Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARCUS WEBER AND HOLGER MEYER

ZIBgridfree – Adaptive Conformation Analysis with qualified Support of Transition States and Thermodynamic Weights

ZIB-Report 05-17 (July 2005)

ZIBgridfree – Adaptive Conformation Analysis with qualified Support of Transition States and Thermodynamic Weights

Marcus Weber and Holger Meyer

Zuse Institute Berlin (ZIB), Germany

Abstract. This paper introduces a new algorithm of conformational analysis based on mesh-free methods as described in [18]. The adaptive decomposition of the conformational space by softly limiting functions avoids trapping effects and allows adaptive refinement strategies. These properties of the algorithm makes ZIBgridfree particularly suitable for the complete exploration of high-dimensional conformational space. The adaptive control of the algorithm benefits from the tight integration of molecular simulation and conformational analysis. An emphasized part of the analysis is the Robust Perron Cluster Analysis (PCCA+) based on the work of Peter Deuflhard and Marcus Weber. PCCA+ supports an almost-characteristic cluster definition with an outstanding mapping of transition states. The outcome is expressed by the metastable sets of conformations, their thermodynamic weights and flexibility.

1 Introduction

The function of drug-like molecules often depends on their 3D conformation. Structural analysis is therefore an important task in molecular modelling. Due to interaction with other molecules or due to a heat bath, a molecule undergoes conformational changes. A huge number of algorithms have been invented in order to generate different conformations for a given molecule. Some examples are: Cerius^2 and InsightII by Accelrys Software Inc., Concord and Corina by Molecular Networks GmbH Computerchemie, FANTOM by the Sealy Center for Structural Biology of the university of Texas and many others. These algorithms mostly include a local optimization routine for the minimization wrt. the rough potential energy landscape of the molecule. Different strategies have been developed in order to assure a complete computation of thermodynamically relevent local minimizers. However, a thermodynamically correct weighting of conformations is a severe problem and needs time-consuming samplings of the conformational space, e.g. [7][1][12][13]. In order to link identification and sampling, Deuflhard, Schütte et al. [2][3][16] described a new, consistent concept for conformation analysis, which they called *conformation dynamics*. In conformation dynamics, one is interested in the identification of conformations, their thermodynamical weights and transition probabilities. ZIBgridfree is based on this concept and provides an adaptive and complete sampling in conformational

space. In ZIBgridfree, conformations are defined as *overlapping* densities, i.e. every point of the position space is assinged to each conformation with a certain grade of membership, which additionally allows the direct identification of transition regions.

2 Conformation Dynamics Approach

2.1 Concept of Molecular Conformations

The aims of conformation dynamics are the identification of "conformations", their thermodynamical weights, and their transition behaviour. For an n-atoms molecule, in literature, conformations are often understood as single points in the 3n-dimensional space Ω of the position coordinates. In most cases, conformations are defined as local minima of the energy landscape of a molecule. For the aims of conformation dynamics this definition of conformations is insufficient, because it is impossible to compute thermodynamic weights, especially entropical information, out of single points $q \in \Omega$. According to the flexibility of a molecule, we define conformations as overlapping partial densities in position space, see Fig. 1 and Fig. 3 for a representation of a conformation via volume rendering in 3D cartesian coordinates, see also [14]. Although the minima of both conformations in Fig. 1 are nearly at the same level they do not have the same thermodynamical weight.

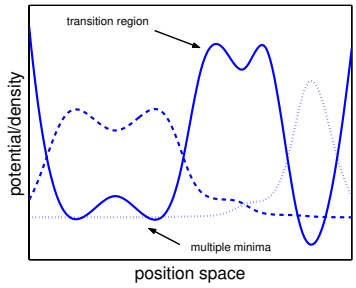

Fig. 1. For the aims of conformation analysis, conformations are not defined as points at local minima, but as overlapping, partial densities of the spatial Boltzmann distribution. Their thermodynamical weights are given by their fraction of the total density, see equation (3).

2.2 Mathematical Description of Molecular Conformations

Since conformations are overlapping densities we map them by fuzzy sets, i.e. conformations are defined via membership functions $\chi_1, \ldots, \chi_{n_C} : \Omega \to [0, 1]$,

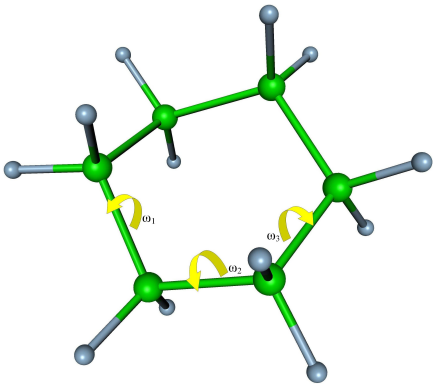

Fig. 2. Cyclohexane with three dihedral angles defining the conformational space Ω. All cyclohexane figures created with amira/amiraMol [14].

