Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

F. CORDES , M. WEBER , J. SCHMIDT-EHRENBERG

Metastable Conformations via successive Perron-Cluster Cluster Analysis of dihedrals

ZIB-Report 02-40 (November 2002)

Metastable conformations via successive Perron-Cluster Cluster Analysis of dihedrals

F. Cordes, M. Weber , J. Schmidt-Ehrenberg

Abstract

Decomposition of the high dimensional conformational space of biomolecules into metastable subsets is used for data reduction of long molecular trajectories in order to facilitate chemical analysis and to improve convergence of simulations within these subsets.

The metastability is identified by the Perron-Cluster Cluster Analysis of a Markov process that describes the thermodynamic distribution. A necessary prerequisite of this analysis is the discretization of the conformational space. A combinatorial approach via discretization of each degree of freedom will end in the so called "curse of dimension".

In the following paper we analyze Hybrid Monte Carlo simulations of small, drug-like biomolecules and focus on the dihedral degrees of freedom as indicators of conformational changes. To avoid the "curse of dimension", the projection of the underlying Markov operator on each dihedral is analyzed according to its metastability. In each decomposition step of a recursive procedure, those significant dihedrals, which indicate high metastability, are used for further decomposition. The procedure is introduced as part of a hierarchical protocol of simulations at different temperatures. The convergence of simulations within metastable subsets is used as an "a posteriori" criterion for a successful identification of metastability. All results are presented with the visualization program AmiraMol.

Key words. metastability, Perron-Cluster Cluster Analysis, curse of dimension, Hybrid Monte Carlo, significant dihedrals

Mathematics subject classification. 65U05, 62H30.

1 Introduction

A molecule can theoretically adopt an infinite number of spatial states, so called conformations. But the probabilities of molecular states are not evenly distributed. Due to the Hamiltonian of a molecule, the state space can be divided into regions of high metastability, in which a molecule tends to stay for a long time, before it changes into another metastable conformation. Conformational analysis aims to identify these regions under the following conditions:

- Conformational space should be divided into a finite number of subsets,
- that are distinguishable according to physical properties, and

- cover almost the whole conformational space.

Cluster analysis [12, 16] of sampling data or exhaustive search of local minima on the energy hypersurface [1] are commonly used for conformational analysis. The latter method is generally combined with clustering to build conformational families and avoid redundancy. The number of conformers depends on the number of energy minima, which is normally very large [9], and the similarity measure of the clustering. Conventional methods are unable to decide a priori, how many conformational subsets are to be expected. Conventional clustering uses the geometry of molecules to distinguish between different clusters. In this context, a certain number of conformers can form a Voronoi coverage of the space. But the borders between Voronoi cells of different clusters do not reflect the cluster membership very well, at least, if the number of generated conformers is low.

Metastability has been introduced as a measure for the dynamical properties of a conformer [4, 3, 6]. The common property of metastable conformations are low transition probabilities between different conformers. Metastable conformations can be identified within equilibrium distributions of states, e.g. the canonical distribution $\mathcal{Q}(q)$ in position space. In contrast to exhaustive search techniques, these conformations are based on statistical sampling data and are connected with thermodynamical weights. In contrast to conventional clustering, the number of metastable conformations can be identified a priori as the number of almost invariant sets $C_i \subset \Omega$, $i = 1, \ldots s$ in the conformational space Ω [6].

A necessary prerequisite for the identification of metastability is a spatial discretization of Ω into N pairwise disjoint boxes $A_i \subset \Omega$, $i = 1, \ldots, N$. One can determine transition probabilities T_{ij} from A_i to A_j by counting transitions between the states of the equilibrium distribution. If the transitions are realized by Hybrid Monte Carlo, the $N \times N$-transition matrix $T = (T_{ij})$ can be interpreted as a discretization of a Markov operator P^τ obtained from the Perron-Frobenius operator by momenta averaging with respect to the given ensemble [4, 3]. Since we are interested in the equilibrium state, the detailed balance condition holds. T is self-adjoint [4], i.e its spectrum is real valued and via Perron-Cluster Cluster Analysis (abbreviated PCCA according to [5]) of T for a given discretization $A_1, \ldots, A_N$ of Ω one can find the number s of conformations $C_1, \ldots, C_s$ and s pairwise disjoint index sets $I(i) \subset \{1, \ldots, N\}$ such that $C_i = \bigcup_{j \in I(i)} A_j$.

A main advantage of the metastability analysis is used in more efficient sampling methods [8]. Since critical slowing down of simulations, either Molecular Dynamics or Monte Carlo, is strongly connected with metastability, the decomposition of conformational space can be used to uncouple the sampling into simulations within separate subsets. Combined with a hierarchical temperature or potential embedding, the efficiency of the Hybrid Monte Carlo simulation is increased, because rare transitions between metastable conformations, which can be identified on a higher temperature level, are avoided on lower temperatures.

