Identification of Biomolecular Conformations from Incomplete Torsion Angle Observations by Hidden Markov Models*

Alexander Fischer Sonja Waldhausen Illia Horenko Eike Meerbach
Christof Schütte†

December 11, 2006

Institute of Mathematics II, Free University Berlin, Arnimallee 2–6, 14195 Berlin, Germany

Abstract

We present a novel method for the identification of the most important conformations of a biomolecular system from molecular dynamics or Metropolis Monte Carlo time series by means of Hidden Markov Models (HMMs). We show that identification is possible based on the observation sequences of some essential torsion or backbone angles. In particular, the method still provides good results even if the conformations do have a strong overlap in these angles. In order to apply HMMs to angular data we use von Mises output distributions. The performance of the resulting method is illustrated by numerical tests and by application to a hybrid Monte Carlo time series of trialanine and to MD simulation results of a DNA–oligomer.

Keywords: Aggregation, Angular Data, Biomolecules, EM Algorithm, HMM, Metastability, Torsion Angle, Viterbi Algorithm, von Mises Distribution

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

† *Correspondence to:* Ch. Schütte; E-mail: schuette@mi.fu-berlin.de

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 [12, 14, 26, 25], while on closer inspection it exhibits a rich temporal multiscale structure [21]. Biophysical research seems to indicate that typical biomolecular systems possess only few dominant conformations that can be characterized in terms of a small number of essential degrees of freedom [1]. In many cases these essential degrees of freedom can be identified from some torsion or backbone angles of the molecule under consideration. Then, metastable conformations can be characterized in terms of these angles and can thus be identified from the molecular dynamics time series projected onto these angles.

In this article we will discuss the application of Hidden Markov Models (HMM) [22, 23] to the analysis of time series from biomolecular simulations. Numerous applications of HMMs can be found in the fields of speech recognition [22, 23], signal processing [13], and bioinformatics [16]. Applications to the analysis of the dynamical behavior of biomolecules is rather rare, and mainly is concerned with the analysis of experimental data, see [15, 29], for example. In this article, HMMs are employed to identify biomolecular conformations observation sequences of a few torsion

angles given by some high dimensional time series. An intriguing feature of HMM analysis is that the set of essential coordinates chosen for analysis need not enable a geometric separation of the important conformations. Instead, the EM [4, 3, 9] and Viterbi algorithm [30] associated with a HMM extract geometric as well as dynamical properties out of the given observation sequences, and thus allow for a separation of overlapping conformations from the dynamics within the sequences.

Most applications of HMM use discrete or Gaussian distributions as output observations. In order to deal with continuous angular data we employ the so-called *von Mises distribution* that is perfectly designed to analyze Gaussian like distributions of angular data [18]. Although this distribution is encountered rarely, it has already been shown that it can easily be included into the HMM framework [28, 27].

The conceptual basis of the proposed approach is new: The information about which and how many biomolecular conformations are important for the description of effective dynamics of the system under consideration is understood as being hidden within the time series. Thus, the available conformations are *hidden conformation states* in the sense of HMMs. In addition, the number of hidden conformation states is automatically determined via aggregation of hidden states by methods proposed in [10, 11] , i.e., an a-priori knowledge about the number of biomolecular conformations is not needed.

Outline. We will proceed as follows: First, we shortly review the Hidden Markov Model and the von Mises distribution. Then, we discuss the associated algorithms that are fundamental for the intended conformational analysis, and particularly introduce the combination of HMMs with the Perron cluster analysis [10, 11] for aggregation of hidden states into conformation states. In the last section, we first illustrate the application of this novel dynamical clustering technique to metastable observation sequences by suitable numerical experiments, and finally discuss the applicability of the resulting concept to two realistic molecular dynamics time series.

Model

Hidden Markov Models

A hidden Markov model (HMM) is a stochastic process with hidden and observable states. The hidden process consists of a sequence $X_1, X_2, X_3, \ldots$ of random variables taking values in some "state space", the value of X_t being "the state of the system at time t". In applications these states are not observable, and therefore called hidden.

In HMM context the sequence of hidden states is assumed to be a Markov process, that is it has the Markov property, meaning that the conditional distribution of the "future" X_{n+1} given the "past", $X_1, \ldots, X_n$, depends on the past only through X_n. Since the HMM state space in general is finite, we thus are concerned with a Markov chain, which is characterized by the so-called transition matrix $P = (p_{ij})$, whose entry p_{ij} correspond to the conditional probability of switching from the (hidden) state i to state j. The sum over j has to be one for each i, which means that the transition matrix is a row-stochastic matrix.

Each hidden state causes a specific output that might be either discrete or continuous. This output is distributed according to a certain conditional distribution (conditioned to the hidden state). Thus, realizations of HMM are concerned with two sequences, an observation sequence and a sequence of hidden states.

In our case the information initially hidden is in which metastable subset (conformation) the molecular system is at a certain time, while the information on the state of the selected torsion or backbone angles is completely known. A HMM then consists of a Markov chain model for the hidden (metastable) states that encodes with which probability one switches from one hidden state to another, and a *conditional probability* of observation of specific torsion angles *if* one is in a certain hidden state. To describe the whole system, we need to know the number of hidden states, the transition matrix between them, an initial distribution, and for each state a certain probability distribution for the observation.

Therefore, a HMM formally is defined as a tuple

$\lambda = (S, V, P, f, \pi)$ where

- $S = \{s_1, s_2, ..., s_N\}$ is a set of a finite number N of states,
- $V \subset \mathbb{R}^k$ is the observation space,
- $P = (p_{ij})$ is the transition matrix, where $p_{ij} = \mathbb{P}(X_{t+1} = s_j | X_t = s_i)$ describes the transition probability from state s_i to state s_j,
- $f = (f_1, f_2, \ldots, f_N)$ is a vector of probability density functions (pdf) in the observation space,
- $\pi = (\pi_1, \ldots, \pi_N)$ is a stochastic vector, that describes the initial state distribution, $\pi_i = P(X_1 = s_i)$.

Often, the short notation $\lambda = (P, f, \pi)$ is used since S and V are implicitly included. Particularly S is identified with the index set $\{1, 2, \ldots, N\}$ in the following.

HMMs can be set up for discrete or continuous observations. For continuous observations the most popular choice is to use (multivariate) normal distributions for the output distributions f_k. However, in the case of circular data (like torsion angle positions in a molecular dynamics simulation) the use of normal distributions often induces crucial problems due to periodicity.

Von Mises Distribution

For circular data, the normal distribution can be replaced with the *von Mises distribution*, whose properties are comparable but regarding the sphere [18] instead of regarding the real axis. In one dimension the von Mises distribution $M(\mu, \kappa)$ is given by the following pdf depending on the two parameters $\mu \in [0, 2\pi]$ and $\kappa > 0$:

$$f_{\mu,\kappa}(\theta) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}, \qquad 0 \leq \theta < 2\pi, \qquad (1)$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero, i.e.,

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} \left(\frac{\kappa}{2}\right)^{2r}. \qquad (2)$$

The parameters characterizing the von Mises distribution are called the *mean direction* μ and the *concentration parameter* κ. It is unimodular, i.e., single-peaked, and symmetrical about the mean direction. The maximum of the pdf, the so-called mode, is at the mean, while the minimum of the pdf (anti-mode) is located at $\mu + \pi$. The ratio of the pdf at the mode to that at the antimode is given by $e^{2\kappa}$, so that the larger the value of the concentration parameter κ, the more pronounced the concentration of the distributed data around the mode. In contrast, for $\kappa \to 0$ the von Mises distribution becomes uniformly distributed. For an illustration see Fig. 1, for more details see [18].

