Multiscale Modelling in Molecular Dynamics: Biomolecular Conformations as Metastable States *

Eike Meerbach, Evelyn Dittmer, Illia Horenko, and Christof Schütte

Institut für Mathematik II, Freie Universität Berlin, Arnimallee 2–6, 14195 Berlin, Germany. Tel.: +49-30-838-75367; fax: +49-30-838-75412. Corresponding Author: schuette@math.fu-berlin.de

Summary. We report on a novel approach to the automatic identification of metastable states from long term simulation of complex molecular systems. The new approach is based on a hierarchical concept of metastability: metastable states are understood as subsets of state or configuration space from which the dynamics exits only very rarely; subsets with the smallest exit probabilities are of most interest, their further decomposition then may reveal subsets from which exiting is less but comparably difficult for the system under investigation. The article gives a survey of the theoretical foundation of the approach and its algorithmic realization that generalizes the well-known concept of Hidden Markov Models. The performance of the resulting algorithm are illustrated by application to a 100 ns simulation of penta-alanine with explicit water. We demonstrate the resulting metastable states allow to reveal the conformation dynamics of the moelcule.

Key words: metastable states, molecular conformation, rare transitions, transition probabilities, transfer operator, Hidden Markov Model, penta-alanine, molecular dynamics

1 Introduction

The macroscopic dynamics of typical biomolecular systems is mainly characterized by the existence of biomolecular conformations which can be understood as metastable geometrical large scale structures, i.e., geometries which are persistent for long periods of time. On the longest time scales biomolecular dynamics is a kind of flipping process between these conformations, while on closer inspection it exhibits a rich temporal multiscale structure. Recent research seems to indicate that the conformations with the most pronounced persistency can be understood as metastable or "almost invariant" sets in state

* Supported by the DFG research center Matheon "Mathematics for key technologies" (FZT 86) in Berlin

or configuration space [1, 2]. In other words, the effective or macroscopic dynamics is given by a Markov jump process that hops between the metastable sets while the dynamics within these sets might be mixing on time scales that are smaller than the typical waiting time between the hops. In many applications this Markovian picture is an appropriate description of the dynamics since typical correlation times in the system are sufficiently smaller than the waiting times between hops (and thus much smaller than the timescale the effective description is intended to cover).

While the problem of computing the transition rates or the transition pathways between two *given* conformations attracted a lot of attention recently [3, 4, 5, 6, 7], the problem of efficient algorithmic identification of the most persistent conformations of a given system still is a challenging open problem. Recently there have been several *set-oriented* approaches to this problem [1, 8, 9]. These approaches are based on the construction of a transition matrix that describes transition probabilities between sets in the state space of the system. The identification of metastable sets then is based on analysis of this transition matrix [2, 10]. For higher dimensional systems this always requires coarse graining of the state space into sets (a partition of state space in disjoint sets that avoids the curse of dimensionality) that has to be designed carefully since the resulting metastable sets are unions of the sets from the partition.

Even more recently, alternative approaches have been introduced that apply appropriate Hidden Markov models (HMMs) to the identification problem [11, 12, 13].

We will herein first explain the background of these two types of approach and comment on their relation. In the second part of this contribution we will discuss the identification of the most persistent conformations of penta-alanine as a numerical example. Other examples can be found in [13, 14].

2 Identification of Metastable States

2.1 Dynamics and Statistics

In classical molecular dynamics, a molecular system with a fixed number of N atoms is given by a state vector $(q, p) \in \mathbf{X} = \mathbb{R}^{3N} \times \mathbb{R}^{3N}$, where $q \in \mathbb{R}^{3N}$ denotes the position vector and $p \in \mathbb{R}^{3N}$ the momentum vector. The dynamical behavior, given a specified potential energy function V, a mass matrix M and initial conditions (q_0, p_0), is described by the Newtonian equations

$$\dot{q} = M^{-1}p, \qquad q(0) = q_0, \tag{1}$$

$$\dot{p} = -\nabla_q V(q), \qquad p(0) = p_0. \tag{2}$$

Eq. (1) models an energetically closed system, whose total energy, given by the Hamiltonian

$$H(q,p) \;=\; \frac{1}{2}p^T M^{-1} p \,+\, V(q), \tag{3}$$

is preserved under the dynamics.

It is well known that for every smooth function $\mathcal{F}:\mathbb{R}\to\mathbb{R}$ the probability measure $\mu(\mathrm{d}x) \propto \mathcal{F}(H)(x)\mathrm{d}x$ is invariant wrt. the Markov process X_t given by the solution of the Hamiltonian system (1). The most frequent choice is the canonical density or *canonical ensemble*

$$f(x) \;\propto\; \exp(-\beta H(x)) \tag{4}$$

for some constant $\beta > 0$ that can be interpreted as inverse temperature. The associated measure $\mu(\mathrm{d}x) \propto f(x)\mathrm{d}x$ is called the *canonical measure*. The canonical ensemble is often used in modeling experiments on molecular systems that are performed under the conditions of constant volume and temperature $\mathcal{T} = \frac{1}{k_B\beta}$, where k_B is Boltzmann's constant. Obviously, a single solution of the Hamiltonian system (1) can never be ergodic wrt. the canonical measure, since it conserves the internal energy H, as defined in (3). There are several approaches in the construction of (stochastic) dynamical systems that allow sampling of the canonical ensemble by means of long-term simulation. Most deterministic methods reduce to the construction of a Hamiltonian system in some slightly extended state space $\hat{\mathbf{X}}$, whose projection onto the lower dimensional state space $\mathbf{X}$ of positions and momenta generates a sampling according to (4). One of the most prominent examples is the Nosé-Hoover thermostat [15]. There are also non-deterministic methods. Amongst them, for example, are the well-known Langevin dynamics models, as well as Hybrid Monte Carlo approaches, cf. [2].

2.2 Metastability and the Transfer operator Approach

Each of the optional dynamical models mentioned above involves a *homogeneous Markov process* $X_t = \{X_t\}_{t\in\mathcal{T}}$ in either continuous or discrete time on some state space $\mathbf{X}$. The motion of X_t is given in terms of the *stochastic transition function*

$$p(t,x,A) \;=\; \mathbb{P}[X_{t+s}\in A\,|X_s = x], \tag{5}$$

for every $t,s\in\mathcal{T}$, $x\in\mathbf{X}$ and $A\subset\mathbf{X}$. We write $X_0\sim\mu$, if the Markov process X_t is initially distributed according to the probability measure μ and denote the corresponding probability function of the process by $\mathbb{P}_\mu$. A Markov process X_t admits an *invariant probability measure* μ, or μ is invariant wrt. X_t, if $\int_{\mathbf{X}} p(t,x,A)\mu(\mathrm{d}x) \;=\; \mu(A)$. In the following we always assume that the invariant measure of the process under investigation exists and is unique. A Markov process is called *reversible* wrt. an invariant probability measure μ if $\int_A p(t,x,B)\mu(\mathrm{d}x) \;=\; \int_B p(t,x,A)\mu(\mathrm{d}x)$ for every $t\in\mathcal{T}$ and $A,B\subset\mathbf{X}$.

Metastability of some subset of the state space is characterized by the property that the Markov process is likely to remain within the subset for a long period of time, until it exits and a transition to some other region of the state space occurs. There are in fact several related but different definitions of metastability in literature (see, e.g., [16, 17, 18, 19, 20]); we will focus on the so-called ensemble concept introduced in (6), for a comparison with, e.g., the exit time concept, see [2].