The critical point of metastability analysis is the spatial discretization of Ω into N pairwise disjoint boxes A_i. In former publications we have shown [15], that uniform discretization suffers from the "curse of dimension". Even if one concentrates only on dihedral angles as the indicators of most significant conformational changes, the division of every angle into for example only three intervals leads for N degrees of freedom to 3^N boxes, thus to exponentially increasing computational costs. Identification of essential degrees of freedom [2, 14] can reduce costs, but is not trivial for multimodal distributions.

Neural networks describe another method, we have applied successfully to reduce the dimension of conformational space [17]. The self organizing box maps (SOBM) designed for this purpose are able to cover almost the complete space according to the underlying distribution [10]. But the resolution of the network is high at the centers of metastability, where most of the data trained the network. In transition regions the density of data and therefore the resolution of the network is low, which leads to an unprecise separation between clusters.

So far, the introduced algorithms have clarified the following points: Discretization of conformational space has to be fine enough to identify and separate metastable subsets, but it has to be coarse enough to avoid the "curse of dimension". Therefore, we first reformulate the identification of high metastability into an optimization of autocorrelation to emphasize the meaning of resolution of the discretization. The reformulation leads to an optimal decomposition of conformational space in terms of membership functions instead of sign structure of eigenvectors [19, 6]. The practical implication of this approach is an algorithm which identifies transition regions via fine discretization at least for projections of the Markov Chain on specific dihedrals and decomposes the whole space via coarse discretization. Finally, some numerical results for small molecules are presented.

2 Methods

Maximizing the autocorrelation. In the following section we will reformulate the problem of identifying high metastability into the optimization of autocorrelation.

We will first look at the simplest case where we have the given Markov operator P^τ and search for a decomposition of Ω into metastable sets $C, C^c \subset \Omega$.

Metastability is defined as the probability $M(C)$ that a trajectory starting in C ends up in C after time τ [4, 3]:

$$M(C) := \frac{\mathbb{E}[\mathbb{1}_C \, P^\tau \mathbb{1}_C]}{\mathbb{E}[\mathbb{1}_C]}, \tag{1}$$

where

$$\mathbb{E}[v] := \int_\Omega v(q)\, \mathcal{Q}(q)\, dq =< \mathbb{1}_\Omega, v >_{\mathcal{Q}} .$$

In equation (5) we will show a relation between autocorrelation $\delta(\mathbb{1}_C)$ for the characteristic function $\mathbb{1}_C$ and the sum $M(C) + M(C^c)$.

For the propagator P^τ we define the autocorrelation value $\delta(v)$

$$\delta(v) := \frac{\mathbb{E}[(v - \mathbb{E}[v])(P^\tau v - \mathbb{E}[P^\tau v])]}{\mathbb{E}[(v - \mathbb{E}[v])^2]}, \tag{2}$$

where $v \in \mathcal{X}$ and $\mathcal{X}$ is the set of membership functions $v : \Omega \to [0, 1]$ with $\mathbb{E}[(v - \mathbb{E}[v])^2] > 0$. In particular $v \in \mathcal{X}$ is a non-constant membership function, cf. [19].

To illustrate the relation between autocorrelation and metastability take a characteristic function $\mathbb{1}_C$. For the function values of $\mathbb{1}_C$ the Markov chain defined by the HMC method jumps inside and between C and C^c and thus is a sequence of 0 and 1. The more this sequence is autocorrelated the more metastability $M(C) + M(C^c)$ we expect for the sets C and C^c.

Further we will proof that the second largest eigenvalue λ_2 of the propagator P^τ, if $1 = \lambda_1 > \lambda_2$ are simple eigenvalues and lie inside the discrete spectrum of P^τ, is the maximal autocorrelation value $\lambda_2 = \delta^*$.

Proof: Since P^τ is self-adjoint and $P^\tau \mathbb{1}_\Omega = \mathbb{1}_\Omega$, we have

$$\mathbb{E}[v] = \mathbb{E}[P^\tau v]. \tag{3}$$

$\delta(v)$ for $v \in \mathcal{X}$ can be transformed with (2) and (3) into

$$\delta(v) = \frac{\mathbb{E}[vP^\tau v] - (\mathbb{E}[v])^2}{\mathbb{E}[v^2] - (\mathbb{E}[v])^2}. \tag{4}$$

Define the set $\tilde{\mathcal{X}} \subset \mathcal{X}$ of all possible characteristic functions $\mathbb{1}_C : \Omega \to \{0, 1\}$ with $0 < \mathbb{E}[\mathbb{1}_c] < 1$.