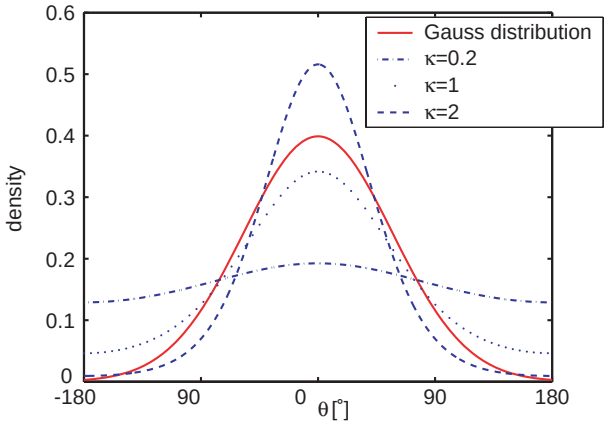

Figure 1: Different von Mises distributions $f_{\mu,\kappa} = f_{\mu,\kappa}(\theta)$ (each with mean $\mu = 0$ and different concentration parameters κ in comparison with a Gaussian distribution with mean 0; here and in the following figures we display angles in terms of degrees in the interval $[-180^o, 180^o)$, simply for the sake of convenience).

Maximum Likelihood Principle

In statistics, likelihood —unlike probability, which estimates the degree of belief in unknown consequences of known causes— works backwards, from observed results to hypothetical models and parameters. The Maximum Likelihood Principle is the standard technique for estimating parameters in a statistical model with respect to observed data. For the context important herein, it can be explained as follows: Let λ

denote a set of parameters, e.g., the specification of an HMM-model as used before, $\Theta = (\theta_1, \theta_2, \ldots, \theta_T)$ an observed sequence of data, and $p(\Theta|\lambda)$ the statistical model, i.e., the probability of the observation sequence, or, depending on the state space, the density of the observation sequence, under the conditions that the parameter set λ is given [1] . The likelihood function then is defined as $L(\lambda) = p(\Theta|\lambda)$, i.e., we consider the observation sequence as being given and ask for the variation of the probability (density function) in terms of the parameters. The maximum likelihood principle then simply states, that the optimal parameters are given by the absolute maximum of L. Thus, the maximum likelihood principle is an optimization problem in parameter space.

Parameter estimation. Like for the mean and variance of a normal distribution there exist maximum likelihood estimators (MLE) for the mean and concentration parameter of the von Mises distribution. We shortly review the results collected in [18].

Suppose that we are given T data points on the sphere with angles $(\theta_1, \theta_2, \ldots, \theta_T)$ in radians, i.e. $\theta_k \in [0, 2\pi[$. The likelihood function $L(\mu, \kappa)$ for the von Mises distribution is then given by

$$\begin{aligned} L(\mu,\kappa) &= \prod_{t=1}^{T} f_{\mu,\kappa}(\theta_t) \qquad (3) \\ &= \frac{\exp\left(\kappa \sum_{t=1}^{T} \cos(\theta_t - \mu)\right)}{(2\pi)^T I_0(\kappa)^T}. \end{aligned}$$

According to the maximum likelihood principle the optimal parameters $(\hat{\mu}, \hat{\kappa})$ are the ones for which $L(\hat{\mu}, \hat{\kappa})$ is maximal. Maximization of $L(\mu, \kappa)$ gives

$$\hat{\mu} = \begin{cases} \text{Arctan}\,(y/x) & \text{if } x > 0, \\ \pi + \text{Arctan}\,(y/x) & \text{if } x < 0, \end{cases} \qquad (4)$$

with

$$x = \frac{\sum_{t=1}^{T} \cos(\theta_t)}{T} \text{ and } y = \frac{\sum_{t=1}^{T} \sin(\theta_t)}{T}, \qquad (5)$$

[1] In the following, to avoid linguistic confusion, we refer to $p(\Theta|\lambda)$ as probability, but always include the meaning of a probability density function.

and $[0, \pi[$ as the range of Arctan.

The formula for the mean estimate $\hat{\mu}$ becomes intuitively clear by observing that the cartesian coordinates $\hat{x}$ and $\hat{y}$ are the mean values of the projection of the data, seen as points on the unit sphere, on the x- and y-axis, respectively. Thus Eq. (4) tells us that $\hat{\mu}$ is the angle of the cartesian mean (x, y), and that the operation Arctan$(\cdot)$ in Eq. (4) means the inverse of the tangent operation in the sense that it gives back the correct angle in polar coordinates of the point (x, y) on the unit circle. Then, setting

$$\hat{r} = (x^2 + y^2)^{\frac{1}{2}},$$

we get the polar coordinate representation $(\hat{\mu}, \hat{r})$ of the cartesian mean (x, y).

The cartesian mean does not necessarily lie on the unit circle. The length $\hat{r}$ can take values between 0 and 1 and measures the clustering of the data around the mean angle $\hat{\mu}$. The closer $\hat{r}$ is to 1, the more concentrated the sample is around the mean.

But, besides the intuitive picture, the MLE for κ can not given explicitly. Instead, by maximization of $L(\mu, \kappa)$ wrt. κ and making use of properties of the Bessel-functions, one can show that the MLE for κ solves the expression

$$\frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{1}{T} \sum_{i=1}^{T} \cos(\theta_i - \mu), \qquad (6)$$

where $I_1(\kappa)$ is the modified Bessel function of the first kind and order one. Eq. 6 can only be solved numerically, which is not a big obstacle as the lhs of (6) is monotonically increasing. Our intuitive picture remains valid in the sense that one can show that the rhs of (6) equals $\hat{r}$.

Multidimensional generalization of the von Mises distribution are obvious as long as we simply use tensor products of one-dimensional von Mises distributions. This means, that the mean direction μ has to be replaced by a vector and the concentration parameter κ simply is replaced by a diagonal matrix. The jth (diagonal) entries are computed from the jth components of the data points $\theta \in \mathbb{R}^d$ due to (4) and (6).

Methods

Application of HMMs typically involves one or all of the three problems (P1) to (P3) stated below. For our specific task of identifying metastable conformations we have to complement these basic problems by also addressing problems (P4) and (P5):

(P1) Calculation of the probability $p(\Theta|\lambda)$ of a certain observation sequence Θ for a given model $\lambda = (P, f, \pi)$.

(P2) Estimation of the best model parameters for a given observation sequence (e.g., torsion angle observations).

(P3) Given the model λ and an observation sequence Θ, finding the most probable hidden state sequence $X^* = (x_1^*, x_2^*, \ldots, x_T^*)$.

(P4) Aggregation of hidden states to conformation states, and clustering of the observation sequence into conformations.

(P5) Avoid combinatorial explosion of computational effort for high dimensional observation sequences.

There are standard algorithms that allow to solve problems (P1) to (P3) efficiently, which we will shortly outline here; our presentation will follow [22, 23]. Most notably, due the Markovian structure of the hidden chain and the independence of observations given the hidden states, all algorithms run linear in the length of the observation sequence. For dealing with problems (P4) and (P5) no standard algorithm exist. Below we will thus also outline the novel algorithms suggested herein.