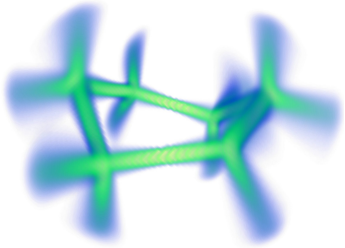

Fig. 3. Volume rendering in 3D cartesian coordinates for a conformation of cyclohexane.

see Deuflhard and Weber [4]. Each point $q \in \Omega$ is assigned to each of the conformations $i = 1, \ldots, n_C$ with a certain degree of membership $\chi_i(q)$. Additionally, a partition of unity constraint has to be satisfied,

$$\forall q \in \Omega \quad \sum_{i=1}^{n_C} \chi_i(q) = 1, \tag{1}$$

which means that the conformations are a kind of soft partitioning of the whole conformational space Ω. Via this concept, conformation analysis can be done in a mathematical rigorous way. For example, if

$$\pi(q) = \frac{\exp(-\beta\, V(q))}{\int_\Omega \exp(-\beta\, V(q))\, dq} \tag{2}$$

denotes the spatial Boltzmann distribution with Boltzmann factor β and potential energy $V : \Omega \to I\!R$, the thermodynamical weights $w_1, \ldots, w_{n_C}$ of the conformations can be calculated as follows:

$$w_i = \int_\Omega \chi_i(q)\, \pi(q)\, dq, \tag{3}$$

where $w_i^{-1}\, \chi_i(q)\, \pi(q)$ is the normalized partial density function corresponding to conformation i.

Since this is a function based definition of conformations, we need a function basis $\phi_1, \ldots, \phi_s : \Omega \to [0, 1]$ for approximation of the conformations. If this function basis has got the same properties like $\chi_1, \ldots, \chi_{n_C} : \Omega \to [0, 1]$, i.e. non-negativity and partition of unity, then conformations are convex combinations of the basis functions, see Weber [18].

In the presence of metastable conformations, the overlap integral matrix $\overline{S} \in I\!R^{s \times s}$

$$\overline{S}(i, j) = \int_\Omega \phi_i(q)\, \phi_j(q)\, \pi(q)\, dq \tag{4}$$

of the basis functions is almost block structured, which can be used to identify the conformations via Robust Perron Cluster Analysis (PCCA+), see Deuflhard and Weber [4] or Weber and Kube [19]. PCCA+ computes the convex combination factors $c_{ij} \in [0, 1], i = 1, \ldots, n_C$, $j = 1, \ldots, s$, for $\chi_i(q) = \sum_{j=1}^{s} c_{ij} \phi_j(q)$, such that the overlap between different conformations is minimized.

Analogous to $w_i^{-1}\, \chi_i(q)\, \pi(q)$, there is a partial density and a thermodynamical weight $v_1, \ldots, v_s$ connected to each of the basis functions. In order to compute the transition behaviour of the molecular conformations, we have to determine the effect of short-time molecular dynamics on these partial densities with respect to the potential function V. The propagation in time span $\tau > 0$ of densities in position space under the influence of a heat bath can be represented by a linear self-adjoint Markov operator P^τ, see Schütte [15]. The corresponding stochastic transition matrix $P \in I\!R^{s \times s}$ is used for an exploration of the transition behaviour in conformation analysis:

$$P(i, j) = \frac{\int_\Omega \phi_i(q)\, P^\tau\, \phi_j(q)\, \pi(q)\, dq}{\int_\Omega \phi_i(q)\, \pi(q)\, dq}. \tag{5}$$

2.3 Sampling Technique for the Function Basis

For a numerical approximation of the desired matrices $\overline{S}$ and P we have to compute the partial densities connected to the basis functions. This can simply be done by an HMC-sampling[1] of modified potential energy functions $V_1, \ldots, V_s : \Omega \to I\!R$, where

$$V_i(q) = V(q) - \frac{1}{\beta}\ln(\phi_i(q)). \tag{6}$$

Via this partial density sampling technique, we assure that each part of the conformational space is sufficiently examined to get a reliable transition matrix P for the exploration of transition paths. The decomposition of the conformational space avoids at the same time trapping-effects. In (6), V denotes the original potential energy function of the molecule and the potential modification $-\beta^{-1}\ln(\phi_i(q))$ is a kind of soft restriction in conformational space.