Since $\mathbb{1}_C^2 = \mathbb{1}_C$ we get from (4) and (1):

$$M(C) = \delta(\mathbb{1}_C)(1 - \mathbb{E}[\mathbb{1}_C]) + \mathbb{E}[\mathbb{1}_C],$$

and with $\mathbb{E}[\mathbb{1}_C] = 1 - \mathbb{E}[\mathbb{1}_{C^c}]$ and $\delta(\mathbb{1}_C) = \delta(\mathbb{1}_{C^c})$, because autocorrelation δ is invariant with respect to scaling and translation, we get

$$M(C) + M(C^c) = 1 + \delta(\mathbb{1}_C). \tag{5}$$

Thus maximizing metastability is equivalent to maximizing δ.

Let δ^* be the maximal value of δ for all $v \in \mathcal{X}$. Since again autocorrelation δ is invariant with respect to scaling and translation, maximization can be transformed into the space of functions u with expectation value $\mathbb{E}[u] = 0$ and normalization $\mathbb{E}[u^2] = 1$:

$$\delta^* = \max_{\mathbb{E}[u]=0, \mathbb{E}[u^2]=1} \mathbb{E}[uP^\tau u] = \max_{<\mathbb{1}_\Omega, u>_\mathcal{Q}=0, <u,u>_\mathcal{Q}=1} < u, P^\tau u >_\mathcal{Q} . \quad (6)$$

For the self-adjoint operator P^τ this optimization problem is the Rayleigh-Ritz term for the 2^{nd} largest eigenvalue λ_2 of P^τ. $\mathbb{E}[u] = 0$ means that u and $\mathbb{1}_\Omega$ are orthogonal, whereas $\mathbb{1}_\Omega$ is the eigenfunction corresponding to the largest, simple eigenvalue $\lambda_1 = 1$ of P^τ (see [4]). And we see that the second eigenvalue of P^τ is the maximal possible autocorrelation, $\lambda_2 = \delta^*$. $\square$

Equation (6) has shown, that the second eigenfunction of P^τ maximizes the autocorrelation. Since autocorrelation is invariant with respect to scaling and translation, an optimal membership function $v^* \in \mathcal{X}$ can be generated by these transformations on the second eigenfunction u of P^τ. In other words: **$\mathbf{v}^*$ can be expressed by a linear combination of the Perron eigenfunctions $\mathbb{1}_\Omega$ and u.** We use this idea for the identification of the almost invariant sets in the next section.

In our algorithm the transition matrix T is a discretization of the propagator P^τ and the eigenvectors of T are piecewise constant discretizations of the eigenfunctions of P^τ. Therefore, the optimal membership function v is approximated with a piecewise constant function $\tilde{v}$ like in Fig. 1. The autocorrelation value of $\tilde{v}$ is lower than $\delta(v)$ and hence the Perron-Cluster of T, which should be near 1, moves away from 1. Mainly in the transition region the difference between $\tilde{v}$ and v is significant. The transition regions of realistic molecules have a non-negligible statistical weight. To get a good approximation of v and therefore a high autocorrelation value we need a fine discretization of Ω.

Decomposition of Ω with Inner Simplex Algorithm. The idea of the last section can be generalized to s metastable conformations, and a discretization of Ω into n boxes. The system of membership functions $\{v_1, \ldots, v_s\} \subset \mathbb{R}^n$ can be expressed as linear combinations of the eigenvectors $\mathbb{1}_\Omega, u_2, \ldots, u_s \in \mathbb{R}^n$ corresponding to the Perron-Cluster of T.

Let

$$U := (\, \mathbb{1}_\Omega \, u_2 \, \ldots \, u_s \,), \quad V := (\, v_1 \, \ldots \, v_s \,)$$

denote the $n \times s$-matrix of eigenvectors and the $n \times s$-matrix of membership functions respectively. Then we have to find an invertible $s \times s$-matrix A such that $V = UA$.

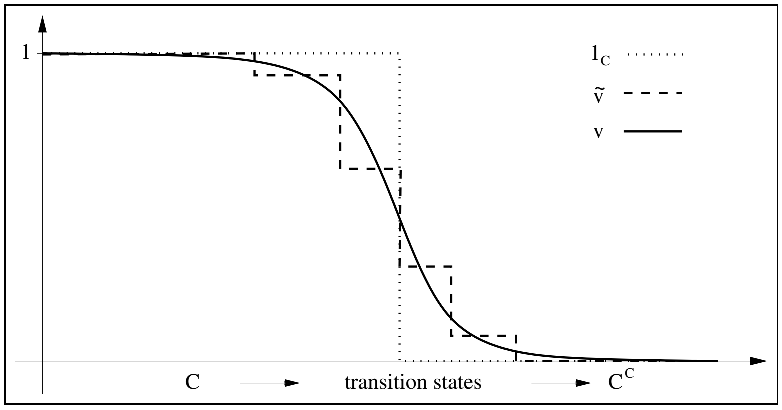

Figure 1: **Transition region between metastable conformations:** Characteristic $1\!\!1_C$ versus membership function v and discretization $\tilde{v}$ of v.