Problem (P1): Forward-Backward Variables

The most straightforward method to calculate the probability $p(\Theta|\lambda)$ of an observation sequence $\Theta = (\theta_1, \theta_2, \ldots, \theta_T)$ given the model λ, is computing every possible hidden state sequence $X = (x_1, x_2, \ldots, x_T)$ of length T and sum up all probabilities conditioned on the hidden sequence:

$$p(\Theta|\lambda) = \sum_{X=(x_1,\ldots,x_T)} p(X|\lambda)p(\Theta|X,\lambda).$$

This method involves on the order of $2TN^T$ calculations therefore becoming infeasible even for small numbers of hidden states N and short observation lengths T.

Fortunately, the effort can be reduced to TN^2 by using the so-called forward and backward variables. These divide the observation sequence Θ recursively in partial subsequences: those from time 1 to time t and those from $t+1$ up to T. The forward variables are given by

$$\alpha_t(i) = p(\theta_1, \theta_2, \ldots, \theta_t, X_t = i \mid \lambda), \qquad (7)$$

denoting the probability of the observation sequence up to time t together with the information that the system is in hidden state i at time t conditioned wrt. the given model λ. Whereas the backward variables are given by

$$\beta_t(i) = p(\theta_{t+1}, \theta_{t+2}, \ldots, \theta_T \mid X_t = i, \lambda), \qquad (8)$$

denoting the probability of the observation sequence from time $t+1$ to T, conditioned that the system is in hidden state i at time t and on the model λ. The computation of the probability $\alpha_T(i)$ for the whole sequence is possible with $N^2\,T$ operations, as recursive formulas can be used:

$$\begin{aligned} \alpha_1(i) &= \pi_i f_i(\theta_1), \qquad 1 \le i \le N, \\ \alpha_{t+1}(j) &= \left[\sum_{i=1}^{N} \alpha_t(i) p_{ij}\right] f_j(\theta_{t+1}), \\ & \qquad 1 \le t \le T-1,\ 1 \le j \le N. \end{aligned}$$

In the case of von Mises distributed output the f_j's denote the probability density functions f_{μ_j,κ_j} we discussed before.

The backward variable $\beta_t(i)$ can be computed with

analogous formula:

$$\begin{aligned}\beta_T(i) &= 1, 1 \le i \le N,\\ \beta_t(i) &= \sum_{j=1}^{N} p_{ij} f_j(\theta_{t+1}) \beta_{t+1}(j),\\ &\quad T-1 \ge t \ge 1,\ 1 \le i \le N\end{aligned}$$

From (7) and (8) one can compute the desired probability for all t as

$$p(\Theta \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i).$$

Forward and backward variables form the basis of the EM algorithm.

Problem (P2): EM Algorithm

The Expectation-Maximization (EM) algorithm [9] is a maximum likelihood approach that improves iteratively an initial parameter set, and converges to a local maximum of the likelihood function. In the case of HMMs this method was already worked out before the general EM formulation, and is often called the Baum-Welch algorithm [4, 3].

To apply the EM algorithm to a given observation sequence, we have to set up an HMM by a parameter set $\lambda = \lambda(P, f, \pi)$, assuming a finite number N of hidden states, an output distribution function for each of the hidden state, and the initial distribution for the hidden Markov chain.

EM steps. There is no known way to analytically determine the model parameters that globally maximize the probability of the given observation sequence. We can, however, estimate a $\hat{\lambda}$ that locally maximizes the likelihood $L(\lambda) = p(\Theta \mid \lambda)$. The EM algorithm is a learning algorithm, it alternately iterates two steps, the Expectation step and the Maximization step. Starting with some initial model $\lambda^{(0)}$ the steps iteratively refine the model:

- The Expectation-step: In this step the state occupation probabilities

$$\gamma_t(i) = p(X_t = i \mid \Theta, \lambda^{(k)}),$$

and the joint probabilities

$$\xi_t(i,j) = p(X_t = i, X_{t+1} = j \mid \Theta, \lambda^{(k)}),$$

are calculated for each time t in the sequence, given the observation Θ and the current model $\lambda^{(k)}$.

- The Maximization-step: This step finds a new model $\lambda^{(k+1)}$ via a set of reestimation formulas. The maximization guarantees that $L(\lambda^{(k+1)}) \ge L(\lambda^{(k)})$.

The two conditional probabilities of the **E-step** can be calculated efficiently by using the forward-backward variables (7)-(8), as

$$\begin{aligned}\xi_t(i,j) &= \frac{p(X_t = i, X_{t+1} = j, \Theta, \lambda^{(k)})}{p(\Theta, \lambda^{(k)})}\\ &= \frac{p(X_t = i, X_{t+1} = j, \Theta \mid \lambda^{(k)})}{p(\Theta \mid \lambda^{(k)})}\\ &= \frac{\alpha_t(i) p_{ij} f_j(\theta_{t+1}) \beta_{t+1}(j)}{p(\Theta \mid \lambda^{(k)})}.\end{aligned}$$

We omitted the superscript k for the forward and backward variables in order to not obscure the formulas by too many indices.

With these values the probability to be in hidden state i at time t can be expressed as

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j).$$

The **M-step** consists of reestimation formulas to obtain an improved model $\lambda^{(k+1)}$. The estimators for the initial distribution and the transition probabilities of the hidden Markov chain are given by

$$\begin{aligned}\pi_i^{(k+1)} &= \gamma_1(i),\\ p_{ij}^{(k+1)} &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)},\end{aligned}$$

using the probabilities computed in the E-step before.

We also need to reestimate parameters for the probability density functions $f_j^{(k+1)}$ via their maximum likelihood estimator. Hereby, the observations

θ_t used to determine $f_j^{(k+1)}$ have to be weighted with the probability $\gamma_t(j)$ to be in the hidden state j. To estimate parameters of von Mises output distributions from torsion angle observation sequences $\theta_1, \ldots, \theta_T$, Eqs. (5) and (6) have to be adopted as follows [27, 28]:

$$x(j) = \frac{\sum_{t=1}^{T} \gamma_t(j) \cos(\theta_t)}{\sum_{t=1}^{T} \gamma_t(j)},$$

and

$$y(j) = \frac{\sum_{t=1}^{T} \gamma_t(j) \sin(\theta_t)}{\sum_{t=1}^{T} \gamma_t(j)},$$

and the mean parameter $\mu_j^{(k+1)}$ of $f_j^{(k+1)} = f_{\mu_j^{(k+1)}, \kappa_j^{(k+1)}}$ is computed from $x(j)$ and $y(j)$ due to (4). Finally, the concentration parameter is computed by fitting

$$\frac{I_1(\kappa_j^{(k+1)})}{I_0(\kappa_j^{(k+1)})} = \frac{1}{\sum_{t=1}^{T} \gamma_t(j)} \sum_{t=1}^{T} \gamma_t(j) \cos(\theta_i - \mu_j^{(k+1)}).$$

Again we avoided the superscript $(k+1)$ for $x(j), y(j)$ and γ_j, to reduce complexity of the stated formulas.

The E- and M-steps are iteratively repeated until a predetermined maximal number of iterations is reached or the improvement of the likelihood becomes smaller than a given limit. As it is a property of the EM algorithm [9] that $L(\lambda^{(k+1)}) \geq L(\lambda^{(k)})$, we iteratively approximate a local maxima $\hat{\lambda}$.

Problem (P3): Viterbi Algorithm

The Viterbi algorithm [30] computes for a given HMM model λ and an observation sequence Θ the most probable hidden path $X^* = (x_1^*, \ldots, x_T^*)$. This path is called the *Viterbi path*.