The *transition probability* $p(t, B, C)$ from a subset $B \subset \mathbf{X}$ to another subset $C \subset \mathbf{X}$ within the time span t is defined as the conditional probability

$$p(t, B, C) \;=\; \mathbb{P}_\mu[X_t \in C \;| X_0 \in B] \;=\; \frac{\mathbb{P}_\mu[X_t \in C \text{ and } X_0 \in B]}{\mathbb{P}_\mu[X_0 \in B]}. \qquad (6)$$

This may be rewritten as

$$p(t, B, C) \;=\; \frac{1}{\mu(B)} \int_B p(t, x, C)\, \mu(\mathrm{d}x). \qquad (7)$$

In other words, the transition probability quantifies the dynamical fluctuations within the stationary ensemble μ. A subset $B \subset \mathbf{X}$ is called *metastable* on the time scale $\tau > 0$ if

$$p(\tau, B, B^c) \approx 0, \quad \text{or equivalently,} \quad p(\tau, B, B) \;\approx\; 1,$$

where $B^c = \mathbf{X} \setminus B$ denotes the complement of B.

The objective of the transfer operator approach is an identification of a *decomposition of the state space into metastable subsets* and the corresponding "flipping dynamics" between these sub-states. By a decomposition $\mathrm{d} = \{D_1, \ldots, D_m\}$ of the state space $\mathbf{X}$ we mean a collection of subsets $D_k \subset \mathbf{X}$ with the following properties: (1) positivity $\mu(D_k) > 0$ for every k, (2) disjointness up to null sets, and (3) the covering property $\cup_{k=1}^m \overline{D_k} = \mathbf{X}$.

The *metastability of a decomposition* d is defined as the sum of the metastabilities of its subsets, supposed that the time scale τ of interest is fixed. Then, for each arbitrary decomposition $\mathrm{d}_m = \{D_1, \ldots, D_m\}$ of the state space $\mathbf{X}$ into m sets we define its metastability measure by

$$\mathrm{meta}(\mathrm{d}_m) = \sum_{j=1}^{m} p(\tau, D_j, D_j)/m.$$

For given m the *optimal metastable decomposition* into m sets can then be defined as that decomposition into m sets which maximizes the functional meta. This means in particular that the appropriate number m of metastable subsets must be identified. Both the determination of m and the identification of the metastable subsets can be achieved via spectral analysis of the so-called transfer operator.

Transfer Operator.

The *semigroup of propagators* or forward transfer operators $P^\tau : L^r(\mu) \to L^r(\mu)$ with $t \in \mathcal{T}$ and $1 \leq r < \infty$ as follows:

$$\int_A P^\tau v(y)\ \mu(\mathrm{d}y) \ = \ \int_{\mathbf{X}} v(x) p(\tau, x, A) \mu(\mathrm{d}x) \tag{8}$$

for $A \subset \mathbf{X}$. As a consequence of the invariance of μ, the characteristic function $\mathbf{1_X}$ of the entire state space is an invariant density of P^τ, i.e., $P^\tau \mathbf{1_X} = \mathbf{1_X}$. Furthermore, P^τ is a Markov operator, i.e., P^τ conserves both norm $\|P^\tau v\|_1 = \|v\|_1$ and positivity $P^\tau v \geq 0$ if $v \geq 0$, which is a simple consequence of the definition. Due to (8), the semigroup of propagators mathematically models the evolution of sub–ensembles in time.

The *key idea of the transfer operator approach* wrt. the identification of metastable decompositions can be described as follows:

> Metastable subsets can be detected via eigenvalues of the propagator P^τ close to its maximal eigenvalue $\lambda = 1$; moreover they can be identified by exploiting the corresponding eigenfunctions. In doing so, the number of metastable subsets in the metastable decomposition is equal to the number of eigenvalues close to 1, including $\lambda = 1$ and counting multiplicity.

This strategy was first proposed by Dellnitz and Junge [9] for discrete dynamical systems with weak random perturbations, and has been successfully applied to molecular dynamics in different contexts [21, 1, 22, 2]. The key idea requires the following two *conditions on the propagator* P^τ: (C1) The essential spectral radius of P^τ is less than one, i.e., $r_{\mathrm{ess}}(P^\tau) < 1$. (C2) The eigenvalue $\lambda = 1$ of P is simple and dominant, i.e., $\eta \in \sigma(P^\tau)$ with $|\eta| = 1$ implies $\eta = 1$.

In our algorithmic strategy we furthermore exploit self-adjointness of the propagator which is inherited from reversiblity of the underlying dynamic and results in a real-valued spectrum. Consider for example: (1) high-friction Langevin processes, and (2) (Nose-Hoover) constant temperature molecular dynamics. For both cases the dynamics is reversible and the transfer operator is self-adjoint. For type (1) examples, conditions (C1) and (C2) are known to be satisfied under rather weak condition on the potential [2]. For type (2) examples, it is unknown whether or not the conditions are satisfied; however, it is normally assumed in molecular dynamics that they are valid for realistically complex systems in solution.

The next result [2, 23] justifies the above key idea:

Theorem 1. *Let $P^\tau : L^2(\mu) \to L^2(\mu)$ denote a reversible propagator satisfying (C1) and (C2). Then P^τ is self–adjoint with spectrum of the form*

$$\sigma(P^\tau) \subset [a, b] \cup \{\lambda_m\} \cup \ldots \cup \{\lambda_2\} \cup \{1\}$$

with $-1 < a \leq b < \lambda_m \leq \ldots \leq \lambda_1 = 1$ and λ_i isolated, eigenvalues that are counted according to their finite multiplicities. Denote by $v_m, \ldots, v_1$ the corresponding eigenfunctions, normalized to $\|v_k\|_2 = 1$. Let Q be the orthogonal projection of $L^2(\mu)$ onto $\mathrm{span}\{\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_m}\}$. *Then the metastability of* *an arbitrary decomposition* $\mathrm{d}_m = \{A_1, \ldots, A_m\}$ *of the state space* $\mathbf{X}$ *can be bounded from above by*

$$p(\tau, A_1, A_1) + \ldots + p(\tau, A_m, A_m) \leq 1 + \lambda_2 + \ldots + \lambda_m,$$

while it is bounded from below according to

$$1 + \kappa_2\lambda_2 + \ldots + \kappa_m\lambda_m + c \leq p(\tau, A_1, A_1) + \ldots + p(\tau, A_m, A_m),$$

where $\kappa_j = \|Qv_j\|^2_{L^2(\mu)}$ *and* $c = a\ ((1-\kappa_2) + \ldots + (1-\kappa_n))$.

Theorem 1 highlights the strong relation between a decomposition of the state space into metastable subsets and a *Perron cluster* of dominant eigenvalues close to 1. It states that the metastability of an arbitrary decomposition d_m cannot be larger than $1+\lambda_2+\ldots+\lambda_m$, while it is at least $1+\kappa_2\lambda_2+\ldots+\kappa_m\lambda_m+c$, which is close to the upper bound whenever the dominant eigenfunctions $v_2, \ldots, v_m$ are almost constant on the metastable subsets $A_1, \ldots, A_m$ implying $\kappa_j \approx 1$ and $c \approx 0$. The term c can be interpreted as a correction that is small whenever $a \approx 0$ or $\kappa_j \approx 1$. It is demonstrated in [23] that the lower and upper bounds are sharp and asymptotically exact.

There is an important message contained in the last theorem: *metastability analysis has to be hierarchical.* Whenever we approximate the optimal metastable decomposition d_2 of state space into, say, two sets, we should always be aware that there could be a decomposition d_3 into three sets for which $\mathrm{meta}(\mathrm{d}_3)$ is almost as large as $\mathrm{meta}(\mathrm{d}_2)$. For example, one or both of the two subsets in d_2 could decompose into two or several metastable subsets from which exit is comparably difficult for the system under investigation.

However, whenever there is a gap in the spectrum of the transfer operator after m dominant eigenvalues, then the results of, e.g., [21, 16] tell us that any decomposition into more than m sets will be associated with a significantly larger drop in metastability as measured by the function meta.