There are many possible linear transformations, but only some are leading to so called *indecomposable cluster solutions*. If we assume, that there is only one unique indecomposable cluster solution, then we can apply the *inner simplex algorithm* to identify this matrix A (cf. [19]).

The rows of U define a set Y of n points $y_i \in \mathbb{R}^s, i = 1, \ldots, n$. By the following algorithm we determine s points out of Y, which are the rows of A^{-1}. This will give us A.

1. Find two points $y_1, y_2 \in Y$ having maximum Euclidean distance with regard to every pair of points in Y.

2. For $k = 3, \ldots, s$ find a point $y_k \in Y$ maximizing the Euclidean distance between point y_k and the hyperplane defined by $y_1 + span\{(y_2 - y_1), \ldots, (y_{k-1} - y_1)\}$.

The s rows of A^{-1} span an inner simplex of Y. After transformation $V = UA$ one obtains the desired membership functions.

Going back from V to characteristic functions, we assign each discretization box $A_i \subset \Omega, i = 1, \ldots, n$ to the cluster $k \in \{1, \ldots, s\}$ where it attains the maximum value with regard to the computed membership functions:

$$V_{ik} = \max_{j=1,\ldots,s} V_{ij}.$$

Successive Perron-Cluster Cluster Analysis of dihedrals. The PCCA of T is improved, if a fine discretization of conformational space is used. But to make T computable, we are restricted to a coarse discretization. The following algorithm is an approach to solve this paradoxon.

1. For every dihedral:

- Project the conformational space represented by all samples of a Hybrid Monte Carlo simulation on the dihedral,
- take a fine, uniform discretization,
- construct the transition matrix T and
- calculate second eigenvalue.

2. select dihedral with highest autocorrelation $\delta > \delta_{min}$

   - apply PCCA to partition the projection of the selected dihedral
   - use the resulting partitioning to decompose the space into metastable subsets.

3. repeat 1 and 2 for all subsets which contain a reasonable number of transitions and a dihedral with autocorrelation $\delta > \delta_{min}$.

4. apply PCCA to the coarse decomposition of the conformational space.

The algorithm generates a hierarchy of decompositions (Fig. 2), which are generated by othogonal cuts on dihedrals.

It is worth to mention, that the dihedrals are not necessarily the cause of metastability. But significant conformational change effects a change in one or more dihedrals.

Hierarchical protocol for temperature embedding. The decomposition of conformational space can be used to start independent HMC simulations within the generated subsets of high metastability. These simulations should not suffer from the critical slowing down, if all metastable conformations have been identified. Therefore, convergence of resimulation serves as an "a posteriori criterion" for the quality of metastability analysis. Moreover, resimulation is combined with a hierachical embedding protocol introduced in previous publications [8]. Starting the simulation at high temperature facilitates the crossing of energy barriers and improves convergence. After analysis of metastability, resimulation is started as a "bridge sampling" between the initial high and a new, lower temperature. The parameters for bridge sampling are automatically generated by the previous simulation [7]. "Bridge sampling" can still overcome barriers at high temperature but delivers also information for the distribution at low temperature. In the present paper we perform initial simulations at 1500K and bridge sampling between 1500K→1000K, 1000K→600K and 600K→300K. All results can be used within a reweighting scheme to produce the thermodynamically correct distribution at the lowest temperature [8].