For an efficient computation we define the highest probability (density) along a single path, for the first t observations, ending in the hidden state i at time t,

$$\delta_t(i) = \max_{x_1, x_2, \ldots x_{t-1}} P(x_1, x_2, \ldots, x_t = i, \theta_1, \theta_2, \ldots, \theta_t | \lambda).$$

This quantity can be computed recursively by

$$\delta_t(i) = \max_{1 \leq j \leq N} [\delta_{t-1}(j) p_{ji}] f_i(\theta_t). \qquad (9)$$

In addition, we assign to $\psi_t(i)$ the argument that maximizes (9), in order to actually retrieve the most likely hidden state sequence. These quantities are calculated for each t and i, afterwards the most probable hidden state sequence can be retrieved from δ and ψ by backtracking, as shown in the Viterbi algorithm scheme:

1) Initialization:

$$\begin{aligned} \delta_1(i) &= \pi_i f_i(\theta_1),\ 1 \leq i \leq N \\ \psi_1(i) &= 0 \end{aligned}$$

2) Recursion:

$$\begin{aligned} \delta_t(i) &= \max_{1 \leq j \leq N} [\delta_{t-1}(j) p_{ji}] f_i(\theta_t), \\ \psi_t(i) &= \operatorname*{argmax}_{1 \leq j \leq N} [\delta_{t-1}(j) p_{ji}], \\ & \quad 2 \leq t \leq T, 1 \leq i \leq N \end{aligned}$$

3) Backtracking:

$$\begin{aligned} x_T^* &= \operatorname*{argmax}_{1 \leq j \leq N} [\delta_T(j)] \\ x_t^* &= \psi_{t+1}(x_{t+1}), \quad t = T-1, T-2, \ldots, 1. \end{aligned}$$

The mapping of the T observation states to N hidden states provided by the Viterbi path X^* results in a dynamical clustering of the data.

Problem (P4): Aggregation into Conformation States

The term "biomolecular conformation" is used in different meanings. We will herein use it in the following sense: A conformation is a kind of "global state" of the molecule, in which the large scale geometric structure is understood to be conserved, while the molecule locally exhibits metastability. Such a conformation is called metastable if a transition from it to other large scale geometric structures is rare. As an intuitive picture for the occurrence of metastable conformations might serve the vicinity of a deep minimum of the (free) energy landscape, defining the large scale geometry, together with all neighboring relative minima, contributing to local flexibility.

The problem of finding these metastable conformations can be formulated as the problem of *aggregating* molecular states into conformation states (i.e., clustering states that belong to the same metastable conformation). The identification of an optimal aggregation from observation of the dynamical behavior, e.g., MD-trajectories, is a difficult algorithmic problem.

However, if a suitable trajectory and a box discretization, which is a collection of nonoverlapping sets (boxes) $B_1, \ldots, B_N$ covering the whole observable space, is at hand, optimal aggregates can be identified via the spectral properties of a stochastic transition matrix $\widetilde{P}$. This transition matrix $\widetilde{P} = (\tilde{p}_{ij})$, containing the overall transition probabilities between all discrete states of the system under consideration, can be obtained by counting transitions in the trajectory between the discrete states:

$$\tilde{p}_{ij} = \frac{\#\{\theta_t \in B_i \ \theta_{t+1} \in B_j\}}{\#\{\theta_t \in B_j\}},$$

where t is meant to vary from 1 to $T-1$.

The identification of an optimal aggregation from the spectrum of the transition matrix $\widetilde{P}$ is possible, as the number of metastable sets, corresponding to metastable conformations, equals the number of eigenvalues close to one, while the metastable sets themselves can be obtained from the structure of the corresponding eigenvectors [8, 25]. This has led to the construction of an aggregation technique called "Perron Cluster Cluster Analysis" (PCCA) [10, 11]. Of course, the technique depends on extracting a suitable set of observables from a given trajectory and finding an appropriate discretization of the observable space. Both problems are non-trivial and solutions often depend on a mixture of insight and preliminary analysis, an algorithmic scheme is described in [7].

While PCCA can detect the number of (hidden) metastable sets by the spectrum of the transition matrix, HMM methods have the drawback that for a given observation sequence one is always confronted with the task to select *in advance* the number of hidden states. On the other hand PCCA crucially depends on geometrical separation of metastable sets in the observable space, as otherwise no appropriate box discretization can be found, while HMM can identify overlapping distributions. To harvest the benefit or both methods, we will use PCCA within the HMM framework as follows: Start the HMM algorithm with some sufficient number of hidden states, say N, that should be greater than the expected number of conformation states (keeping in mind that the algorithmic effort will increase like $\mathcal{O}(N^2T)$). After termination of the EM algorithm, take the resulting transition matrix $\hat{P}$ and aggregate the N hidden states into $M \leq N$ conformation states by means of PCCA, i.e., by means of analyzing the spectrum of $\hat{P}$. Note, that we replaced the transition matrix $\widetilde{P}$, obtained from counting transitions in the trajectory, by the transition matrix $\hat{P}$, obtained from the estimated HMM model. By aggregating the Viterbi path according to the outcome of the PCCA analysis , we obtain a clustering of the observation states into metastable sets. This novel combination of HMMs with PCCA will be exemplified in the next section.

Problem (P5): High Dimensions and Viterbi-Clustering

In large biomolecular systems the dimension of each state in the observation sequence may be large, say d. The number of free parameters of each output distribution will typically increase with d more than linearly. A full d-dimensional Gaussian or von Mises distribution, for example, has $\mathcal{O}(d^2)$ free parameters in the covariance matrices. However, for increasing dimensions of the parameter space the likelihood maximization via the EM algorithm will converge more and more slowly (and eventually fail to detect the maximum) if the dimensions get too large. In order to overcome this "curse of dimension" we suggest the following novel algorithmic scheme for which the name *Viterbi-clustering* seems appropriate:

First, the high dimensional state space V of the observation sequence is decomposed into a sequence of low-dimensional subspaces $V^{(1)}, \ldots, V^{(k)}$. For example, V could be the state space of all torsion angles of the system under consideration, and $V^{(j)}$ the subspace of a single torsion angle. By choice of the $V^{(j)}$, $j = 1, \ldots, k$, and projection onto each one, we get k low-dimensional time series $\Theta^{(j)} = \theta_1^{(j)}, \ldots, \theta_T^{(j)}$ out

of the original time series $\theta_1, \ldots, \theta_T$.

Second, each of these low-dimensional time series is separately analyzed by means of the above HMM procedure, i.e. by approximating the model with the EM-algorithm and afterwards obtaining the Viterbi path with the Viterbi algorithm, and finally aggregating with PCCA. This results in k aggregated Viterbi paths $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \ldots, x_T^{(j)})$ that represent the conformational dynamics as detected from the information contained in the single time series $\Theta^{(j)}$, $j = 1, \ldots, k$. Observe, that by using HMMs we are able to discretize the low-dimensional projections even if the metastable sets are overlapping in the projected space, as the output distributions of the HMM are allowed to overlap.

Third, these single Viterbi paths can be combined into a *global* Viterbi path X: at instance t its state is $x_t = (x_t^{(1)}, \ldots, x_t^{(k)})$. This global Viterbi path can be seen as a time series originating from a system with discrete states from $S = \times_{j=1}^{k} S^{(j)}$, where $S^{(j)}$ denotes the state space of the jth Viterbi path. The combined global Viterbi path has finitely many states such that we can directly compute the associated transition matrix. Fourth, the metastable states of this transition matrix are identified, again by means of PCCA. Based on these metastable states, the global Viterbi path is aggregated into an clustered global Viterbi path whose resulting discrete states finally are interpreted as the global conformation states of the original full-dimensional time series $\theta_1, \ldots, \theta_T$.