Example.

The easiest nontrivial example is a time-discrete Markov chain on a discrete state space. For example, take the chain with state space $S = \{1, 2, 3, 4\}$ and one-step transition probabilities as illustrated in Fig. 1.

This chain has a unique invariant measure, $\mu = (0.25, 0.25, 0.25, 0.25)$, and is reversible. Its transition matrix is given by the following stochastic matrix:

$$\mathcal{T} = \begin{pmatrix} 1-\alpha & \alpha & 0 & 0 \\ \alpha & 1-\alpha-\epsilon & \epsilon & 0 \\ 0 & \epsilon & 1-\beta-\epsilon & \beta \\ 0 & 0 & \beta & \mathbf{1}-\beta \end{pmatrix}.$$

In this specific case $\mathcal{T}$ is the transfer operator of the process.

Let us consider two cases defined in Table 1 below: In case 1, we obviously have a metastable decomposition into the subsets $\{1, 2\}$ and $\{3, 4\}$

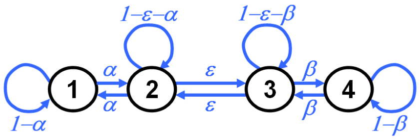

Fig. 1. Markov chain with states 1,2,3,4. The numbers on the arrows linking the states are the one-step transition probabilities. Any transitions that is not represented by an arrow is not allowed.

with metastability measure 0.995 while any further decomposition significantly lowers the metastability measure. Table 2 shows that the spectrum of the transfer operator exhibits a corresponding gap after the first two eigenvalues, and that the upper bound from our theorem, $(\lambda_1 + \lambda_2)/2 = 0.995$, is a very good approximation of meta$(\{1,2\},\{3,4\})$. The second eigenvector $v_2 = (0.51, 0.49, -0.49, -0.50)$ clearly exhibits almost constant levels on the two sets $\{1,2\}$ and $\{3,4\}$ of the metastable decomposition. In case 2, the decreased value of α introduces an additional, milder metastability that separates state 1 from state 2. We can also see this in the spectrum, see Table 2. Still, $\{1,2\}$ and $\{3,4\}$ is the metastable decomposition into two sets with metastability measure 0.995, and associated second eigenvector $v_2 = (0.55, 0.45, -0.49, -0.50)$. But this time, this is the top level of the hierarchy of metastable decompositions only: we can further decompose the set $\{1,2\}$ and the resulting decomposition $\{1\}$, $\{2\}$ and $\{3,4\}$ into three sets also has a significantly high metastability measure of 0.962. For both decompositions, the upper bounds on the metastability measure computed from the eigenvalues are very close to the true values.

case	α	β	ϵ	meta$(\{1,2\},\{3,4\})$	meta$(\{1\},\{2\},\{3,4\})$
1	0.25	0.4	0.01	0.995	0.828
2	0.05	0.4	0.01	0.995	0.962

Table 1. Different parameter sets for the Markov chain considered herein, and metastability measures of the two different decompositions discussed in the text.

case	α	β	ϵ	λ_1	λ_2	λ_3	λ_4
1	0.25	0.4	0.01	1.000	0.990	0.495	0.195
2	0.05	0.4	0.01	1.000	0.991	0.895	0.195

Table 2. Different parameter sets for the Markov chain considered herein, and spectrum of the associated transfer operators.

2.3 Discretization and PCCA

In the typical case the dynamical process under investigation lives on a continuous state space such that the transfer operator does not have the form of a nice stochastic matrix. Therefore, discretization of the transfer operator is needed to yield a stochastic matrix with which one can proceed as in the example above.

Let $\chi = \{\chi_1, \ldots, \chi_n\} \subset L^2(\mu)$ denote a set of *non–negative* functions that are a partition of unity, i.e., $\sum_{k=1}^n \chi_k = \mathbf{1}_\mathbf{X}$. The *Galerkin projection* $\Pi_n : L^2(\mu) \to \mathcal{S}_n$ onto the associated finite dimensional ansatz space $\mathcal{S}_n = \mathrm{span}\{\chi_1, \ldots, \chi_n\}$ is defined by

$$\Pi_n v \;=\; \sum_{k=1}^{n} \frac{\langle v, \chi_k\rangle_\mu}{\langle \chi_k, \chi_k\rangle_\mu}\, \chi_k .$$

Application of the Galerkin projection to $P^\tau v = \lambda v$ yields an eigenvalue problem for the discretized propagator $\Pi_n P^\tau \Pi_n$ acting on the finite-dimensional space $\mathcal{S}_n$. The matrix representation of this finite dimensional operator is given by the $n \times n$ *stochastic transition matrix* $\mathcal{T} = (\mathcal{T}_{kl})$, whose entries are given by

$$\mathcal{T}_{kl} \;=\; \frac{\langle P^\tau \chi_k, \chi_l\rangle_\mu}{\langle \chi_k, \chi_k\rangle_\mu}. \tag{9}$$

The transition matrix inherits the main properties of the transfer operator: it is a stochastic matrix with invariant measure given by the invariant measure μ of P^τ, it is reversible if P^τ is self-adjoint, and (if the discretization is fine enough) it also exhibits a Perron cluster of eigenvalues that approximates the corresponding Perron cluster of P^τ, and with eigenvectors that approximate the dominant eigenvectors of P^τ [2]. It thus allows to compute the metastable sets of interest by computation of the dominant eigenvectors of $\mathcal{T}$ and by realization of the identification strategy of page 5 based on these (discrete) eigenvectors. This has led to the construction of an aggregation technique called "Perron Cluster Cluster Analysis" (PCCA) [8, 10, 24].

If $x_0, \ldots, x_N$ denote a time series obtained from a realization of the Markov process with time stepping τ, then the entries of $\mathcal{T}$ can be approximated from the relative transition rates computed by means of this time series:

$$\mathcal{T}_{kl} \;\approx\; \mathcal{T}_{kl}^{(N)} \;=\; \frac{\sum_{j=1}^{N} \chi_k(x_j)\cdot \chi_l(x_{j+1})}{\sum_{j=1}^{N} \chi_k(x_j)}. \tag{10}$$

Although it looks extremely simple, using equation (10) algorithmically may become problematic. There are two main reasons for potential difficulties.

Trapping problem.

The *rate of convergence* of $\mathcal{T}_{kl}^{(N)} \to \mathcal{T}_{kl}$ depends on the smoothness of the partition functions χ_k as well as on the mixing properties of the Markov process [25]. The latter property is crucial here: The convergence is geometric with a rate constant $\lambda_1 - \lambda_2 = 1 - \lambda_2$ where λ_2 denotes the second largest eigenvalue (in modulus). In the case of metastability with λ_2 being very close to $\lambda_1 = 1$, we will have dramatically slow convergence. However we will *not* go into the depth of the discussion on overcoming the trapping problem, but instead assume in all of the following that we have already generated or can directly generate a time series that is “long enough” in the sense that it contains statistically significant information about more than one –if not all– interesting metastable states of the system under consideration. The interested reader may refer to the vast literature [26, 27, 28].

Curse of Dimension.

Any discretization of the transfer operator will suffer from the curse of dimension whenever it is based on a uniform partition of all of the hundreds or thousands of degrees of freedom in a typical biomolecular system. Fortunately, chemical observations reveal that—even for larger biomolecules—the curse of dimensionality can be circumvented by exploiting the hierarchical structure of the dynamical and statistical properties of biomolecular systems: only relatively few *essential degrees of freedom* may be needed to describe the conformational transitions; furthermore, the canonical density has a rich spatial multiscale structure induced by the rich structure of the potential energy landscape, which again underlines the necessity of a hierarchical approach.