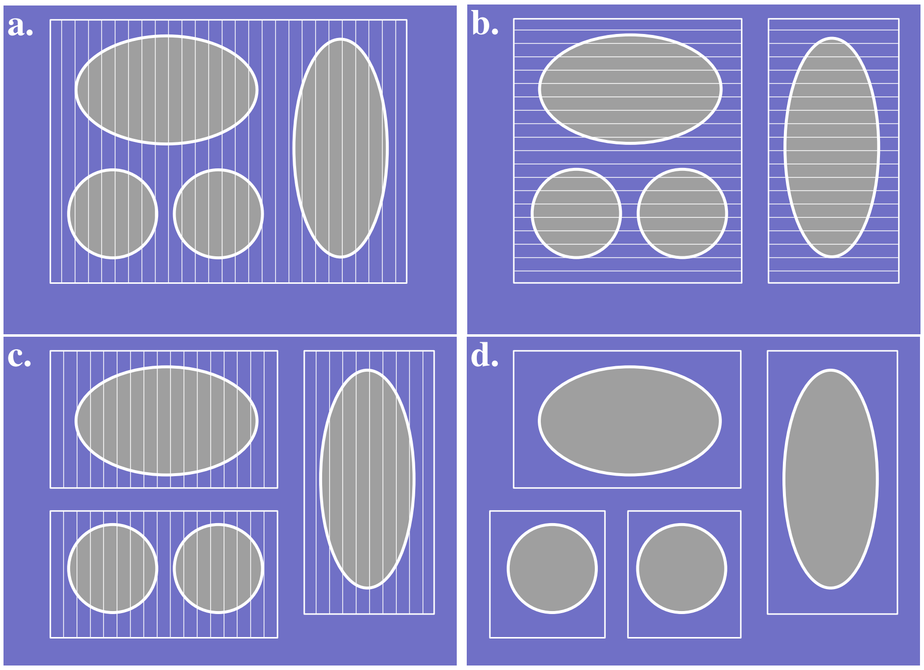

Figure 2: **Algorithmic scheme for successive PCCA of dihedrals:** Four metastable regions are drawn as ellipses in a 2-dimensional dihedral space. Thin lines show the fine discretization of the dihedral, which indicates highest metastability according to the second eigenvalue of the corresponding transition matrix. Figures a. to d. illustrate the alternation of fine discretization followed by coarse decomposition.

3 Numerical Results

Parameter for simulation and analysis. We have performed HMC-Simulations for small biomolecules. Results are presented for the uncharged form of the amino acid glycine and the tripeptide ala-gly-gly with charged termini. The molecules were parameterized by the MMFF force field [13]. The propagation of dynamics within the HMC simulation was performed with a timestep of 1.4 femtoseconds. The length of the MD trajectory was chosen randomly between 40 to 80 integration steps. 5 independent Markov chains were started in every subset of metastable conformations. Convergence of HMC was controlled by the Gellmann and Rubin criterion, which evaluates the mixing of the chains [11]. All simulations have generated between 20000 and 200000 molecular states to reach convergence. The temperatures of the hierarchical embedding protocol were 1500K, 1000K, 600K and 300K. All simulations were performed in vacuum. The discretization per dihedral was performed under a resolution of $5.0°$, which results in 72 boxes in the intervall $[-180.0°, +180.0°]$. A cluster of eigenvalues near 1 is

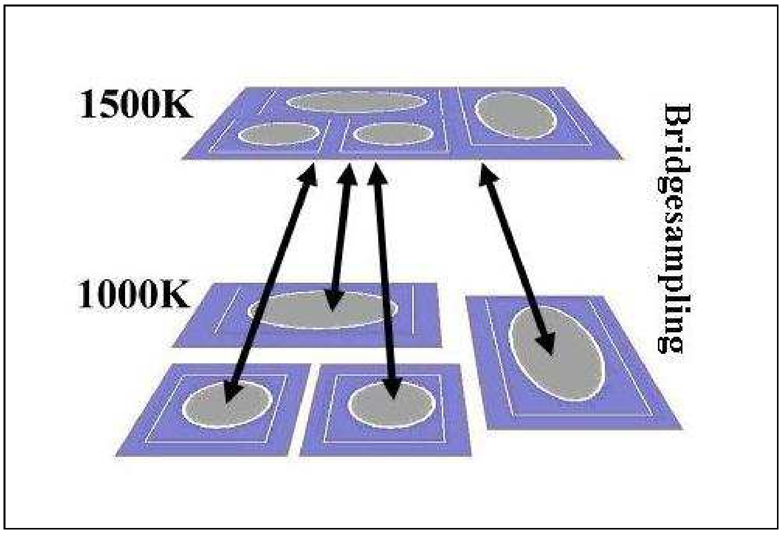

Figure 3: **Hierarchical simulation protocol:** After decomposition, the metastable subsets of the conformational space are sampled independently at a lower temperature level. Two temperature level are connected via bridge sampling, i.e. the HMC simulation samples a generalized distribution, which is constructed from the canonical distributions of both temperatures [7].

identified as a Perron-Cluster if

- the second eigenvalue is greater 0.9,
- the difference between two successive eigenvalues < 1.0 in the cluster is smaller than the difference between the lowest eigenvalue of the cluster and the next eigenvalue of the spectrum,
- the lowest eigenvalue of the cluster is greater than 0.8.

