STABLE COMPUTATION OF PROBABILITY DENSITIES FOR METASTABLE DYNAMICAL SYSTEMS[*]

MARCUS WEBER[†], SUSANNA KUBE[†], LIONEL WALTER[†], AND PETER DEUFLHARD[†]

Abstract. Whenever the invariant stationary density of metastable dynamical systems decomposes into almost invariant partial densities, its computation as eigenvector of some transition probability matrix is an ill-conditioned problem. In order to avoid this computational difficulty, we suggest applying an aggregation/disaggregation method which addresses only well-conditioned subproblems and thus results in a stable algorithm. In contrast to existing methods, the aggregation step is done via a sampling algorithm which covers only small patches of the sampling space. Finally, the theoretical analysis is illustrated by two biomolecular examples.

Key words. dynamical systems, metastability, molecular conformations, cluster analysis, sampling, aggregation/disaggregation, domain decomposition

AMS subject classifications. 65F15, 82B80, 62H30, 65P99

DOI. 10.1137/060675423

Introduction. Consider general dynamical systems in equilibrium with a unique stationary density. Among them, *metastable* dynamical systems are characterized by the fact that there exist *almost invariant* subsets wherein the system remains for "a long time," once it is in there. In molecular dynamical systems, which represent the main class of interest here, the almost invariant subsets are the *metastable conformations*; for general dynamical systems, we also speak of *metastable clusters*. Throughout this paper, we fix the stationary density to be the Boltzmann distribution without loss of generality. Within each of the conformations, the dynamics is often said to be "rapidly mixing," whereas between them it is "slowly mixing." Any of the available sampling strategies will, when applied to such systems, suffer from the undesirable fact that they are prone to be "trapped" within conformations. Difficulties of this kind also arise in a special multiscale approach called *conformation dynamics*, which has been developed in recent years (for surveys see [8, 9, 30, 31]) and which will represent our algorithmic frame here.

In this paper, we deal with the trapping problem as it shows up in computational sampling techniques for metastable dynamical systems. We design an *efficient* sampling strategy, especially for high-dimensional systems, which both exploits the fast mixing within the conformations and, at the same time, safely covers the rare transitions between the conformations. For our computations, we use the meshless implementation ZIBgridfree [41, 39] of conformation dynamics. In the presence of metastabilities, the arising Markov chains will be *nearly completely decomposable* or *nearly reducible* [34, 5, 23, 6]. The metastable conformations to be identified are analyzed via a stochastic transfer matrix, using our recently developed robust Perron cluster analysis (PCCA+) [10, 39]. The basic reasoning behind the approach to be advocated herein is the following: As we will show below, the computation of the stationary density via the transfer matrix is highly ill-conditioned, whereas

the identification of the metastable conformations is well-conditioned. On the basis of this insight, we suggest a *domain decomposition* of the configuration space into subdomains, ideally the conformations or certain computationally available approximations of them. Our method may be regarded as an infinite-dimensional analogue of *aggregation/disaggregation* methods known from finite-dimensional problems. The disaggregation step comprises the determination of the partial Boltzmann densities, using our algorithm ZIBgridfree. The aggregation step realizes the construction of the overall stationary density, using our newly developed algorithm ConfJump, which works efficiently only when applied to the already identified conformations.

The paper is organized as follows. Existing sampling algorithms and their difficulties in the treatment of metastable dynamical systems are analyzed in section 1. In section 2, we derive the basic idea of aggregation/disaggregation in function space and show how our proposed method can overcome the described difficulties. Finally, we present three numerical examples, one trivial but elucidating and the other two from biochemistry.

1. Difficulties of existing sampling methods. In recent years, the efficient exploration of molecular configuration spaces and the identification of metastable conformations have been the focus of our research; for surveys see [8, 9, 30, 31]. Our algorithms are embedded in the mixed deterministic/stochastic approach called *conformation dynamics*. This approach is based on the analysis of a stochastic transfer operator corresponding to a specific Markov chain and, after some spatial discretization, on a Perron cluster analysis of a stochastic matrix. In the present section, we will briefly revisit conformation dynamics algorithms as they have been used up to now. In the sampling part of these algorithms, some computational dilemmas arise unavoidably, as we will show. However, the same kind of difficulty would equally arise with any other sampling technique when applied to metastable dynamical systems.

