step for the Markovian dynamics is 50 ps for
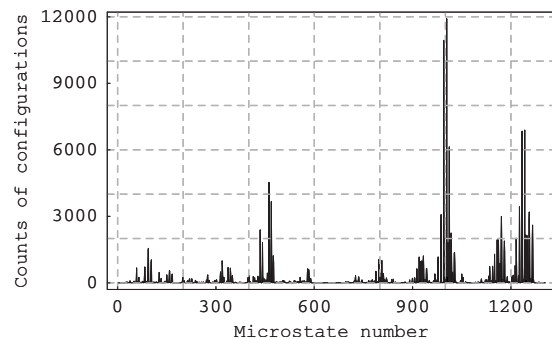

FIG. 3. Microstate statistics. Each microstate is a nonempty cell of the four-dimensional hypercube filled with points representing the molecular dynamics trajectory. In total, there are 1306 nonempty cells (or microstates), and this figure shows the points (or conformations) occupancy of each such cell.

the system studied here. This value depends on the partitioning of the state space; therefore, we performed a test for our partitioning that confirmed the results in Ref. 9. The count matrix was thus constructed using the 50 ps time sampling. The sampling rate was also varied to check for the robustness of the Markovian property. No significant differences were found, and we conclude that the resulting metastable states do not depend critically on the sampling rate. The count matrix showed that the minimum number of transitions into a microstate was 1, whereas the maximum was almost 300, and on average each microstate had 30 transitions in and out. The statistics of the microstate transitions is shown in Fig. 5.

IV. RESULTS AND DISCUSSION

We start with analyzing the number of metastable states using the spectrum of the original transition matrix P. As it can be seen from Fig. 6, there are four eigenvalues close to the Perron–Frobenius eigenvalue of 1. These eigenvalues are separated from the rest of the spectrum by a gap. From this, we conclude that the system exhibits four metastable states. We call these eigenvalues *significant*.

The spectrum has a large number of complex eigenvalues. This indicates that the Markov chain is nonreversible. The reversibility condition for a Markov transition matrix is
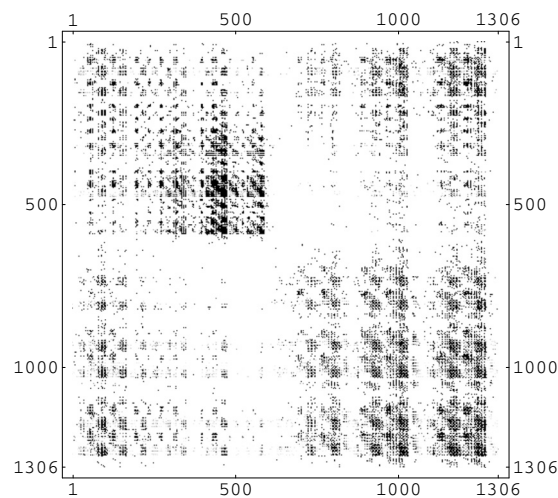

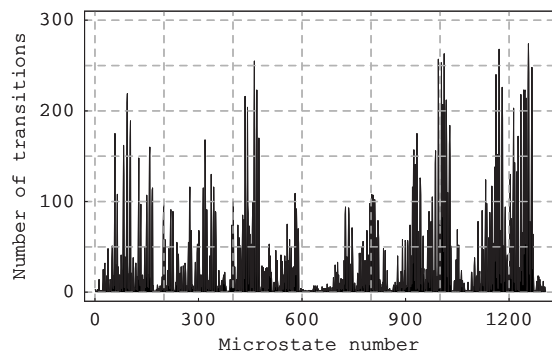FIG. 4. The initial P matrix. Darker points correspond to more transitions.

FIG. 5. Microstate transitions statistics. Following the trajectory, we counted the number of transitions between the microstates. This figure shows the number of the transitions to all other microstates in total for each microstate.

FIG. 7. The first significant eigenvector of P, u, whose elements, i, are sorted to reveal the level structure.

that there exists a vector u_{0i} such that $P_{ji}u_{0i}=P_{ij}u_{0j}$. By summing over i, it is clear that u_0 is the stationary distribution, that is, $Pu_0=u_0$. Intuitively, this means that we observe the transition $i \rightarrow j$ and its reverse transition $j \rightarrow i$ with equal probability, which is often referred to as the detailed balance condition. If we define a diagonal matrix $D_{ii}=1/\sqrt{u_{0i}}$, the similarity transformation $D^{-1}PD$ transforms a reversible transition matrix P into a symmetric matrix with real eigenvalues. Since the similarity transformation does not change the eigenvalues, it follows that the eigenvalues of a reversible Markov chain are real. In Fig. 6, we see that this is not true for our system. Another sign of nonreversibility is that there exists 38 012 cases when $P_{ij}>0$ while $P_{ji}=0$, which contradicts the reversibility condition. It should be stressed that the nonreversibility is the results of finite sampling and not a feature of the underlying system, which is in thermal equilibrium. Numerically, however, the nonreversibility can have a significant impact on the stability of the spectral analysis of the transition matrix. In our case, this does not happen however, as can be shown by comparing the result obtained directly from the transition matrix to the results of using a more elaborate method introduced in Ref. 15.