2.4 Relaxations, Transitions, and Effective Dynamics

Assume that we successfully identified a metastable decomposition into the sets $D_1, \dots, D_m$ for given lag time τ. Due to our above results the dynamics is jumping from sets D_k to set D_j with probability $p(\tau, D_k, D_j)$ during time τ. Then, it is an intriguing idea to describe the effective dynamics of the system by means of the Markov chain with discrete states $D_1, \dots, D_m$ and transition matrix $P = (P_{kj})$ with $P_{kj} = p(\tau, D_k, D_j)$. This ”effective dynamics” is Markovian and thus cannot take into account that there may be memory in the system that is much longer than the time span τ used to compute the metastable decomposition.

In order to categorize thimetastable states, molecular conformation, rare transitions, transition probabilities, transfer operator, Hidden Markov Model, penta-alanine, molecular dynamicss more precisely, let us denote the typical (mean) exit time from D_j to D_k by T_{jk}, and the typical relaxation timescale within D_j by τ_j (that is, when the system enters D_j at $t = 0$ it has lost almost all its memory at $t = \tau_j$).

The simplest case is that we have τ being comparable to the largest τ_j and $\tau_j \ll \min_k T_{jk}$ for all $j = 1, \dots, m$. Then the above construction is a good model of the effective dynamics, and the system on average samples its restricted invariant density $\mu|_{D_j}$ in D_j before exiting from D_j.

The more complicated case is that still $\tau \ll \min_k T_{jk}$ but no longer $\tau_j \ll \min_k T_{jk}$ for all j. Then the above construction constitutes a misleading model for the effective dynamics.

Comparable problems appear for other types of coarse graining. In the above, the coarse graining is given by the metastable decomposition. But we could also try to realize the transfer operator approach, by means of discretizing only a subspace instead of the full state space. For example the subspace spanned by the essential degrees of freedom, or the torsion angles space. Then, what is assumed to be a Markov process in state space not longer might be Markovian on the chosen subspace.

2.5 The Hidden Markov Model Approach

We are now going to consider the case of a given timeseries $(O_t)_{t=t_1,\ldots,t_N}$, with constant sampling time $\tau = t_{j+1} - t_j$. Here, the O_t do not necessarily denote the state of the molecule at time t but rather some low-dimensional observable, for example, some or all torsion angles or the set of essential degrees of freedom (if this should be available). We assume that there is an unknown metastable decomposition, say, into m sets $D_1, \ldots, D_m$. Using the notation introduced above, we furthermore assume $\tau \ll \min_k T_{jk}$ but do not specify the relation between τ and the τ_j. We then can premise that, at any time t, the system is in one of the metastable states D_{j_t} to which we simply refer by j_t in the following. However, the time series (j_t) is *hidden*, i.e., not known in advance, while the series (O_t) is called the output series or the *observed* sequence.

This design can be represented by a Hidden Markov Model (HMM). A HMM abstractly consists of two related stochastic processes: a hidden process j, that fulfills the Markov property and an observed process O_t that depends on the state of the hidden process j_t at time t. A HMM is fully specified by the initial distribution π, the rate matrix R of the hidden Markov process j, as well as by the law that governs the observable O_t depending on the respective hidden state j_t.

In the standard versions of HMMs the observables are i.i.d. random variables with stationary distributions that depend on the respective hidden states [13]. Within the scope of molecular dynamics this means, that one considers the simple case where τ is comparable to the τ_j and $\tau_j \ll \min_k T_{jk}$, i.e., the process samples the restricted invariant density before exiting from a metastable state, and the sampling time of the time series is long enough to assume statistical independence between steps. Nevertheless, if this is not the case, only a slight modification of the model structure is required to represent the relaxation behavior: Instead of i.i.d. random variables one can use an Ornstein-Uhlenbeck (OU) process as a model for the output behavior in each hidden state. The HMM then gets the form [11]:

$$\dot{Y}_t = -DV^{(q)}(Y_t) + \sigma^{(j_t)}\dot{W}_t, \tag{11}$$

$$j_t \ : \ \mathbf{R}^1 \to \{1, 2, ..., m\}, \tag{12}$$

where j_t are the realizations of the hidden Markov process with discrete state space, W_t is standard "white noise", and $\{V^{(j)}, \sigma^{(j)}\}$ is a set of the state-specific model parameters with harmonic potentials $V^{(j)}$ of the form

$$V^{(j)}(Y) = \frac{1}{2}(Y - \mu^{(j)})^T D^{(j)}(Y - \mu^{(j)}) + V_0^{(j)}, \tag{13}$$

where $\mu^{(j)}$ and $D^{(j)}$ are *equilibrium position* and Hesse-matrix of the OU process within conformation j. This process is therefore given by the parameters $\Theta^{(j)} = (\mu^{(j)}, D^{(j)}, \sigma^{(j)})$. Since the output process is specified by a stochastic differential equation we will refer to this model modification as HmmSde . Its entire parameter set will be denoted $\Theta = (\Theta^{(1)}, \dots, \Theta^{(m)}, R)$ in the following, where R denotes the rate matrix of the Markov chain in (12).

Assume for a moment that the hidden state is fixed, i.e. $j_t = j$. Then, the evolution of probability density $\rho(t, Y|j)$ under the dynamics given by (11) can be obtained as solution of the corresponding Fokker Planck equation:

$$\partial_t \rho = \triangle_Y V^{(j)}(Y)\rho + \nabla_Y V^{(j)}(x) \cdot \nabla_Y \rho + \nabla_Y \cdot B^{(j)} \nabla_Y \rho, \quad \rho(t=0, Y|j) = \rho_0(Y|j) \tag{14}$$

where $B^{(j)} = (\sigma^{(j)})^2 \in \mathbf{R}^1$ denotes the variance of the white noise (for $\mathbf{R}^d$ it is a positive definite selfadjoint matrix). In the subsequent we will denote the partial differential operator on the RHS of (14) by $L_{\Theta^{(j)}}$. Then, the solution of (14) can be written as $\rho(t, Y|j) = (\exp(tL_{\Theta^{(j)}})\rho_0)(Y)$.

We have the following aim: For a given observed sequence $O = (O_t)$ determine those parameters in the dynamical model (11&12) for which the *probability that O is an output of these parameters is maximal.* To this end, one has to know the probability $p(O, j|\Theta)$ of the observed sequence O and a specific hidden sequence j for given parameters Θ. First assume that the dynamics (11) yields output $Y_1 = O_{t_{k+1}}$ at time $t = t_k + \tau = t_{k+1}$ after starting at $Y_0 = O_{t_k}$ while remaining in $j = j_{t_k}$ during the evolution form t_k to t_{k+1}. Due to the above this probability obviously is

$$\rho(O_{t_{k+1}}|j, O_{t_k}) = \Big(\exp(\tau L_{\Theta^{(j)}})\delta_{O_{t_k}}\Big)(O_{t_{k+1}}), \tag{15}$$

where δ_O denotes the Dirac measure supported at state O. With this the total probability that the observed sequence O and a given hidden sequence j is an output of (11), parametrized by Θ, is

$$p(O, j|\Theta) = \pi(j_{t_0})\nu(O_{t_0}|j_{t_0}) \prod_{k=1}^{N-1} \mathcal{T}(j_{t_k}, j_{t_{k+1}})\rho(O_{t_{k+1}}|j_{t_k}, O_{t_k}), \tag{16}$$

where $\mathcal{T} = \exp(\tau R)$ denotes the transition matrix of the Markov jump process, specified in (12), in time τ and π and ν are initial distributions that have to chosen in addition.

In putting everything together, we have to face these algorithmic problems: (1) determine the optimal parameters Θ by maximizing the probability

$p(O,j|\Theta)$ —this is a nonlinear global optimization problem—, (2) determine the optimal sequence of hidden metastable states $j=(j_t)$ for given optimal parameters, and (3) determine the number of important metastable states; which we, up to now, simply assumed to be identical with the number of hidden states.