Results and Discussion

We first illustrate the fundamental features and performance of the algorithms. Then, we demonstrate how conformation states can be identified from a hybrid Monte Carlo simulation of trialanine and of a specific DNA-oligomer by analyzing torsion angle time series.

Illustrative Examples

In order to generate data sets for illustrative means, we computed time series from direct realizations of fully specified HMM models, i.e., for given parameters (P, f, π). Based on the observation sequence of such realizations we tried to re-identify the parameters by application of the EM and Viterbi algorithms. A Realization of an observation sequence $\Theta = \theta_1, \ldots, \theta_T$ for a given HMM (P, f, π) is easily obtained: One starts in a single hidden state randomly chosen from π, say j, and draws the first output state θ_1 from the associated distribution f_j. Then the next hidden state is chosen according to the transition probabilities given by the j-th row of the transition matrix P. This procedure is iterated T times.

We considered four parameter sets, in which P always is a 2×2 stochastic matrix, i.e., there are two hidden states, and the output pdfs $f_{\mu_i,\kappa_i}, i = 1, 2$ are von Mises distribution pdfs. The parameters of the four test cases are given in Table 1.

case	(μ_1, κ_1)	(μ_2, κ_2)	P
(a)	$(-\pi/2, 1)$	$(\pi/2, 3)$	$\begin{pmatrix} 0.95 & 0.05 \\ 0.02 & 0.98 \end{pmatrix}$
(b)	$(-\pi/2, 1/2)$	$(\pi/2, 1)$	$\begin{pmatrix} 0.95 & 0.05 \\ 0.02 & 0.98 \end{pmatrix}$
(c)	$(-0.5, 1)$	$(0.5, 3)$	$\begin{pmatrix} 0.95 & 0.05 \\ 0.02 & 0.98 \end{pmatrix}$
(d)	$(-\pi/2, 1)$	$(\pi/2, 3)$	$\begin{pmatrix} 0.6 & 0.4 \\ 0.45 & 0.55 \end{pmatrix}$

Table 1: Von Mises parameters and transition matrices of the four HMMs used for generation of the observation sequences.

These four cases (a)-(d) have been chosen to illustrate the following scenarios (see Fig. 2):

(a) In this case the two von Mises distributions show no significant overlap since the means are well separated and the concentration parameters are large enough. In addition, the transition matrix is clearly metastable such that the period of time in which the system remains within one hidden state is long. Therefore this is expected to be an easy case for the identification algorithm.

(b) This case differs from the first case by smaller concentration parameters broadening the two von Mises distributions, thus introducing sa ignificant overlap.

(c) This case differs from the first only by closer means of the von Mises distributions, thus also introducing a significant overlap.

(d) The last case differs from the first only by the fact that the transition matrix is no longer metastable.

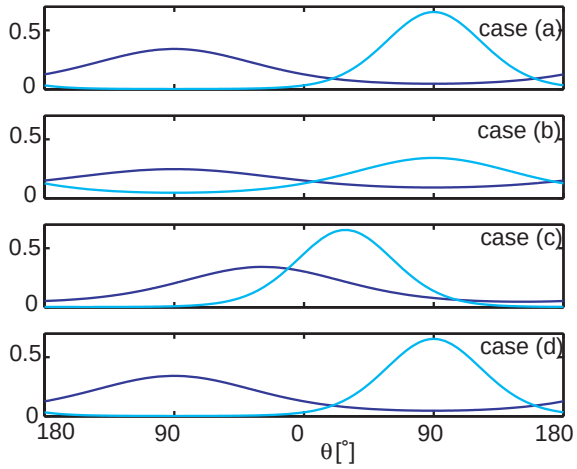

Figure 2: Von Mises probability density functions f_{μ_1,κ_1} and f_{μ_2,κ_2} for each of the four test cases (a)-(d), specified in Table 1.

The problematic cases for parameter identification by the EM algorithm should be obvious: If there are two (or more) hidden states with overlapping distributions, then, with input of the observation sequence alone, the algorithm may have trouble to detect or reliably identify the two states.

Based on an observation sequence of length $T = 2000$ for each of the four HMMs, the EM algorithm has been started with the same generic initial parameters independent of the observation sequence. The EM iteration has been automatically stopped after some iteration steps (denoted by M in Table 2) according to a pre-selected termination criterion that is based on the increase in the likelihood function. In all four cases the estimated parameters are good approximations of the original model parameters (see Table 2).

To furthermore illustrate the performance of the Viterbi algorithms for our four test cases, we compared for each case the information about the original hidden state sequence with the Viterbi path computed by the Viterbi algorithm based on the estimated HMM given in Table 2. Figure 3 illustrates the differences in the hidden state sequences for the scenarios (a) and (b). The quality of the Viterbi paths is astonishingly good; most mismatches only occur during transitions between hidden states.

case	$(\hat{\mu}_1, \hat{\kappa}_1)$	$(\hat{\mu}_2, \hat{\kappa}_2)$	$\hat{P}$
(a) $M = 8$	$(-1.54, 1)$	$(1.57, 3.1)$	$\begin{pmatrix} 0.96 & 0.04 \\ 0.02 & 0.98 \end{pmatrix}$
(b) $M = 26$	$(-1.58, 0.5)$	$(1.59, 1.01)$	$\begin{pmatrix} 0.95 & 0.05 \\ 0.02 & 0.98 \end{pmatrix}$
(c) $M = 17$	$(-0.53, 1)$	$(0.51, 32)$	$\begin{pmatrix} 0.96 & 0.04 \\ 0.02 & 0.98 \end{pmatrix}$
(d) $M = 13$	$(-1.54, 1.3)$	$(1.62, 2.6)$	$\begin{pmatrix} 0.60 & 0.40 \\ 0.45 & 0.55 \end{pmatrix}$

Table 2: Parameters estimated (up to two digits) by the EM algorithm after M steps based on the observation sequences from the four test cases (a)-(d).

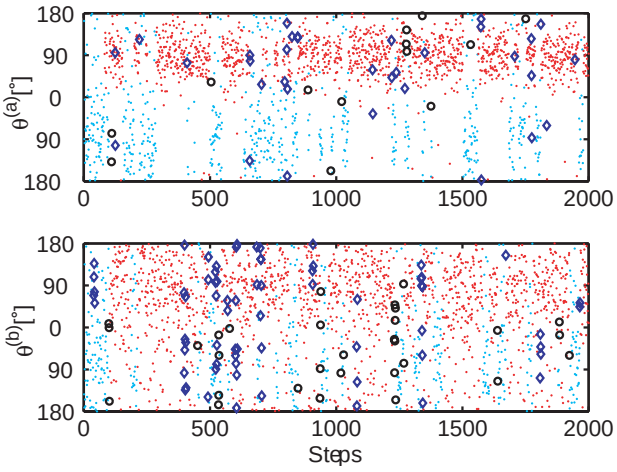

Figure 3: Observation sequences $\theta^{(a)}$ and $\theta^{(b)}$ for the test cases (a) and (b). The colors (or greyscales) represent the two hidden states according to the Viterbi path; all instances with mismatches between the original hidden state sequence and the Viterbi path have been marked with circles and diamonds.