In Fig. 7, the first significant eigenvector of P is shown, where the elements are sorted to reveal the level structure. To find the sets of similar values in the eigenvectors, we applied the K-means clustering method[22] in the three-dimensional space formed by the second, third, and fourth eigenvectors (the first eigenvector can be omitted since it is constant). The
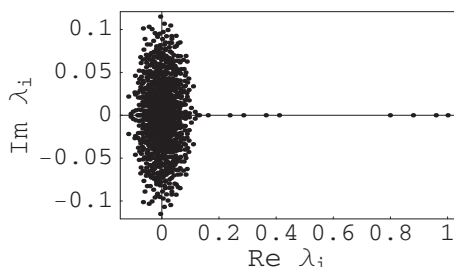
clustering result is shown in Fig. 8. The numbers of points in each cluster are 250, 480, 207, and 369 for the corresponding metastable states.

Using the clustering result, the initial transition probability matrix was permuted (rows and columns sorted so that the states in the same cluster appear in sequence), and the hidden block dominant structure was revealed (Fig. 9). The transitions within the metastable states defined by the blocks are more probable than the transitions between the metastable states. Not all states belong to a metastable state. The states outside the metastable blocks represent transitions between the states, that is, the conformations of the peptide that are visited only when the molecule changes from one metastable state to another.

In addition to the visual representation of the results (Fig. 9), we also present a numerical one by means of the reduced probability matrix (4). This matrix was constructed from the trajectory in the following way. Each point of the trajectory belongs to some microstate; at the same time, each microstate belongs to some metastable state. So, following the trajectory, the number of the transitions between the metastable states was counted and the total probability of the row was normalized to one,

FIG. 8. Eigenvector clustering results. Three eigenvectors of the P matrix corresponding to the significant eigenvalues were clustered into four clusters. Each point represents one microstate, each group of points of the same color belongs to one cluster, and each cluster corresponds to one metastable state.

FIG. 6. The real and imaginary parts of the eigenvalues of the P matrix, λ. Four eigenvalues are separated from the others by a spectral gap, from which we conclude that the peptide has four metastable states (we call these eigenvalues significant to distinguish them from the others).
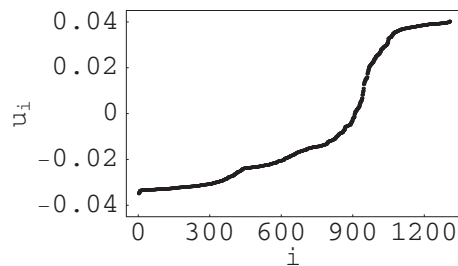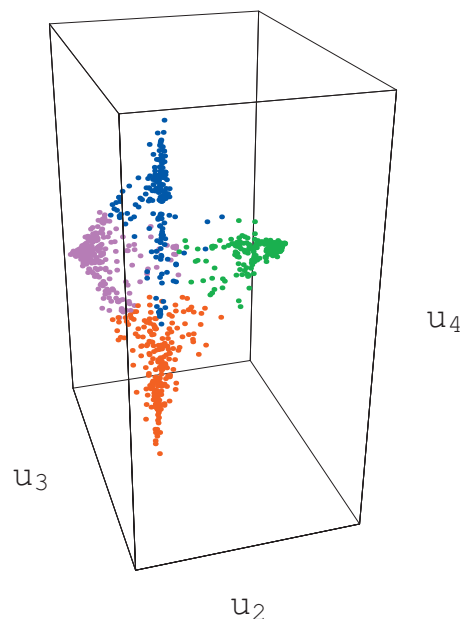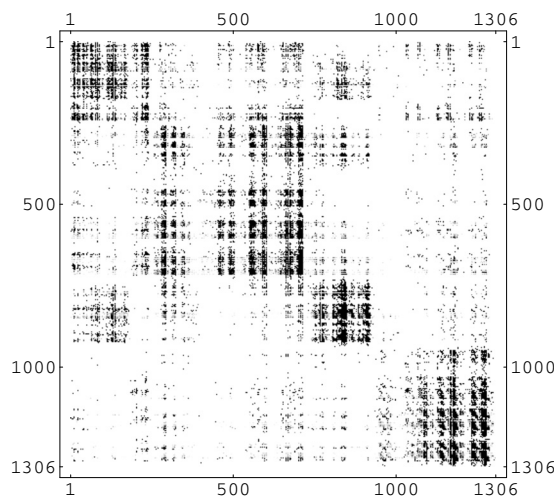
FIG. 9. The blocked (permuted) P matrix.