1.1. Conformation dynamics revisited. Throughout the paper, we consider a canonical ensemble where the stationary density of configurations q in a continuous space Ω is given by the Boltzmann density $\pi(q)$. Conformation dynamics aims at a description of the continuous dynamical process in terms of a jump process on the discrete set of conformations. Given a molecule's geometry, we identify n_C metastable conformations $C_1, \ldots, C_{n_C} \subset \Omega$ together with their life times and transition patterns. Conformations are given by measurable functions ξ_i defined in Ω (e.g., characteristic functions of the sets $\{C_i\}$), which, together with the Boltzmann density, determine the *conformational weights* or, more general, *cluster weights*

$$(1) \qquad \sigma_i^{\text{exact}} = \int_\Omega \xi_i(q)\pi(q)\,dq, \qquad i = 1, \ldots, n_C.$$

In order to make the computations feasible, the continuous configuration space is decomposed into a number of N discrete "spatial" states, $E = \{1, \ldots, N\}$, and the discrete dynamical process is described as a Markov chain on these states. The stationary density of the corresponding transition probability matrix is denoted by $\mathbf{w} = (w_1, \ldots, w_N)$, in contrast to the stationary density $\pi(q)$ of the continuous process. The entry w_i of the discrete stationary density is denoted as *statistical weight* of state $i \in E$. The vector $\mathbf{w}$ arises as normalized left eigenvector of the transition probability matrix P, i.e.,

$$(2) \qquad\qquad\qquad \mathbf{w}^\top P = \mathbf{w}^\top.$$

As a result of our cluster algorithm PCCA+, membership vectors $\{\chi_i\}_{i=1}^N$ are obtained, in contrast to the continuous functions $\xi_i(q)$. This can be considered as a fuzzy clustering [40]. The entry $\chi_i(j)$ is a number between 0 and 1 and denotes the grade of membership of state j w.r.t. conformation i. Given these values and the stationary density $\mathbf{w}$, the continuous cluster weights in (1) are approximated by discrete cluster weights

$$(3) \qquad \sigma_i = \sum_{j=1}^N \chi_i(j) w_j \approx \sigma_i^{\text{exact}}.$$

The information about the discrete cluster dynamics is condensed in the so-called *coupling matrix* $\mathcal{P}$, the stochastic matrix describing the transitions between the metastable conformations. With diagonal matrices $D = \text{diag}(\{w_i\}_{i=1}^N)$ and $\widetilde{D} = \text{diag}(\{\sigma_i\}_{i=1}^{n_C})$, this matrix can be computed via

$$(4) \qquad \mathcal{P} = \widetilde{D}^{-1} \chi^\top D P \chi.$$

From [8], the characteristic life times of conformations are known to behave (in first order) like

$$(5) \qquad \tau_i \doteq \frac{\tau}{1 - \mathcal{P}(i, i)},$$

where τ is the molecular dynamics simulation time step. Note that these times are extremely sensitive to perturbations of the diagonal entries $\mathcal{P}(i, i) < 1$ when $\mathcal{P}(i, i) \approx 1$.

Summarizing, both the computation of the relevant dynamical information $\mathcal{P}$ and the cluster weights $\{\sigma_i\}$ depend on a well-conditioned determination of $\mathbf{w}$ in (2) and a reasonable clustering χ.

Sampling strategy. Within conformation dynamics, a hybrid Monte Carlo (HMC) sampling technique [11] is used, which is known as a quite efficient method for the sampling of canonical ensembles. However, as with other sampling strategies, its main problem is the so-called *"trapping"* effect, which means that the sampling remains for "too long" a time in the neighborhood of some potential well, an effect also denoted as *"critical slowing down."* As a consequence, the method fails to further visit physically relevant parts of the conformational space—a significant lack of reliability. One possibility to circumvent this undesirable effect is to modify the potential energy surface accordingly and to reweight the generated sampling points. However, such techniques require some deep knowledge about the energy landscape and tend to accumulate too many low-weighted sampling points in transition regions. Therefore, we will advocate below the use of a domain decomposition method instead. Its realization requires our recently developed software package ZIBgridfree to be presented next.