Application to Molecular Data

Time series from molecular dynamics and Metropolis Monte Carlo simulations of biomolecules are often analyzed in terms of some essential torsion angles. We will demonstrate that even in the case of overlapping marginal distributions such torsion angle time series possess sufficient characteristics to make an identification of conformational substates by HMM models possible. It should be noted that it is in general a difficult task to extract a meaningful set of collective observables or reaction coordinates from a molecular system. In many cases system-dependent insight is needed. This difficulty can not be solved in general by the algorithmic scheme we propose, but the use of HMM methods allows to extract information even if the observables chosen *only reflect* somehow the global changes. So recent research suggests that it is possible to get information about water clusters from the torsion angles of a solvated molecule by a HMM based analysis [20].

The use of von Mises output distributions in application of HMMs to torsion angle time series, is motivated by (a) the fact that the von Mises distribution is the simplest peaked distribution for circular data, in the sense that it is completely determined by the first two moments (as in the planar case the Gaussian distribution is), (b) the observation that torsion angle distributions can almost ever be interpreted as mixtures of Gaussian-like peaks; since the HMM embedding allows to couple single von Mises distributions into mixtures this seems to be an appropriate choice, and (c) by the observation that HMMs with Gaussian output distributions fail to reproduce certain aspects of the system whenever periodicity is relevant.

In addition, other (non-stationary) output distributions are well possible, as described in, e.g., [19].

Trialanine. In the following we present results from the application of the proposed HMM framework to trialanine, a small peptide composed of three alanine amino acid residues.

For the simulation of trialanine we employed the Gromos96 vacuum force field [17], in which trialanine is represented by 21 united atoms. The structural and dynamical properties of this molecule are mainly determined by two central peptide backbone angles Φ and Ψ. In addition, at very high simulation temperatures the otherwise planar peptide bond angle Ω, may also undergo some conformational transition (see Fig. 4).

The time series of 50000 steps has been generated by means of a Hybrid Monte Carlo (HMC) [6] scheme at a temperature of 700 K. That means at each step momenta were randomly drawn from a maxwell distribution, according to the temperature, and proposal steps were generated by integration in time for 500 fs, using the Verlet integration scheme based on a 1 fs time step. This yielded an acceptance rate of about 93 percent.

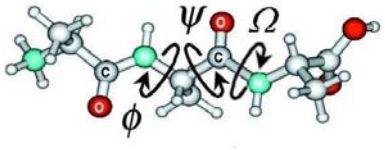

Figure 4: The trialanine molecule shown in ball-and-stick representation. Typically, the overall structure of trialanine is sufficiently described by the two torsion angles Φ and Ψ, but at higher temperatures one should also take into consideration changes of the peptide bond angle Ω.

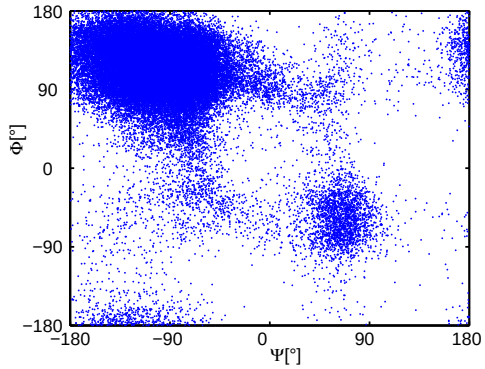

Figure 5: Ramachandran plot in Φ and Ψ of the time series obtained from HMC simulation at 700 K.

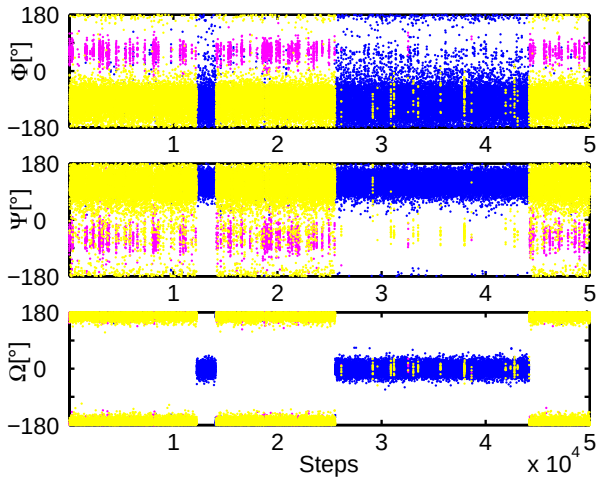

Figure 6: Observation sequences of the three torsion angles Φ, Ψ, and Ω (from top to bottom) of trialanine at 700 K. We observe conformational changes in Ω that are correlated to the dynamical behavior of Ψ and Φ. The three colors (or greyscales) correspond to the aggregated Viterbi path from the HMM analysis of the two-dimensional observation sequence given by Φ and Ψ.

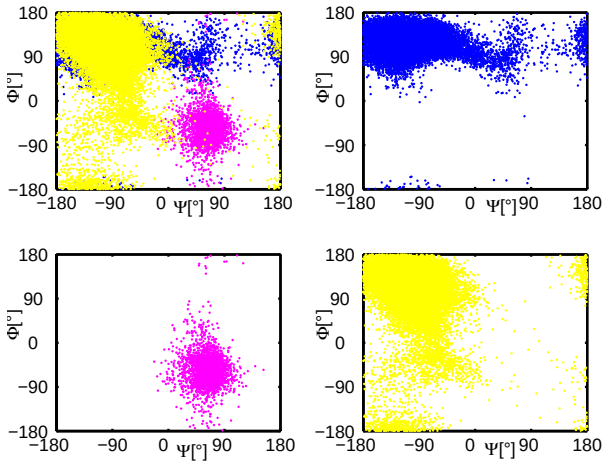

Figure 7: Ramachandran plot in Φ and Ψ colored according to the three aggregated hidden states as computed by the EM and Viterbi algorithms. Top left: all observation points colored according to the aggregated Viterbi path. Top right, bottom left, and bottom right: output from each of the aggregated hidden states plotted separately.

From this time series we computed the observation sequences in Φ, Ψ, and Ω. Fig. 5 shows the obtained data points projected on the (Φ, Ψ)-plane, the so-called Ramachandran plot. We observe that the (Φ, Ψ)-plane has been explored rather intensively due to the high temperature of 700 K. Furthermore we observe transitions between substates of the torsion angle Ω (see Fig. 4) that we do not observe in 300 K simulations of this length. As one can see in Fig. 6, these transitions introduce additional features of conformation dynamics. Thus, by raising the simulation temperature, we obtained an intriguing test of the novel HMM technique whether it can identify this additional structure based on the observation sequences of the peptide angles Φ and Ψ alone.