2.4 Meshfree Partitioning of the Conformational Space

Because of the partition of unity constraint, the function basis $\phi_1, \ldots, \phi_s$ decomposes the whole conformational space. In order to get a good start decomposition of Ω, we apply Shepard's partition of unity method [17] to a set of Gaussian functions. The maxima of the Gaussian functions are called "nodes". In ZIBgridfree the nodes of the function basis are located evenly according to a short presampling at high temperature, for a technical description of ZIBgridfree see Meyer [10], for a mathematical introduction see Weber [18]. We can assume that these functions decompose the part of conformational space sufficiently fine at all locations, where the density of molecular states is relevant above zero at lower temperature. This is the first step of an adaptive discretization of a high-dimensional conformational space.

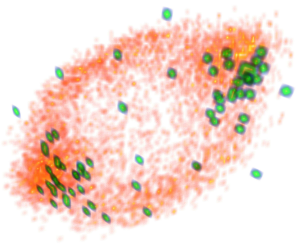

Fig. 4. Density plot of a 1000K presampling of cyclohexane in the 3-dimensional space spanned by the dihedrals ω_1, ω_2 and ω_3 (projected to 2D). The presampling is used as basis for the node selection. In green: Selected nodes for the basis functions after the refinement routine.

[1] For HMC sampling theory see Duane et al. [5] and Fischer [6].

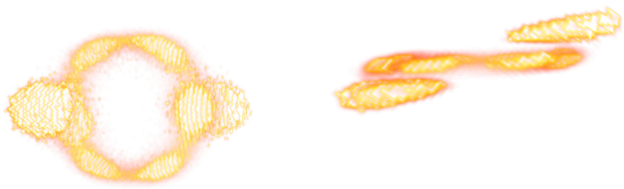

Fig. 5. Two different perspectives of a density plot of cyclohexane in its conformational space spanned by the three dihedrals of the ring for a 300K sampling. Obviously, there are eight accumulation regions in that plot representing eight conformations of cyclohexane.

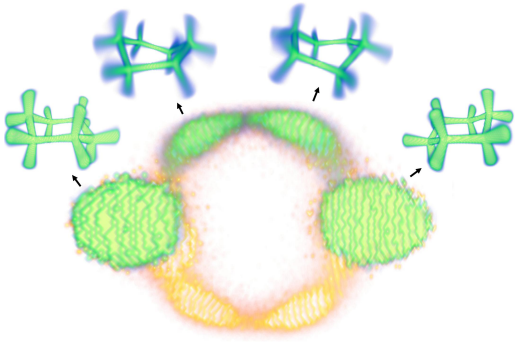

Fig. 6. This figure shows cyclohexane in its conformational space spanned by the three dihedrals of the ring, and above four corresponding conformations in the 3D cartesian space. These are some of the conformations identified via ZIBgridfree. They cover a certain part of the conformational space, which is marked in green.

2.5 Adaptive Refinement of the Partitioning

After generating a start decomposition, we have to adapt the function basis to the given molecular potential. Recursively, we test if a given node leads to a converging subsampling according to (6) or has to be replaced with a new set of nodes. In ZIBgridfree covergence is tested according to an error estimation of the weight computation. Convergence is reached, if the overlap integral values between different samplings become stationary, see Weber [18]. Numerical examples have shown, that this criterion leads to an effective sampling wrt. computational effort. Via refinement of the function basis, not only the error of weight computation is minimized; it can be shown, that especially this kind of refinement indicator at the same time leads to good approximation properties according to the conformations.

3 Numerical Results

As an example for the results of ZIBgridfree [11] we present cyclohexane. Like in Fig. 2, three dihedral angles can give a sufficient description of its conformational space Ω. A presampling at 1000K leads to a Boltzmann distribution in this 3-dimensional space shown as projection in Fig. 4. This pre-sampling is the basis for the node selection algorithm. In Fig. 4, the 75 nodes, which have been selected after applying the refinement algorithm with the above convergence indicator, are shown in green. After selecting the nodes, ZIBgridfree starts the $300K$ HMC-subsamplings for the modified potentials (6) and computes the corresponding thermodynamical weights $v_1, \ldots, v_s$. Plotting the partial densities according to $\phi_1, \ldots, \phi_s$ with regard to their thermodynamical weights ends up in Fig. 5, which represents the spatial Boltzmann distribution of cyclohexane at $300K$ in the conformational space Ω. Robust Perron Cluster Analysis [4] applied to $\overline{S}$, see (4), identifies the corresponding eight accumulation regions, which represent different conformations of cyclohexane, see Fig. 6. The two chair conformations (see Fig. 6 left and right) of cyclohexane together have a total thermodynamical weight of about 99.99%. Finally, with the aid of the transition matrix P, see (5), we examined the transition paths between the two chair conformations of cyclohexane. The result was, that once cyclohexane has reached a non-chair conformation it will leave this conformation in time span $\tau = 40fs$ with a probability of about 30% at $300K$ room temperature. In contrast to that, leaving a chair conformation is only done with a probability of about 0.01%. In other words: Only the chair conformations are metastable, there is a high fluctuation in the transition region. Figure 5 right provides a good overview of the two metastable chair conformations and the ellipsoid transition region.