The above formulation of the two first problems seems to contain a considerable contradiction: How can we determine optimal parameters without knowing the optimal hidden sequence? Fortunately the solution is already available from the standard HMM framework: The parameter optimization is carried out by the Expectation Maximization (EM) algorithm that *iteratively* determines the optimal parameters Θ_* via maximizing the expectation

$$Q(\Theta;\Theta_k) \;=\; \mathbb{E}\left(\log p(O,j|\Theta) \;|\; O,\Theta_k \right) \tag{17}$$

of the complete probability $p(O,q|\lambda)$ wrt. the hidden sequence j given the observation sequence and the current parameter estimate Θ_k. It is a classical result [29] (Chap. 4.2) that this can be rewritten as a sum over all hidden sequences:

$$Q(\Theta;\Theta_k) = \sum_{j=(j_t)} p(O,j|\Theta_k)\log\left(p(O,j|\Theta)\right). \tag{18}$$

The expectation-step of the EM algorithm evaluates the expectation value Q based on the given parameter estimate Θ_k, while the maximization-step determines the refined parameter set $\Theta_{k+1} = \mathrm{argmax}_{\Theta}\, Q(\Theta;\Theta_k)$. The expectation step is standard but the maximization step can also be realized algorithmically, see [11] or [12] for different realizations.

For the identification of the optimal sequence of hidden metastable states we can use the well-known Viterbi algorithm [30], which exploits dynamic programming techniques to resolve in a recursive manner the optimization problem

$$j_* = \mathrm{argmax}\, p(O,j|\Theta_*).$$

The obtained optimal sequence j_* is called "Viterbi path". For technical details see [11].

The parameter fitting step requires the specification of the number of hidden states, which, whenever the hidden states should be metastable states, is in general not apriori known. One policy to overcome this problem is to assume a sufficient large number of hidden states, perform the parameter fitting and conduct a further aggregation of the resulting transition matrix. This can be done by Perron cluster cluster analysis (PCCA), e.g., by the spectral properties of the resulting transition matrix $\mathcal{T}$ as proposed in the transfer operator approach (we will illustrate this procedure on an example in the next section), see [11] for details.

As the numerical effort of the used algorithmic process scales wrt. to the dimension of the observable d as $\mathcal{O}(d^3)$ the three steps above provide a tool for metastability analysis only on low-dimensional observables. But this obstacle can be circumvented by first applying the algorithm *separately* to several low-dimensional projections. We end up with an aggregated Viterbi path j_* for each projection (for example, when the observed sequence contains all peptide angles along a polypeptide chain, we could apply HmmSde to the peptide angle of all single nucleotides of the sequence first). Then, simple combinatorics allows to combine these Viterbi paths into a "combined" Viterbi path, that contains every occurred state combination. A transition matrix for the full-dimensional system can be obtained by counting relative frequencies in the combined Viterbi path and another aggregation via PCCA finally allows for the identification of metastable sets.

We want to underline two important points on this algorithmic scheme: (1) The philosophy of HMM models gives a justification to work on low-dimensional projections of an observable, because the observable is not meant to specify the occupied metastable state at a certain time, but to reflect different states by different dynamical behavior which seems quite reasonable. (2) We not only obtain an optimal sequence of hidden metastable states, but also optimal parameters for a simple, but physically motivated, reduced model (11&12) of the dynamics. We think that the extraction of such models is quite important to gain more insight in the mechanisms behind metastable behavior.

3 Numerical Example: Penta-Alanine

As illustration we demonstrate the performance of the proposed algorithmic procedure in application to the analysis of a peptide molecule. The global (secondary) structure of a peptide is determined by the so-called peptide angles. For each alanine amino acid we have to consider two of these backbone torsion angles. These peptide angles pairs can not take arbitrary values due to steric interaction, but will adopt values in definite regions, belonging to various secondary structures. In Fig. 2 the backbone torsion angles of penta-alanine are shown. As usual the pair of angles belonging to the same amino acid residue is labelled by Φ and Ψ. Illustration of quantities as functions of Φ and Ψ are called Ramachandran plots. The Ramachandran plot in Fig. 2 exhibits the values that a pair of peptide angles typically takes in certain secondary structures.

Our analysis is based upon a time series of the 10 backbone torsion angles of penta-alanine, extracted from the long time simulation that has already been discussed in [31] (courtesy of G. Stock, Frankfurt). The simulation was done in explicit water using a thermostat of 300K over an interval of 100ns, while the coordinates were written out every 0.1ps, resulting in 1000000 data points. Fig. 10 below shows a histogram Ramachandran plot of the entire time series for each $\Phi\backslash\Psi$ pair. It reveals that the fifth angle pair has a substantial different behavior than the other pairs. This will not be our concern here since we just take the timeseries for demonstration of the algorithmic procedure of extracting information about metastable states from a given time series.

As penta-alanine is a short peptide it will not have a stable β-sheet conformation, but as we will see in our analysis it exhibits a stable α-helix conformation and several other conformations which can be characterized by certain flexibility patterns.

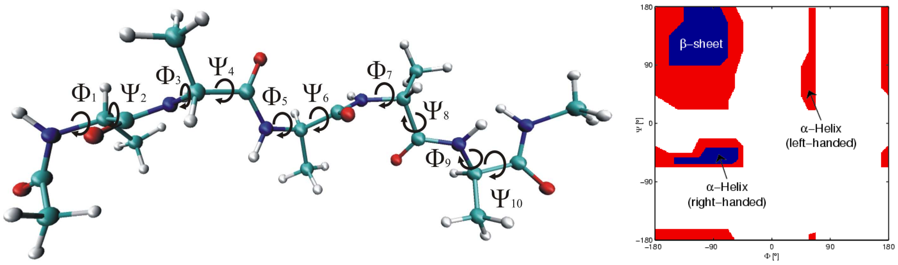

Fig. 2. *Left:* The Penta-alanine peptide in ball-and-stick representation. The ten peptide angles determining the secondary structure are marked by $\Phi_1, \Psi_2, \ldots, \Phi_9, \Psi_{10}$. *Right:* Ramachandran plot, showing the energetically preferred regions of a $\Phi\backslash\Psi$ pair with the associated secondary structures (simplefied plot due to [32]).

The first step of our analysis consists in analyzing each of the peptide angles *separately* by using the HMMSDE techniques to determine the Viterbi path for each of the angles. Beforehand an initial guess must specify the number of hidden states, which we set to 4. As result each peptide angle is represented by a discrete time series with 4 states, the Viterbi path of this angle, assigning each instance in the time series to a state of a hidden Markov process. Fig. 3 provides an example. Note the fundamental difference to a direct transfer operator approach, where we have to specify a box discretization of the ten dimensional state space.