Therefore we used the two-dimensional observation sequence given by Φ and Ψ to estimate a HMM with 7 hidden states, each with a two-dimensional von Mises output distribution. After successful termination of the EM algorithm (after 28 iteration steps) we got the following transition matrix:

$$\hat{P} = \begin{pmatrix} 0.81 & 0.01 & 0.01 & 0.00 & 0.00 & 0.17 & 0.00 \\ 0.20 & 0.59 & 0.01 & 0.00 & 0.02 & 0.16 & 0.02 \\ 0.00 & 0.00 & 0.77 & 0.02 & 0.00 & 0.00 & 0.21 \\ 0.01 & 0.00 & 0.19 & 0.72 & 0.05 & 0.00 & 0.03 \\ 0.00 & 0.01 & 0.01 & 0.03 & 0.95 & 0.00 & 0.00 \\ 0.50 & 0.05 & 0.00 & 0.00 & 0.00 & 0.45 & 0.00 \\ 0.00 & 0.00 & 0.45 & 0.01 & 0.00 & 0.00 & 0.54 \end{pmatrix}$$

Now we use the PCCA algorithm [10, 11] as explained in the discussion of problem (P4), which yields into aggregation of the hidden states S_1, S_2, S_6 and S_3, S_4, S_7 into two new hidden aggregated states. This aggregation results in the following 3×3 transition matrix between the aggregated hidden states:

$$\mathrm{aggr}(\hat{P}) = \begin{pmatrix} 0.99 & 0.00 & 0.01 \\ 0.01 & 0.95 & 0.04 \\ 0.01 & 0.00 & 0.99 \end{pmatrix}$$

As it can be seen from the diagonal, all three aggregated states are metastable in the sense that there is a high probability to stay within a state the next

time step. We thus end up with three hidden states, each being the combination of some of the 7 original states. Therefore, we can simply aggregate the 7-state Viterbi path into a new 3-state Viterbi path. This path is shown in Fig. 6 by coloring all states in the observation sequence of Φ, Ψ, and Ω according to their assignment to one of the hidden aggregated states. We observe that the conformations and the associated jumps between different phases of the simulation perfectly agree with the rare transitions in Ω, while Fig. 7 shows the overlap of conformations in the (Φ, Ψ) plane.

Torsion dynamics of a B-DNA oligomer in water. Finally, we consider a time series originating from a long term molecular dynamics (MD) simulation of a 15-AT B-DNA oligonucleotide with the sequence $GT(AT)_6C$, see Fig. 8. The 100 ns AMBER96 force field simulation with explicit water and potassium ions was conducted in the group of J. Maddocks (EPFL), the computation protocol is described in detail in [5]. The MD simulation delivers a time series of the cartesian coordinates of all atoms (about 23.000 atoms, including solvent). The MD trajectory was sampled every picosecond, which is rather large compared to the fastest time scales (1 fs) of the underlying dynamics, resulting in a time series of 10^5 data points. As observables we have chosen the physically motivated projections of the original data onto a sets of 84 torsion angle coordinates [24], arising from six backbone torsion angles for each of the fourteen base-pairings (junctions) in the sequence, see Fig. 8.

Fig. 9 (top panel) shows, as an example, three of the extracted torsion angles of a single base-pair junction. As it can be seen the torsion angles exhibit a metastable behavior with sharp transitions between the metastable states. We started our analysis, by analyzing each of the 84 dimensional torsion angle time series separately, i.e. doing an HMM analyis assuming three hidden states for *each* of the torsion angles and aggregating afterwards with PCCA. Thus, we get, by application of our HMM-PCCA analysis, a set of 84 aggregated Viterbi paths describing the conformational change between the metastable states (for examples, see Fig. 9). As a threshold value for

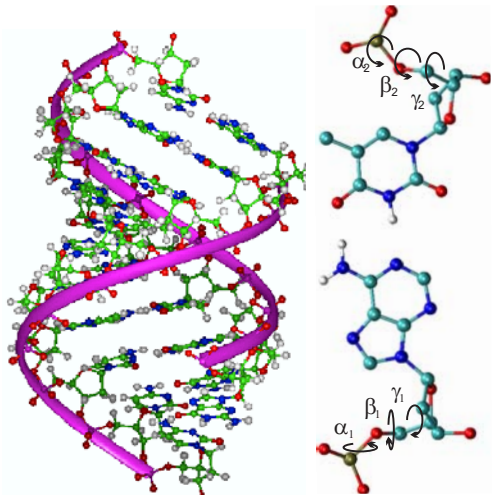

Figure 8: Left: Illustration of the 15-AT B-DNA oligonucleotide in atomic resolution. The attached violet (grey) stings indicate the backbones of the helix (strands). The molecular dynamics simulation referred to herein includes solvent (water and counter-ions) which is not shown here. Right: A single base pair junction (adenine-thymine) attached to a backbone piece. Marked are the six (backbone) torsion angles which we used as observables in the analysis.

the PCCA analysis, i.e. the all eigenvalues of the transition matrix above the threshold are supposed to indicate metastability, we used 0.97 (we tested our analysis with less strict threshold values and obtained similar results).

The full set of 84 aggregated Viterbi paths is a coarse–grained description of the original time series and encapsulates the dynamical information about the process of the conformation change in the full 84-dimensional torsion angle space. As explained in Problem (P5) we extract a global Viterbi path of the 84 aggregated paths by superposition of all paths, i.e. we assign to each different pattern $(x_t^{(1)}, \ldots, x_t^{(84)}), 1 \leq t \leq 100000$ a different global state. This yields 496 different global states, which is quite impressive, as theoretically there could be about $4 \cdot 10^{11}$ different states.

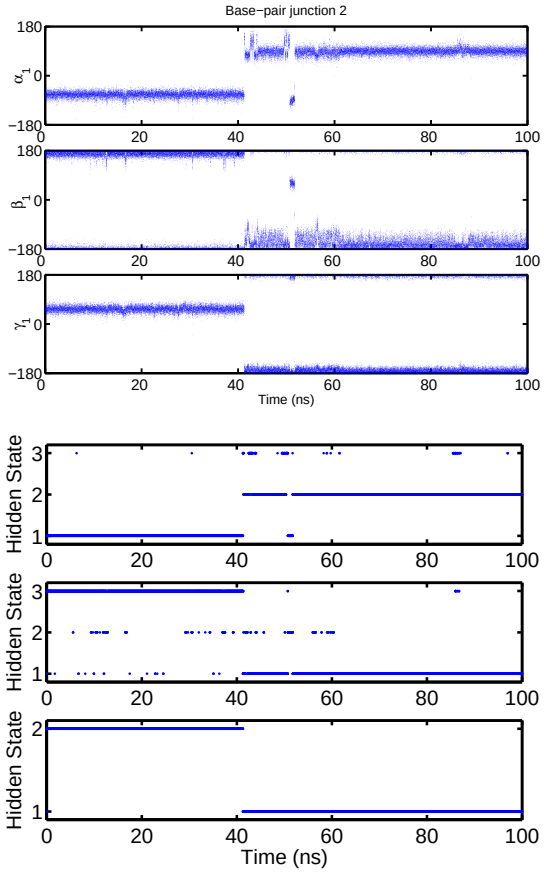

Figure 9: Time series of the 1st strand α, β, γ-backbone torsion angles for junction 2 (upper panel). Lower panel: Corresponding aggregated Viterbi paths as derived from the independent HMM analysis of these three torsion angle time series with an eigenvalue threshold of 0.97.

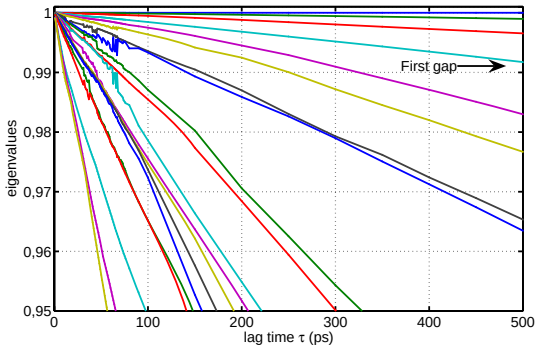

Figure 10: The largest twenty eigenvalues of the global transition matrix as a functions of the lag time τ along the respective global Viterbi path.