Meshless algorithm ZIBgridfree [41, 39]. The name of this algorithm stands for a *gridfree* (or *meshless*) partitioning of the sampling space such that different regions can be sampled independently. The decomposition of high-dimensional spaces itself is a difficult problem. Regular grids are well known to be inappropriate, because they suffer from the "curse of dimensionality"; for an illustration see Figure 1, left, where the lines represent the conformation domain. In order to overcome this problem, the algorithm realizes a decomposition into "domains" $\{\Omega_i\}_{i=1}^N$ via overlapping sigmoidal functions $\{\phi_i\}_{i=1}^N$; see Figure 1, right:

$$(6) \qquad \Phi_i(q) = \frac{\exp(-\alpha\, d^2(q, \bar{q}_i))}{\sum_{j=1}^N \exp(-\alpha\, d^2(q, \bar{q}_j))}.$$
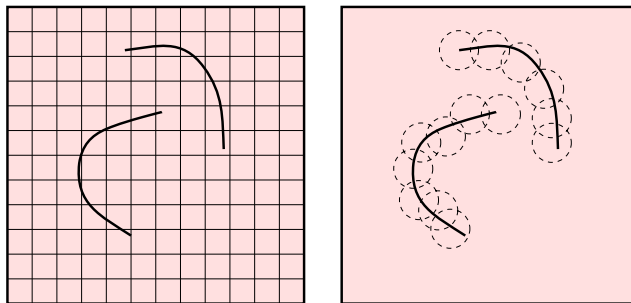
FIG. 1. *Different discretizations of conformations (lines) in configuration space Ω. Left: Regular grids require too many unnecessary boxes, giving rise to the "curse of dimensionality." Right: Meshless methods realize an adaptive, function-based, and overlapping discretization. The circles indicate the "domains" $\{\Omega_i\}$ corresponding to the strongly localized Gaussian radial basis functions $\{\phi_i\}$.*

The parameter $\alpha > 0$ is a shape parameter and $d : \Omega \times \Omega \to \mathbb{R}$ denotes a Euclidean distance function in some lower dimensional space, more precisely in the dihedral space. Details can be found in [39]. By definition, the basis functions form a partition of unity:

$$\sum_{i=1}^{N} \Phi_i(q) = 1 \qquad \forall\, q \in \Omega.$$

This property makes them particularly accessible to umbrella sampling [37] and reweighting strategies. The points $\{\bar{q}_i\}_{i=1}^{N}$ are the centers of the functions which should be localized in dynamically interesting regions of configuration space, i.e., within conformations or in transition regions. To find these regions beforehand, one can perform simulations at a high temperature which do not suffer from trapping problems. In contrast to regular grids, the number of basis functions does not depend on the dimension of the integrand but on its variance [29].

To sum up, the algorithm realizes a *decomposition in function space* rather than in configuration space. However, even though the support of the basis functions is the whole configuration space Ω, they are strongly localized in the conformation domain. Hence, we may as well loosely speak of domain decomposition in Ω. Throughout the paper, we assume that we have already realized some decomposition by means of N basis functions $\phi_1, \ldots, \phi_N : \Omega \to [0, 1]$, where $\phi_i(q)$ determines the grade of membership of configuration $q \in \Omega$ w.r.t. Ω_i. For later purposes, we already mention here that this algorithm may be regarded as realizing a *disaggregation step* in infinite dimension.

In detail, the dynamic is described via a Markov operator T^τ that propagates densities in Ω for some time τ. Here, the complete dynamic in phase space (configurations q and momenta p) comes into play, but the trajectories are projected onto the configuration space Ω [30]. Thus, the conformations to be identified are defined in configuration space but are indeed characterized from a dynamical point of view. The corresponding transition probabilities between domains Ω_i can be expressed in terms of a stochastic transition matrix $P \in \mathbb{R}^{N \times N}$ as follows:

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

It can easily be shown that the integral can be rewritten in terms of local densities

$$\pi_i(q) = \frac{\phi_i(q)\pi(q)}{\int_\Omega \phi_i(q)\pi(q)} \tag{7}$$

as

$$P(i,j) = \int_\Omega T^\tau \phi_j(q)\, \pi_i(q)\, dq.$$

The local densities are Boltzmann densities corresponding to modified potentials $\{V_i(q)\}$, which depend on the basis functions $\{\phi_i(q)\}$, so that

$$\pi_i(q) = \frac{\exp(-\beta V_i(q))}{\int_\Omega \exp(-\beta V_i(\bar q))\, d\bar q}$$