This way we obtain 10 Viterbi paths. In the next step these are now combined to pairwise Viterbi paths by simple superposition of the Viterbi paths which belong to a $\Phi\backslash\Psi$ combination. This produces five *pairwise* Viterbi paths, each with 16 states. Each of these Viterbi paths now is understood as the output time series of a Markov process with discrete state space. For each pair the corresponding stochastic transition matrix can be computed by counting the transitions between different states. Following the transfer operator approach the spectra of these transition matrices contain information about metastability in the dynamics of each peptide angle pair. For example the eigenvalues of the transition matrix $\mathcal{T}_4$, extracted from the fourth pairwise Viterbi path (time lag $\tau = 0.1ps$), are

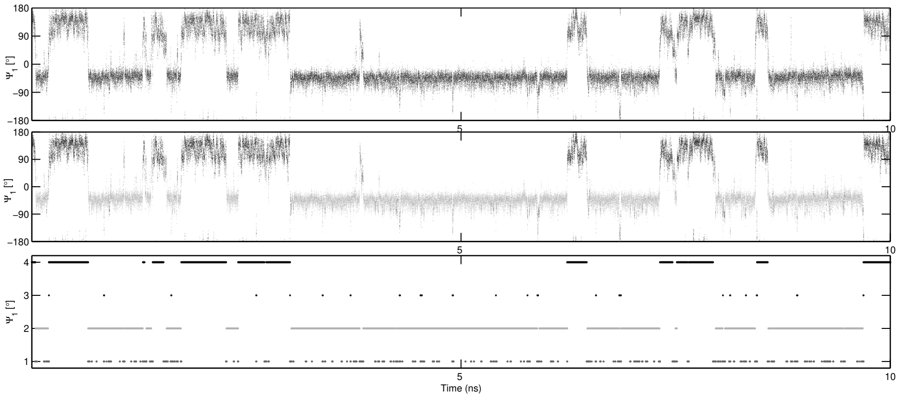

Fig. 3. Results of HMMSDE for the Ψ_2 time series. *Top:* Illustration of the first 10ns of the Ψ_2 time series. *Middle:* The same picture colored according to the association with the assumed 4 hidden states. *Bottom:* The Viterbi path displayed as a discrete time series which specifies only the hidden states.

$\lambda_1(T_4)$	$\lambda_2(T_4)$	$\lambda_3(T_4)$	$\lambda_4(T_4)$	$\lambda_5(T_4)$	$\lambda_6(T_4)$	$\lambda_7(T_4)$	$\lambda_8(T_4)$
1	0.998	0.994	0.991	0.979	0.975	0.959	0.947

$\lambda_9(T_4)$	$\lambda_{10}(T_4)$	$\lambda_{11}(T_4)$	$\lambda_{12}(T_4)$	$\lambda_{13}(T_4)$	$\lambda_{14}(T_4)$	$\lambda_{15}(T_4)$	$\lambda_{16}(T_4)$
0.9337	0.898	0.8798	0.839	0.176	0.154	0.139	-0.001

,

which suggests four metastable subsets for this pairwise Viterbi path. Obviously other interpretations of the spectrum are also reasonable, as one could argue that the clearest gap occurs after the twelfth eigenvalue. This is more or less a decision of how much detail one wants to or can afford to preserve at this stage of analysis, but it turns out that if we turn to the global analysis these details are filtered out anyway.

The next step is to cluster all pairwise Viterbi paths according to the structure of the eigenvectors belonging to the dominant eigenvalues of the associated transition matrix. This yields five *clustered* pairwise Viterbi paths with 4 or 5 states each, allocating each instance in the corresponding $\Phi\backslash\Psi$ pair time series to a *metastable* hidden set. Plotting this information in the form of a Ramachandran plot reveals a similar, although not equal, structure for each $\Phi\backslash\Psi$ pair, see Fig. 4. Note again that the different metastable sets are not disjoint sets in the $\Phi\backslash\Psi$ plane, as the HMMSDE analysis assumes the given data to be a projection of some hidden full process.

By another superposition of the five clustered pairwise Viterbi paths a *global* Viterbi path is obtained. This path contains 1114 different states, due to the fact that the states of the clustered pairwise Viterbi paths can be combined in any way giving a theoretical maximum of $5 \cdot 4 \cdot 5 \cdot 4 \cdot 4 = 1600$ possible global states (of which only 1114 actually occur). Setting up the transition matrix again yields a sparse stochastic matrix in which more than 99% of the entries are equal to zero.

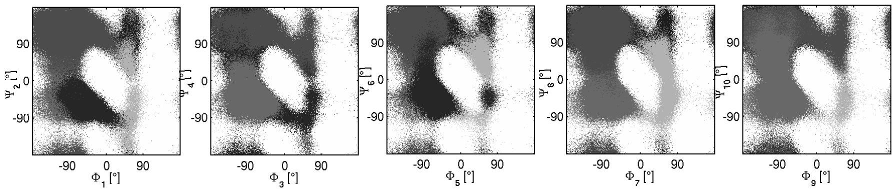

Fig. 4. Ramachandran plots of the five $\Phi\backslash\Psi$ pairs colored according to the associated metastable sets. The numbers of metastable sets differs for the different $\Phi\backslash\Psi$ pairs (it is from left to right 5,4,5,4,4).

It is instructive to compare the eigenvalues of transition matrices obtained for different lag times τ. That is, we do not count transitions on a timescale of $0.1ps$ which means to observe transitions from one instance of the time series to the next, but count transitions on a timescale of, say, $1ps$ which is between every tenth step in the global Viterbi path. For all time lags, Fig. 5 clearly indicates two dominant eigenvalues after which we find a gap, followed by other gaps after 4, 9, or 16 eigenvalues. This yields 2, respectively 4, 9, or 16 metastable sets. To avoid confusion we call these metastable sets (molecular) conformations.

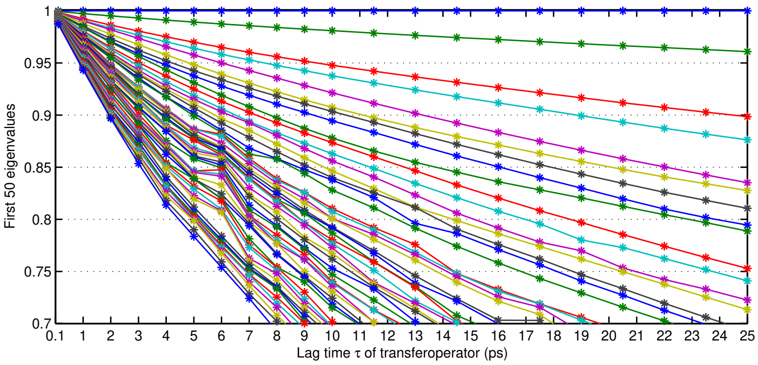

Fig. 5. Illustration of the 50 largest eigenvalues of the transition matrix obtained from the global Viterbi path vs. lag time τ. (A time step of 10ps means that transition are considered that occur from time x to time $x + 10ps$). On clearly observes that the structure of the spectrum does not depend on τ.

To gain more insight into the metastability analysis we will now compare the results of the procedure based on the first two eigenvectors with the results based on the first four ones. The outcome will identify 2, resp. 4, conformations in the discrete global Viterbi path. This allows to associate each data point of the ten dimensional peptide angle time series with one of the 2, resp. 4, conformations.

In Fig. 6 the results based on the first two eigenvectors are displayed ($M = 2$). Each of the two resulting conformations is represented by five histogram plots that belong to the five $\Phi\backslash\Psi$ pairs. These histograms are based on the assignment of data points to the conformation. Comparing the positions of the histogram peaks in the first conformation with the classifications given in Fig. 2 shows that this conformation corresponds to an α-helix structure. In contrast, the other conformation allows no assignment to a specific secondary structure, as every angle pair is very flexible and adopts regions of the α-helix structure and the β-strand structure.

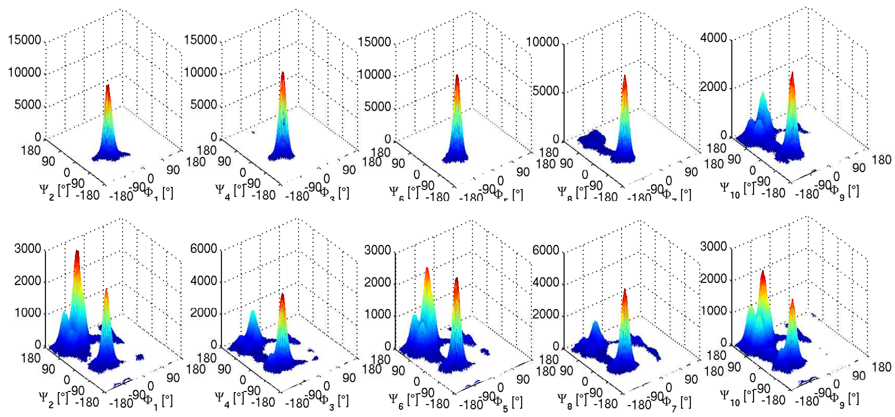

Fig. 6. Histogram plots for the two conformations (as resulting from the $M = 2$ analysis) displayed in the Ramachandran plane of each of the five $\Phi\backslash\Psi$ pairs. *Top row:* This conformation constitutes a clear helix structure. *Bottom row:* This conformation corresponds to no clear secondary structure.