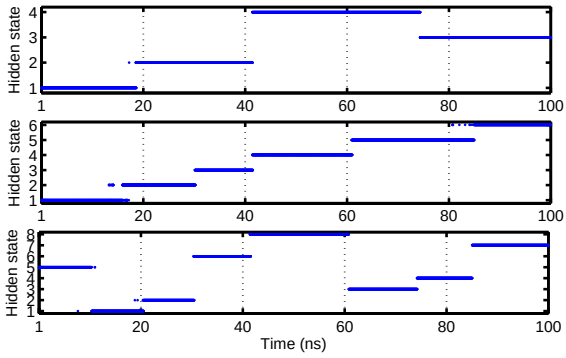

Figure 11: Different global Viterbi paths, resulting from the first four, resp. six and eight (top to bottom) eigenvalues of the global transition matrices.

Upon this global state space we set up a transition matrix by counting the transitions between the global states in the given time series. Fig. 10 represents the dependence of the eigenvalues of the transition matrix as a function of the lag time τ used to count transitions, i.e. lag time $\tau = l\Delta t$ means that the transition matrix counts l-step transitions between instances t and $t + l\Delta t$ along the given time series. Increasing τ means decreasing correlation between successive time steps and therefore a more informative spectrum. As it can be seen, the first eigenvalue gap can be identified after the first four dominant eigenvalues for all θ, which indicates a presence of four metastable sets in 84 dimensional space of torsion angles. However, for a more detailed level of description a choice of 6 or 8 eigenvalues would be also reasonable.

In Fig. 11 we present the aggregated global Viterbi path based upon the assumption of 4 metastable states, in comparison with the global Viterbi path based upon the assumption of 6 and 8 metastable sets. It can be seen how the assumption of more metastable states hierarchically resolves the time series. In Fig. 12 we visualize two metastable states, as obtained from the analysis assuming 4 methastable states.

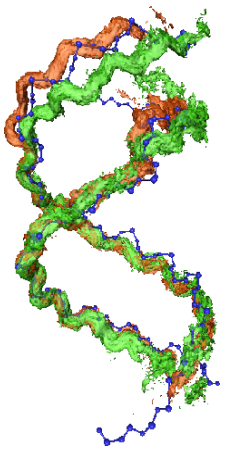

Figure 12: Flexibility plots of 2 (out of 4) obtained global states for the DNA molecule. The plot is obtained by aligning the molecule wrt. the backbone (blue) and coloring the regions where to find the backbone with high probability. The different states are distinguished by different coloring (red and green, belonging to index 1, resp 2, in Fig. 11). As it can be seen, the two shown conformations differ in that the strands depart at the end. Visualization based on AMIRA-software [2].

Conclusion

We presented an HMM algorithm with von Mises distributed observations that can be employed to identify metastable conformations from molecular dynamics time series based solely on observation of some torsion angles. An outstanding feature of the HMM approach is a reliable detection of biomolecular conformations even if the conformations strongly overlap in the chosen set of torsion angles. Since computational complexity only depends on the number of torsion angles to be analyzed and the number of hidden states in the model, the presented technique is in principle applicable to arbitrary high dimensions. Yet, for a high number of hidden states the EM step of the algorithm will become computationally expensive, and for an increasing number of torsion angles the complexity of the likelihood landscape (i.e., parameter space) is hard to judge. Therefore, an additional scheme for use with high dimensions has been devised. Its performance has been illustrated by application to a DNA oligomer.

References

[1] A. Amadei, A. Linssen, and H. Berendsen. Essential dynamics on proteins. *Proteins*, 17, 1993.

[2] Amira—advanced visualization, data analysis and geometry reconstruction, user's guide and reference manual. Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Indeed–Visual Concepts GmbH and TGS Template Graphics Software Inc., 2000. S

[3] L. E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.

[4] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, 41(1):164–171, 1970.

[5] D.L Beveridge, G. Barreiro, K.S Byun, D.A Case, T.E Cheatham III, S.B Dixit, E. Giudice, F. Lankas, R. Lavery, J.H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K.M. Thayer, P. Varnai and M.A. Young, Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonuleotides. I.Research Design and Results on d(CpG) Steps. Biophysical Journal, 87 (2004), 3799-3813

[6] A. Brass, B. J. Pendleton, Y. Chen, and B. Robson. Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics. *Biopolymers*, 33:1307–1315, 1993.

[7] F. Cordes, M. Weber, and J. Schmidt-Ehrenberg. Metastable conformations via successive Perron cluster analysis of dihedrals. ZIB-Report 02-40, Zuse Institute Berlin, 2002.

[8] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Num. Anal.*, 36(2):491–515, 1999.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39(1):1–38, 1977.

[10] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.

[11] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. ZIB-Report 03-19, Zuse Institute Berlin, 2003.

[12] R. Elber and M. Karplus. Multiple conformational states of proteins: A molecular dynamics analysis of Myoglobin. *Science*, 235:318–321, 1987.

[13] R. Elliot, L. Aggoun, and J. Moore. *Hidden Markov Models: Estimation and Control*. Springer, 1995.

[14] H. Frauenfelder, P. J. Steinbach, and R. D. Young. Conformational relaxation in proteins. *Chem. Soc.*, 29A:145–150, 1989.

[15] D. R. Fredkin and J. A. Rice. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. Roy. Soc. Lond. B*, 249:125–132, 1992.

[16] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. Predicting transmembrane protein topology with a Hidden Markov Model: Application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, 2001.

[17] E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, 7:306–317, 2001.

[18] K. V. Mardia. *Statistics of Directional Data*. Academic Press, New York, 1972.

[19] E. Meerbach, E. Dittmer, I. Horenko and C. Schütte Multiscale Modelling in Molecular Dynamics: Biomolecular Conformations as Metastable States In M. Ferrario, G. Ciccotti, and K. Binder, editors, *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology. Volume I*, Lecture Notes in Physics 703: 475–497, 2006.

[20] E. Meerbach, C. Schütte, I. Horenko and B. Schmidt. Metastable conformational structure and dynamics: Peptide between gas phase via clusters and aqueous solution. In O. Kühn and L. Wöste, editors, *Analysis and Control of Ultrafast Photoinduced Reactions*, Series in Chemical Physics 87: 798–808, 2007.

[21] G. U. Nienhaus, J. R. Mourant, and H. Frauenfelder. Spectroscopic evidence for conformational relaxation in Myoglobin. *PNAS*, 89:2902–2906, 1992.

[22] L. R. Rabiner. A tutorial on Hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, 1989.

[23] L. R. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[24] W. Saenger. Principles of Nucleic Acid Structure. Springer Verlag, New York, 1984.

[25] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.

[26] C. Schütte and W. Huisinga. Biomolecular Conformations can be Identified as Metastable Sets of Molecular Dynamics. In P. G. Ciaret and J.-L. Lions, editors, *Handbook of Numerical Analysis X*, Special Volume Computational Chemistry, pages 699–744, 2003.

[27] H. Shatkay. Learning hidden Markov models with geometrical constraints. In *Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence*, pages 602–611, 1999.

[28] H. Shatkay and L. P. Kaelbling. Heading in the right direction. In *Proc. of the 15th Int. Conf. on Machine Learning*, pages 531–539, 1998.

[29] K. M. Thayer and D. L. Beveridge. Hidden Markov models from molecular dynamics simulations on DNA. *Proc. Natl. Acad. Sci.*, 99(13), 2002.

[30] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. on Inform. Theory*, IT-13:260–269, 1967.