4 Conclusions

Due to a partition of unity method with an adaptively refined function basis, ZIBgridfree provides a thermodynamically *complete* and *reliable* sampling of the

conformational space of a drug-like molecule. Via Robust Perron Cluster Analysis, it identifies conformations as metastable, partial densities of the Boltzmann distribution in Ω. In contrast to other sampling approaches, the modified potential samplings of ZIBgridfree are additionally focussed into transition regions of the molecule, such that the rare event information of transition paths and probabilities is simulated reliably.

Acknowledgement. The authors want to thank Peter Deuflhard for many fruitful discussions and mathematical input. We also want to acknowledge Frank Cordes and Johannes Schmidt-Ehrenberg for inspiring contributions to our work.

References

1. B. Berne and J. Straub. Novel methods of sampling phase space in simulation of biological systems. *Curr. Opinion Struct. Biol.*, 7:181, 1997.
2. P. Deuflhard. From molecular dynamics to conformational dynamics in drug design. In M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, editors, *Trends in Nonlinear Analysis*, pages 269–287. Springer, 2003.
3. P. Deuflhard and Ch. Schütte. Molecular conformation dynamics and computational drug design. In J.M. Hill and R. Moore, editors, *Applied Mathematics Entering the 21th Century*. ICIAM 2003, Sydney, Australia, 2004.
4. P. Deuflhard and M. Weber. Robust Perron Cluster Analysis in Conformation Dynamics. In M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte, editors, *Lin. Alg. App. – Special Issue on Matrices and Mathematical Biology*, volume 398C, pages 161–184. Elsevier Journals, 2005.
5. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.
6. A. Fischer. Die Hybride Monte–Carlo Methode in der Molekülphysik. Diploma thesis, Department of Mathematics and Computer Science, Free University Berlin, (in German), 1997.
7. D. Frenkel and B. Smit. *Understanding Molecular Simulation – From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Press, A Division of Harcourt, Inc., www.academicpress.com/computationalscience, 2nd edition, 2002.
8. T.A. Halgren. *J. Am. Chem. Soc.*, 114:7827–7843, 1992.
9. T.A. Halgren. Merck molecular force field. *J. Comp. Chem.*, 17(I-V):490–641, 1996.
10. H. Meyer. Die Implementierung und Analyse von HuMFree – einer gitterfreien Methode zur Konformationsanalyse von Wirkstoffmolekülen. Master's thesis, Free University Berlin, 2005. In German.
11. H. Meyer, M. Weber, and A. Riemer. ZIBgridfree. Software package for HMC-simulation and conformation analysis based upon C++ classes of amiraMol [14] using the Merck Molecular Force Field [8][9] implemented by T. Baumeister and parametrized by F. Cordes. Robust Perron Cluster Analysis [4] implemented by M. Weber and J. Schmidt-Ehrenberg. VERX (extrapolation method based on Verlet) implemented by U. Nowak, Status: January 2005. Software owned by the Zuse Institute Berlin.
12. A. Mitsutake, Y. Sugita, and Y. Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*, 60:96–123, 2001.

13. T. A. Knotts N. Rathore and J J. de Pablo. Configurational temperature density of states simulations of proteins. *Biophysical Journal*, 85:3963–3968, 2003.
14. J. Schmidt-Ehrenberg, D. Baum, and H. Ch. Hege. Visualizing dynamic molecular conformations. In *IEEE Visualization 2002*, pages 235–242. IEEE Computer Society Press, 2002.
15. Ch. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Department of Mathematics and Computer Science, Free University Berlin, 1998.
16. Ch. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In P. G. Ciarlet and C. Le Bris, editors, *Handbook of Numerical Analysis, Special Volume Computational Chemistry*, volume X, pages 699–744. North-Holland, 2003.
17. D. Shephard. A two-dimensional function in irregularly spaced data. *In Proc. 23rd AMC National Conference*, pages 517–523, 1968.
18. M. Weber. *Mehless Methdos in Conformation Dynamics*. PhD thesis, Free University Berlin, 2005. In preparation.
19. M. Weber and S. Kube. Robust Perron Cluster Analysis for Various Applications in Computational Life Science. CompLife05 in Konstanz, Germany, September 2005.