Redoing the analysis with the $M = 4$ leading eigenvectors yields that the α-helix structure is still identified as a conformation, while the other conformation of the previous analysis splits-up into three conformations. Each of these three conformations can be uniquely described by the dynamical behavior of the peptide angle pairs, some of them are fixed to α-helix, resp. β-strand regions, while others remaining flexible in the sense that they alternate between these regions, see Fig. 7.

Taking more leading eigenvectors into account would resolve more flexible angle pairs by separating α-helix and β-strand parts, but it is important to note that this means resolving metastability on a faster timescale, cf. Fig. 5. As illustration we show representatives of each conformation $C_1, \ldots, C_4$ resulting from the $M = 4$ analysis in Fig. 8, This figure also includes the conditional transition probabilities $p(\tau, C_j, C_k)$ of the Markov switching process for lag time $\tau = 0.1$ps. Note that in accordance with our definition of metastability

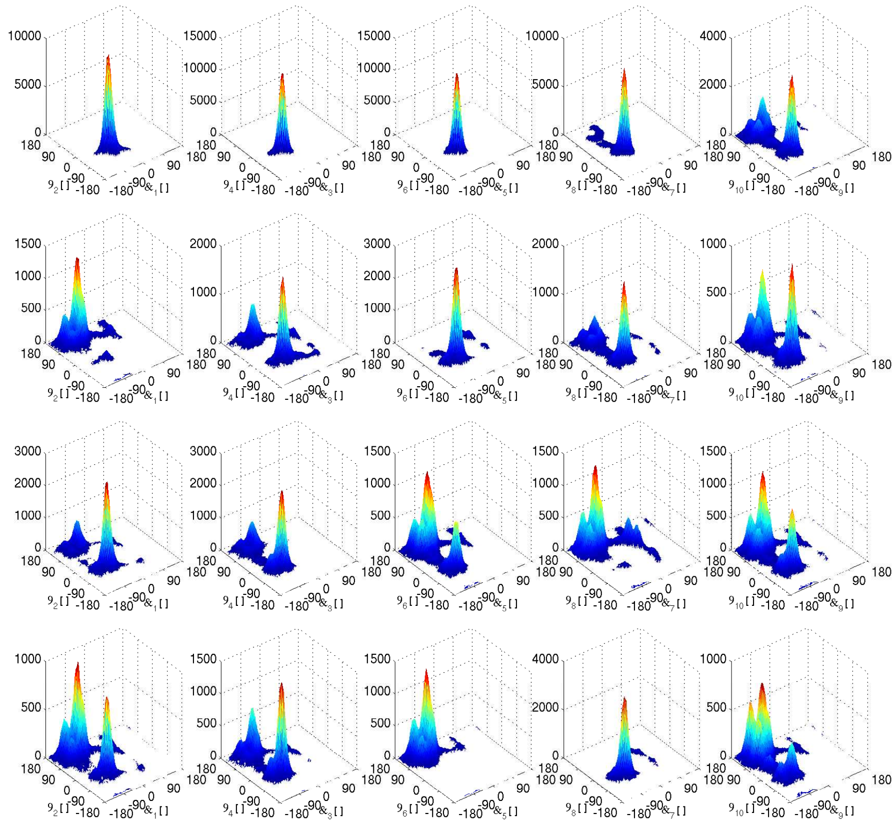

Fig. 7. Conformations as resulting from $M = 4$ analysis represented as histogram plots displayed in the Ramachandran plane of each of the five $\Phi\backslash\Psi$ pairs. *Top:* The helix structure of the first conformation is unchanged compared to the $M = 2$ analysis. *Below:* The other conformations exhibit mixed structures, with some angle pairs fixed to α-helix or β-strand regions while others are flexible.

the conditional probability to stay within one conformation is nearly 1 for each conformation.

Finally we want to try to verify the results of the HmmSde based procedure by comparison with a direct transfer operator based analysis. We will do this by reducing the dimensionality of the system considerable by noting that the relevant dynamical information is contained in the first four Ψ angles. Therefore we can reduce the ten-dimensional to a four-dimensional peptide angle space by skipping the other dimensions. The four-dimensional space can be partitioned directly by discretizing each dimension in 10 equidistant boxes. This yields $10^4 = 10000$ discretization boxes from which 6551 have been visited by the time series under consideration. Computing the associated transition matrix and evaluating the dominant spectrum is easily feasible,

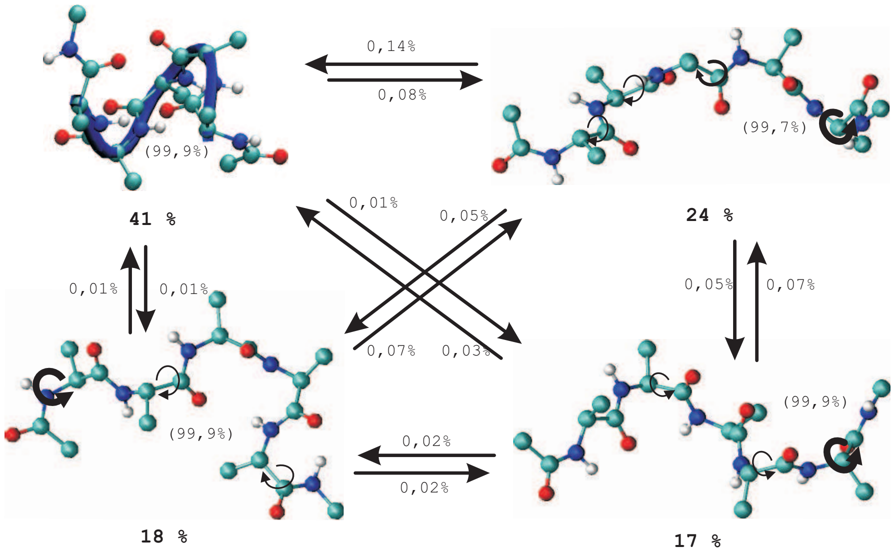

Fig. 8. Representatives of the four conformations obtained in the $M = 4$ analysis and the conditional transition probabilities between them (lag time $\tau = 0.1$ps). Fat numbers indicating the statistical weight of each conformation, numbers in brackets the conditional probability to stay within a conformation. Flexibility in peptide angles is marked with arrows, cf. Fig. 7. *Top left:* For the helix conformation the backbone is colored blue for illustrative purpose. It should be obvious from Fig. 5 that for significantly larger lag time τ only two eigenvalues will correspond to metastability such that only the helical conformation and a mixed flexible and partially unfolded one remain with significantly high conditional probability to stay within.

particular as it is a sparse matrix. The results are obtained by analyzing this transition matrix, based upon a direct partition of the state space, are similar to the results we obtained with HmmSde . Without giving details, we indicate this by showing the eigenvalues plotted against different lag times τ in Fig. 9, which reveals a very similar spectral structure as Fig. 5.