P_{red}

$$= \begin{pmatrix} 0.795\,988\,0 & 0.097\,460\,6 & 0.056\,272\,2 & 0.050\,279\,3 \\ 0.018\,174\,1 & 0.937\,747\,0 & 0.031\,854\,5 & 0.012\,224\,0 \\ 0.077\,611\,4 & 0.048\,468\,0 & 0.861\,560\,0 & 0.012\,360\,7 \\ 0.003\,598\,3 & 0.017\,442\,8 & 0.004\,486\,9 & 0.974\,472\,0 \end{pmatrix}.$$

(4)

High values of the diagonal elements indicate good quality of the results since it means that the transitions are most probable within the metastable states.

Finally, in Fig. 10, we show the density Ramachandran plots illustrating the found metastable conformations with the notation defined in Fig. 2. Our result shows that the VPAL molecule has metastable states corresponding to the following dihedral angle combinations: $A_1B_2+B_1B_2$, A_1A_2, B_1A_2, and A_1C_2 (B_1A_1 was not populated by the trajectory). Compared to what was obtained in an earlier investigation conducted by visual inspection, the clustering is almost identical. The only difference is that the method has concatenated two states into one, namely, $A_1B_2+B_1B_2$. This essentially shows one of the advantages in using the method, which is that the clustering is based on the dynamics of the system and not just on where the density of states is high. The method is therefore much better adapted to distinguish real metastable states from spurious ones.

The computation of the dominant eigenvectors of the transition matrix is done using a power method.[23] The computational complexity is therefore directly proportional to the number of nonzero elements in the matrix. The total number of microstates increases dramatically with the box size used when we partition the phase space (since we have four angles, it scales with the number of bins l), and consequently, the number of nonzero elements in the transition matrix could possibly grow as $(l^4)^2$. However, in practice, there are two effects that prevent this scaling to occur. First, only a relatively small part of the phase space is actually visited. Second, the transitions are typically localized in the phase space, resulting in a sparse transition matrix where the total
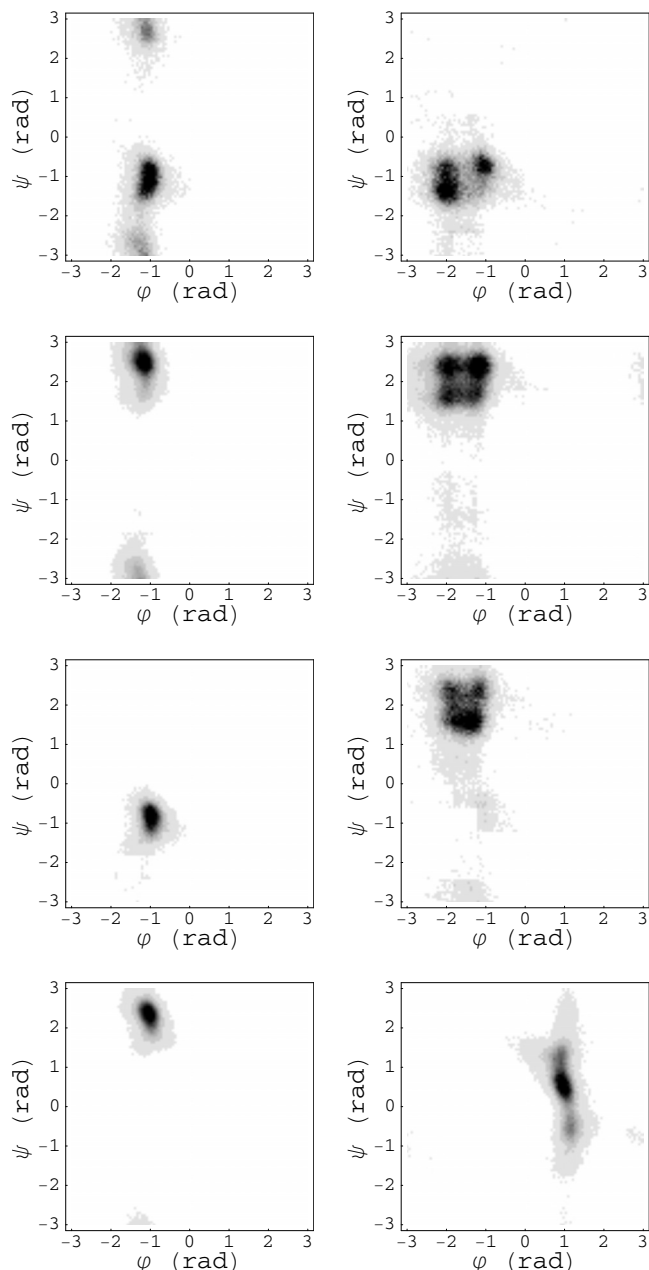
FIG. 10. Ramachandran density plots. Each pair of plots shows the area that corresponds to one metastable conformation. We obtained the following combinations (rows from top to bottom): $A_1B_2+B_1B_2$, A_1A_2, B_1A_2, and A_1C_2.