Results glycine. Figure 4 shows 3 conformations out of 6 for glycine at 300K. The other 3 conformations are chemically identical but differ mathematically due to an exchange of hydrogens at the amino group. The most stable conformation Fig. 4 c is stabilized by a hydrogen bond between N- and C- terminus. The corresponding energy minimum is responsible for the identification of the first metastability at 1500K, reflected in the high second eigenvalue 0.963 of dihedral α (Tab. 1). Although metastable regions of this small molecule should be clearly separable, the autocorrelation still depends on the coarseness of the discretization which results in a lower second eigenvalue of 0.953 for the decomposition into only two subsets. The identified conformations at 1500K serve as starting points for the subsequent bridge samplings between 1500K and 1000K. The PCCA of dihedrals

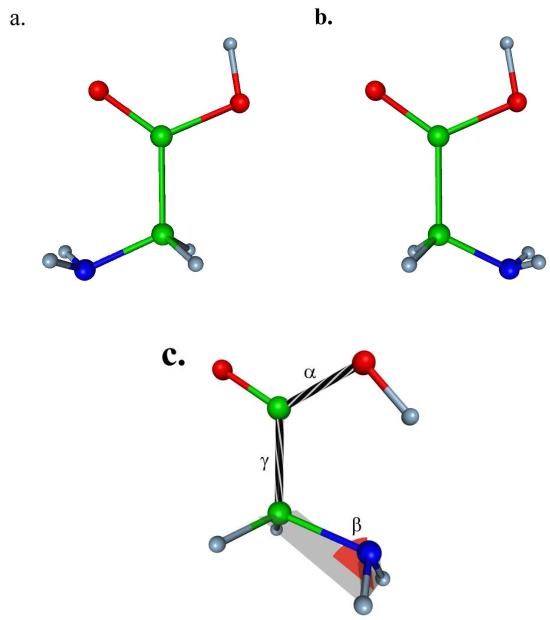

Figure 4: **Mean molecules calculated from 3 metastable subsets of the uncharged form of the amino acid glycine:** The thermodynamical weights of conformation a,b,c at 300K are $0.48\%, 0.29\%, 99.23\%$

reveals metastability of the improper dihedral β, which is connected to the inversion of the amino group. The energy barrier for inversion is about 20 kJ/mol [18] and significant higher than the thermal energy at 300K. Crossing this barrier at 300K is nearly impossible, but at 1000K rare transitions occur. Each of the metastable subsets generated with respect to β is further splitted according to Another metastability of the dihedral γ. The conformational space decomposes into a coarse discretization four subsets, but the fina PCCA on the coarse discretization offers only two metastable conformations. A visual inspection of the coarse decomposition clarifies, that only the inversion contributes to the split as the dominant metastability. Within the following bridge sampling between 1000K and 600K the metastability indicated by γ increases (Tab. 1) and leads to the final split, which results in conformations **a** and b of Fig. 4. Since all results of the hierarchical simulation protocol are connected via reweighting between successive temperature level, thermodynamical weights can be calculated after the last bridge sampling between 600K and 300K. The set of metastable conformations is dominated by conformation **c** of Fig. 4 due to the strong hydrogen bond between N- and C-terminus.

T[K]	coarse spectrum	coupling matrix	spectrum α	spectrum β	spectrum γ
1500	1.000 0.953	0.985 0.015 0.031 0.969	1.000 0.963 0.373		
1000	1.000 0.894 0.837	0.946 0.054 0.052 0.948		1.000 0.911 0.318	1.000 1.000 0.901 0.902 0.740 0.729
1000	1.000 0.967 0.875	0.983 0.017 0.017 0.983		1.000 0.971 0.350	1.000 1.000 0.919 0.916 0.630 0.641
600	1.000 0.942	0.969 0.031 0.027 0.973			1.000 0.969 0.779
600	1.000 0.944	0.971 0.029 0.026 0.974			1.000 0.968 0.773

Table 1: **Hierarchical temperature embedding for glycine :** The spectra of fine discretization of different dihedrals and the resulting coarse discretization are presented.