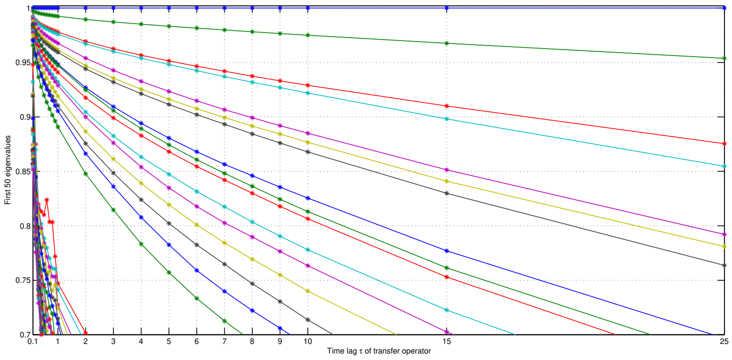

Fig. 9. The 50 largest eigenvalues of the transition matrix obtained from a direct discretization of the $\Psi_2, \Psi_4, \Psi_6, \Psi_8$-subspace versus the lag time τ.

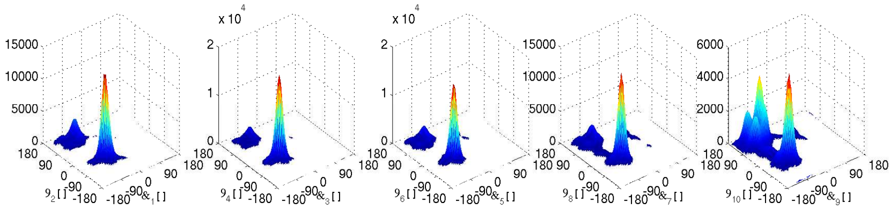

Fig. 10. Pairwise Ramachandran plots of the histogram of the entire time series.

Acknowledgements

We are indebted to G. Stock for the courtesy of making the simulation data of penta-alanine available to us.

References

1. C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, "A direct approach to conformational dynamics based on hybrid Monte Carlo," *J. Comput. Phys., Special Issue on Computational Biophysics*, vol. 151, pp. 146–168, 1999.
2. C. Schütte and W. Huisinga, "Biomolecular conformations can be identified as metastable sets of molecular dynamics," in *Handbook of Numerical Analysis* (P. G. Ciaret and J.-L. Lions, eds.), vol. Computational Chemistry, North-Holland, 2003.
3. D. Chandler, "Finding transition pathways: Throwing ropes over rough montain passes, in the dark," in *Classical and Quantum Dynamics in Condensed Phase Simulations* (B. Berne, G. Ciccotti, and D. Coker, eds.), pp. 51–66, Singapure: World Scientific, 1998.
4. W. E, W. Ren, and E. Vanden-Eijnden, "String method for the study of rare events," *Phys. Rev. B*, vol. 66, no. 052301, 2002.

5. W. E, W. Ren, and E. Vanden-Eijnden, "Finite temperature string method for the study of rare events," *J. Phys. Chem. B*, 2005.
6. A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of the United States of America*, 2002.
7. R. Elber and M. Karplus, "Multiple conformational states of proteins: A molecular dynamics analysis of Myoglobin," *Science*, vol. 235, pp. 318–321, 1987.
8. P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, "Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains," *Lin. Alg. Appl.*, vol. 315, pp. 39–59, 2000.
9. M. Dellnitz and O. Junge, "On the approximation of complicated dynamical behavior," *SIAM J. Num. Anal.*, vol. 36, no. 2, pp. 491–515, 1999.
10. P. Deuflhard and M. Weber, "Robust Perron cluster analysis in conformation dynamics." ZIB-Report 03-19, Zuse Institute Berlin, 2003.
11. I. Horenko, E. Dittmer, A. Fischer, and C. Schütte, "Automated model reduction for complex systems exhibiting metastability." Submitted to Multiscale Modeling and Simulation.
12. I. Horenko, E. Dittmer, and C. Schütte, "Reduced stochastic models for complex molecular systems," *Computing and Visualization in Science*, 2005.
13. A. Fischer, S. Waldhausen, I. Horenko, E. Meerbach, and C. Schütte, "Identification of biomolecular conformations from incomplete torsion angle observations by hidden Markov models," *Journal of computational Physics*, 2004.
14. I. Horenko, E. Dittmer, F. Lankas, J. Maddocks, P. Metzner, and C. Schütte, "Macroscopic dynamics of complex metastable systems: Theory, algorithms, and application to b-dna," *J. Appl. Dyn. Syst., submitted*, 2005.
15. S. D. Bond and B. B. L. Benedict J. Leimkuhler, "The Nosé–Poincaré method for constant temperature molecular dynamics," *JCP*, vol. 151, no. 1, pp. 114–134, 1999.
16. A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, "Metastability in stochastic dynamics of disordered mean–field models," *Probab. Theor. Rel. Fields*, no. 119, pp. 99–161, 2001.
17. E. B. Davies, "Metastable states of symmetric Markov semigroups I," *Proc. London Math. Soc.*, vol. 45, no. 3, pp. 133–150, 1982.
18. C. Schütte, W. Huisinga, and P. Deuflhard, "Transfer operator approach to conformational dynamics in biomolecular systems," in *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems* (B. Fiedler, ed.), pp. 191–223, Springer, 2001.
19. G. Singleton, "Asymptotically exact estimates for metatstable Markov semigroups," *Quart. J. Math. Oxford*, vol. 35, no. 2, pp. 321–329, 1984.
20. W. E and E. Vanden-Eijnden, "Metastability, conformation dynamics, and transition pathways in complex systems," *preprint*, 2005.
21. W. Huisinga, S. Meyn, and C. Schütte, "Phase transitions and metastability in Markovian and molecular systems," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 419–458, 2004.
22. C. Schütte and W. Huisinga, "On conformational dynamics induced by Langevin processes," in *EQUADIFF 99 - International Conference on Differential Equations* (B. Fiedler, K. Gröger, and J. Sprekels, eds.), vol. 2, (Singapore), pp. 1247–1262, World Scientific, 2000.
23. W. Huisinga and B. Schmidt, "Metastability and dominant eigenvalues of transfer operators," in *New Algorithms for Macromolecular Simulation* (C. Chipot,

R. Elber, A. Laaksonen, B. Leimkuhler, A. Mark, T. Schlick, C. Schütte, and R. Skeel, eds.), vol. 49 of *Lecture Notes in Computational Science and Engineering*, Springer, to appear 2005.
24. M. Weber, "Improved Perron cluster analysis." ZIB-Report 03-04, Zuse Institute Berlin, 2004.
25. P. Lezaud, "Chernoff and Berry–Esséen inequalities for Markov processes," *ESIAM: P & S*, vol. 5, pp. 183–201, 2001.
26. B. J. Berne and J. E. Straub, "Novel methods of sampling phase space in the simulation of biological systems," *Curr. Opinion in Struct. Biol.*, vol. 7, pp. 181–189, 1997.
27. D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar, eds., *Monte Carlo Methods in Chemical Physics*, vol. 105 of *Advances in Chemical Physics*. New York: Wiley, 1999.
28. A. Fischer, C. Schütte, P. Deuflhard, and F. Cordes, "Hierarchical uncoupling-coupling of metastable conformations," in *Computational Methods for Macromolecules: Challenges and Applications* (T. Schlick and H. H. Gan, eds.), vol. 24 of *Lecture Notes in Computational Science and Engineering*, pp. 235–259, Springer, 2002.
29. J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and Hidden Markov Models," tech. rep., International Computer Science Institute, Berkeley, 1998.
30. A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Informat. Theory*, vol. IT-13, pp. 260–269, 1967.
31. Y. Mu, P. H. Nguyen, and G. Stock, "Energy landscape of a small peptide revealed by dihedral angle principal component analysis," *Proteins: Structure, Function, and Bioinformatics*, vol. 58, no. 1, pp. 45–52, 2004.
32. G. N. Ramachandran and V. Sasiskharan, "Conformations of polypeptides and proteins," *Advan. Prot. Chem.*, vol. 23, pp. 283–427, 1968.