number of transitions scales essentially linearly with the number of states. In Fig. 11, the black line shows how the total number of nonzero transitions scales in relation to the total number of possible transitions. We also show how sparse the transition matrix is measured in percentage filling (the red curve in the plot). From this we conclude that the computational complexity of the algorithm suggested here remains reasonable even for relatively fine partitions of the state space.

V. CONCLUSIONS

We have used a spectral method to identify the metastable conformational states of a biomolecular system represented by a Markov chain. The system consisted of a four-
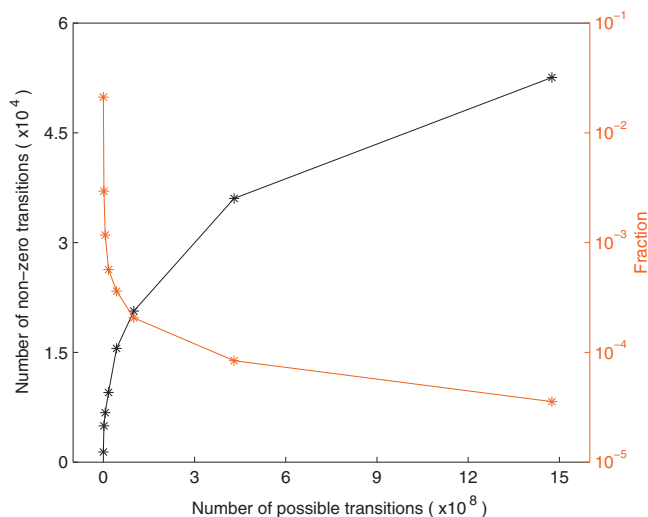
FIG. 11. Black: the number of observed transitions between the microstates vs the number of possible transitions, $(l^4)^2$, for the given number of bins l. Red: the fraction of nonzero elements in the corresponding transition matrix.

residue peptide simulated using Molecular Dynamics in a bath of explicit water. The conformational dynamics of the peptide represented by two pairs of the dihedral angles reveals four well separated metastable states. The states coincide with the previously empirically defined ones,[9,10] thus confirming the intuitive chemical picture by the rigorous mathematical analysis.

Interestingly, one of the found metastable states is equivalent to the sum of two previous ones, $A_1B_2+B_1B_2$ (Fig. 2). Our more elaborate analysis[11] shows that a significant part of the trajectories passing through the state A_1B_2 indeed needs to be separated and joined with the state B_1B_2 in order to compensate for slight non-Markovianity of the initial five states.

In the example discussed in this paper, the transition matrix showed a clear sign of metastability by having a visible gap in its spectrum. When this is the case, it is easy to immediately conclude how many metastable states to search for in the eigenvector clustering algorithm, as well as how many eigenvectors to include in the analysis. More complex peptides or proteins are not expected to have such clear spectral gaps, and in addition, we expect metastable states to organize into hierarchies or relevant time scales where metastability can be identified. It should be noted, however, that this situation does not necessarily prevent the application of the methods presented here. In the absence of a spectral gap, it is possible to successively extend the number of eigenvectors and metastable states and monitor how well the method manages to identify metastability, i.e., by measuring how dominant the diagonal of the reduced transition matrix is. An example of this type of blind identification was made in Ref. 15. An alternative method for identifying metastable states from the level structure in the eigenvectors in the absence of a spectral gap in the spectrum was presented.[24]

The identification of metastable states discussed in this paper is based on a discretization of the state space. For a large protein with a backbone defined by hundreds or thousands of dihedral angles, a naive partitioning of the states space leads to very large number of states in the Markov chain. As discussed in Sec. IV, the actual number of states visited is typically very small. Still, however, the resulting Markov chain may be too large to allow efficient analysis. In these cases, the discretization of the state space can be based on secondary structure or other properties (see, e.g., Ref. 25). It should be noted that the spectral method is independent of the partitioning of the state space as long as the partitioning is fine grained enough for the metastable states to be resolved.