Results ala-gly-gly. Glycine is a simple example to demonstrate the features of the introduced algorithm. To get an impression, how the method works with more degrees of freedom, we present data of a tripeptide, sampled under the same protocol. The only difference concerns the restriction of the analysis to only those dihedrals, that contain no hydrogens, because we are not interested in chemically identical conformations, which only differ in rotations of the methyl or amino group. Therefore, we concentrate on the remaining backbone dihedrals as descriptors for the essential conformational flexibility of the tripeptide. Figure 5 shows mean molecule representations of 8 out of 11 metastable conformations at 300K. For presentation reasons, we have skipped the remaining 3, which possess thermodynamical weights below 0.05%. The pictures are orientated with the amino and carboxylate group in the background. All conformations show these groups in a close neighborhood, which results from their strong electrostatic interaction. The selected metastable subsets reflect the global conformational changes indicated by ϕ, ψ transitions of the backbone. Again, only a few conformations (a,b) dominate the thermodynamic distribution. Since, the number of metastable conformations is too high to reproduce the history of the sampling in a table, we summarize only some observations. The first run at 1500K decomposes into 2 subsets weighted with 90.0% and 10.0% respectively. After bridge sampling down to 1000K, these conformations are

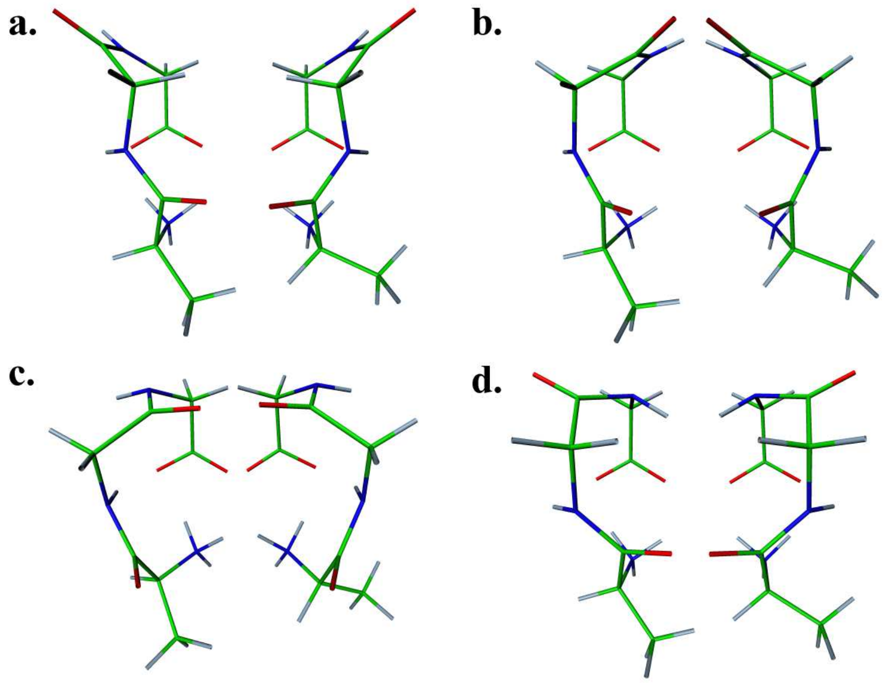

Figure 5: **Mean molecules calculated from 8 metastable subsets of the peptide ala-gly-gly:** The molecules are arranged in pairs to emphasize the partial symmetry of the conformations. The thermodynamical weights of the conformations in a,b,c, and d at 300K are 63.07% and 33.18%, 1.52% and 1.66%, 0.19% and 0.08%, 0.09% and 0.05%

further separated into 12 and 3 subsets each. The sampling down to 600K results in the final splitting into 22 clusters, which means, that only a few of the 15 conformations at 1000K decompose further. The analysis of the final 22 metastable conformations exhibits 11 pairwise identical conformations, which only differ in a 180.0° rotation of the carboxylate group.

Perron-Cluster derived from fine discretization per dihedral are comparable to the results of glycine. In general, gaps are clearly observable and cluster analysis generates splittings into 2 to 5 intervalls. Perron-Cluster from the coarse discretization are denser, at least in cases, in which the spaces decomposes into a very high number of subsets. For one coarse discretization derived from a 1000K→600K sampling we get the following spectrum:

1.00000	0.999999	0.999549	0.998979	0.998138
	0.998109	0.998104	0.997246	...

The weights of the corresponding subsets are very low. A visualization of these subsets suggests, that the observed metastabilities are connected with an unsufficient sampling of conformations with low thermodynamical weight. This is a typical effect of importance sampling in Monte Carlo. To circumvent the generation of a lot of artificial metastable subsets with low thermodynamical weights, we have experimented with a heuristic filter, which neglects clusters with a weight below a given threshold of for instance 0.01%.

4 Conclusion

The proposed successive PCCA of dihedrals satisfies the criteria of conformational analysis mentioned in the introduction. Metastable conformations describe a finite number of separable conformational subsets, which cover the whole conformational space, at least with respect to the quality of the underlying sampling. Metastable conformations do not reflect every local minimum of the energy hypersurface, but only those, which are responsible for critical slowing down. Therefore, sampling the whole conformational space with the hierarchical simulation protocol increases the efficiency of the HMC simulation. The uncoupling of simulations is further used for a parallel implementation of the simulation. With the coupling of the different bridge samplings [8] thermodynamical weights of every metastable subsets can be calculated. Therefore, the algorithm provides the opportunity to compare results with experimental data of the occupancy of states e.g. NMR - data.