The sparsity of the transition matrix, illustrated in Fig. 11, allows for numerically efficient calculation of the dominant eigenvalues and eigenvectors needed for the analysis presented in this paper. Still standard linear algebra software can typically not handle matrices much larger than $10^4 \times 10^4$. It is interesting to note that the problem of identifying metastable states in Markov chains is very similar to the recently very active topic of finding community structure in large networks.[12,26] In this area, several specially designed algorithms have been constructed to achieve this task in very large networks. For example, in Ref. 26, 1684 communities (metastable states) were identified in a network with 4×10^5 nodes (states) and 2.5×10^6 edges (observed transitions). In more recent studies, networks with millions of nodes have been analyzed.[27] It would be interesting to investigate if some of these dedicated algorithms could be transferred over and applied to protein folding data.

Identifying proper Markov states for molecular systems is considered one of the most important challenges in the Markov state model framework. On the one hand, strict Markov property is the necessary prerequisite for the theory. The states of the system should be related to physically meaningful conformations of the molecules in order to be useful and experimentally verifiable. Currently, these two criteria are met by using chemical intuition as the main tool for building the states. Even though the process sometimes is automated (see, for example, the very recent work[8]), a formally exact, unsupervised approach is highly desirable. Our suggested framework is of this sort: it is mathematically sound and does not require chemical input. We hope it can be used as a basis for efficient practical implementations.

ACKNOWLEDGMENTS

[1] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, J. Comput. Phys. **151**, 146 (1999), http://www.sciencedirect.com/science/article/B6WHY-45N4M0N-7/2/df732f263c903a7f5d544c3b77805635.

[2] W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004), http://pubs.acs.org/doi/pdf/10.1021/jp037421y, http://pubs.acs.org/doi/abs/10.1021/jp037421y.

[3] W. 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), http://pubs.acs.org/doi/pdf/10.1021/jp037422q.

[4] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007), http://link.aip.org/link/?JCP/126/155102/1.

[5] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, J. Chem. Phys. **126**, 155101 (2007), http://link.aip.org/link/?JCP/126/155101/1.

[6] S. Park and V. S. Pande, J. Chem. Phys. **124**, 054118 (2006), http://link.aip.org/link/?JCP/124/054118/1.

[7] J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, Multiscale Model. Simul. **5**, 1214 (2006), http://link.aip.org/link/?MMS/5/1214/1.

[8] G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, J. Chem. Phys. **131**, 124101 (2009), http://link.aip.org/link/?JCP/131/124101/1.

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

[10] C. H. Jensen, D. Nerukh, and R. C. Glen, J. Chem. Phys. **129**, 225102 (2008), http://link.aip.org/link/?JCP/129/225102/1.

[11] D. Nerukh, C. H. Jensen, and R. C. Glen, J. Chem. Phys. **132**, 084104 (2010).

[12] M. Fiedler, Czech. Math. J. **23**, 298 (1973).

[13] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, Linear Algebra Appl. **315**, 39 (2000), citeseer.ist.psu.edu/deuflhard98identification.html.

[14] M. Nilsson Jacobi and O. Görnerup, Adv. Complex Syst. **12**, 1 (2009).

[15] M. Nilsson Jacobi, "A robust spectral method for finding lumpings and meta stable states of non-reversible Markov chains," e-print: arXiv:0810.1127v1.

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

[17] H. J. C. Berendsen, J. Postma, W. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331–342.

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

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

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

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

[22] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, United Kingdom, 2003), available from http://www.inference.phy.cam.ac.uk/mackay/itila/, citeseer.ist.psu.edu/mackay03information.html.

[23] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, MD, 1996).

[24] P. Deuflhard and M. Weber, Linear Algebra Appl. **398**, 161 (2005).

[25] S. Muff and A. Caflisch, J. Chem. Phys. **130**, 125104 (2009).

[26] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004).

[27] Y. Zhang, J. Wang, Y. Wang, and L. Zhou, "Parallel Community Detection on Large Networks with Propinquity Dynamics," Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD'09), Paris, France, June 28–July 1, 2009, pp. 997–1005.