Improvement of the algorithm has to be concentrated on correlations between dihedrals. If two or more dihedrals are correlated and form a metastability, then an orthogonal cut as described in Fig. 2 can be impossible. A discretization of such dihedrals, which indicate a"correlation metastability" will be missed. Critical slowing down in a subsequent sampling will be the consequence. This "a posteriori" indicator of metastability is time consuming. The molecules we have analyzed so far have not shown this effect. Metastability behind two correlated dihedrals for example in rings has been resolved, because a third dihedral separated the metastable regions.

Another future task concerns symmetrically related conformations, which are not recognizable in the current version of the algorithm. If such conformations are detectable, they should be removed from the simulation hierarchy to save computer time.

Acknowledgement The work of F. Cordes and J. Schmidt-Ehrenberg has been supported by the German Federal Ministry of education and research (grant no. 031U109A/031U209A, Berlin Center for Genome Based Bioinformatics). The authors would like to thank P. Deuflhard for helpful discussions.

References

[1] P. Towbin A. Smellie, S. L. Teig. Poling: Promoting conformational variation. *J. Comp. Chem.*, 16:171–187, 1995.

[2] A. Amadei, A.B.M. Linssen, and H.J.C. Berendsen. Essential dynamics of proteins. *Proteins*, 17, 1993.

[3] W. Huisinga Ch. Schütte, A.Fischer and P. Deuflhard. A direct approach to conformational dynamics based on hybrid monte carlo. *J. Comput. Phys., Special Issue on Computational Biophysics*, 151:pp. 146–168, 1999.

[4] Ch.Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm and Application to Biomolecules*. Habilitation Thesis, Dept. of mathematics and computer science, Free University Berlin, 1998.

[5] P. Deuflhard. From molecular dynamics to conformational dynamics in drug design. In M. Kirkilionis, S. Krömker, R. Rannacher, and F. Toni, editors, *Trends in Nonlinear Analysis*, pages 269–287. Springer, 2002.

[6] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in nearly uncoupled markov chains. *Linear Algebra and its Applications 315*, pages 39–59, 2000.

[7] A. Fischer, F. Cordes, and Ch. Schütte. Hybrid Monte Carlo with adaptive temperature in mixed-canonical ensemble: Efficient conformational analysis of RNA. *J. Comput. Chem.*, 19:1689–1697, 1998.

[8] A. Fischer, C. Schtte, P. Deuflhard, and F. Cordes. Hierarchical uncoupling-coupling of metastable conformations. In T. Schlick and H.H. Gan, editors, *Computational methods for macromolecules: Challenges and applications*, Proceedings of the third international workshop on algorithms for macromolecular modeling, pages 235–261. Springer, 2002.

[9] H. Frauenfelder, S.G. Sligar, and P.G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.

[10] T. Galliat, P. Deuflhard, R. Roitzsch, and F. Cordes. Identification of metastable conformations via self-organized neural networks. In T. Schlick and H.H. Gan, editors, *Computational methods for macromolecules: Challenges and applications*, Proceedings of the third international workshop on algorithms for macromolecular modeling, pages 262–284. Springer, 2002.

[11] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992.

[12] H.L. Gordon and R.L. Somorjai. Fuzzy cluster analysis for molecular dynamics trajectories. *Proteins*, 14:249–264, 1992.

[13] T.A. Halgren. Merck molecular force field.i-v. *J. Comp. Chem.*, 17(5&6):490–641, 1996.

[14] S. Hayward and H.J.C. Berendsen. Systematic analysis of domain movements in proteins from conformational change: New results on citrate synthase and t4 lysozyme. *Proteins*, 30:144–154, 1998.

[15] W. Huisinga, C. Best, R. Roitzsch, C. Schütte, and F. Cordes. From simulation data to conformational ensembles: Structure and dynamic based methods. *J. Comp. Chem.*, 20(16):1760–1774, 1999.

[16] M.E. Karpen, D.J. Tobias, and C.L. Brooks III. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: Analysis of 2.2-ns trajectories of ypgdv. *Biochemistry*, 32:412–420, 1993.

[17] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, 3rd edition, 2001.

[18] J. B. Lambert. *Topics Stereochem.*, 6:19, 1971.

[19] M. Weber and T.Galliat. Characterization of transition states in conformational dynamics using fuzzy sets. Technical report, ZIB-Report 02-12, March 2002.