with

$$V_i(q) = V(q) - \frac{1}{\beta}\log(\phi_i(q)). \tag{8}$$

The first step of the sampling routine comprises the generation of sampling points $\{q_k^{(i)}\}_{k=1}^M$ according to the local density $\pi_i(q)$. This is similar to the famous *umbrella sampling* technique introduced by Torrie and Valleau [37]. The second step includes the application of a propagator T^τ to these sampling points and the evaluation of the integral by Monte Carlo quadrature:

$$P(i,j) \approx \frac{1}{M}\sum_{k=1}^M \phi_j(T^\tau q_k^{(i)}). \tag{9}$$

Actually we compute an approximation of P, because Monte Carlo integration introduces truncation errors, which, within ZIBgridfree, are controlled by means of certain stopping criteria. For more details see [39].

*Robust Perron cluster analysis (*PCCA+*).* Transition matrices of completely decoupled Markov chains with n_C independent clusters can be reordered to block diagonal structure and have an n_C-fold Perron eigenvalue $\{\lambda_i\}_{i=1}^{n_C} = 1$. In the presence of metastabilities, we deal with nearly completely decomposable Markov chains. In this case, there exists a *Perron cluster* of eigenvalues near the Perron eigenvalue $\lambda_1 = 1$. The perturbation is measured as the gap to the second largest eigenvalue [10], i.e., as

$$\varepsilon = 1 - \lambda_2, \tag{10}$$

which is understood to be a small number, positive by construction.

In contrast to a reordering of states according to clusters, PCCA+ aims at an assignment of states $i \in \{1, \ldots, N\}$ to clusters $j \in \{1, \ldots, n_C\}$ with certain grades of membership $0 \le \chi_j(i) \le 1$, $\sum_j \chi_j(i) = 1$. Let $X \in \mathbb{R}^{N \times n_C}$ denote the matrix of the eigenvectors corresponding to the Perron eigenvalue cluster. In [10, 39] it has been shown that χ can be expressed as a linear combination of the eigenvectors

$$\chi = X\mathcal{A},$$

where $\mathcal{A} \in \mathbb{R}^{n_C \times n_C}$ is a regular matrix. The algorithm PCCA+ computes this transformation matrix via maximizing the metastability, which is the sum of the diagonal entries of the coupling matrix.

Sampling algorithm ConfJump. In [38], we described the sampling algorithm ConfJump that uses the results of an a priori analysis of the potential energy surface to enhance the sampling by additional intermediate jumps between low-energy regions. This method allows a much faster sampling than the HMC algorithm. There is a class of alternative methods, so-called *smart* methods, which have been designed to overcome the trapping problem; see [3] for a survey. ConfJump can be regarded as an extension of the *smart darting Monte Carlo* [1] and the *jump between wells* [32, 33] methods. However, ConfJump achieves larger acceptance ratios than these methods; i.e., within the Monte Carlo setting the part of proposed states which are accepted as next state in the Markov chain is larger. For algorithmic details see [38]. Since ConfJump combines jumps with the HMC method, it is not only a correct sampling method for the Boltzmann distribution but also provides dynamical information.

1.2. Stability analysis. In section 1.1 we arranged the basic tools for conformation dynamics. On this basis we are now ready to compare the condition of global versus local density computation in the presence of metastable conformations, which is the case of interest here.

Ill-conditioned global density computation. If we want to compute the probabilities for the system to stay within each of the conformations, then we have to solve two problems. The first one is the clustering problem to find the hidden block structure of a transition matrix P and thus to determine the metastable conformations. This problem can be solved by ZIBgridfree and PCCA+ as described in the preceding section. The second problem is the determination of the stationary density $\mathbf{w}$ required to compute the discrete cluster weights $\{\sigma_i\}_{i=1}^{n_C}$ as defined in (3). Unfortunately, a careful analysis reveals that this problem is ill-conditioned. For a deeper understanding, we will briefly recall the line of argument as given in [39].

Assume we want to compute the stationary density via solving the eigenvalue problem (2). In this case, the truncation error of the transition matrix sampling leads to the estimate

$$(11) \qquad \|\mathbf{w} - \mathbf{w}_{tr}\|_\infty \leq \kappa \, \|P - P_{tr}\|_\infty,$$

where $\mathbf{w}_{tr}, P_{tr}$ indicate the result of the algorithm after truncated sampling, and $\mathbf{w}, P$ the true values. As for the choice of condition number, we select $\kappa = \kappa_{ME}$ suggested by Meyer [24], since it can be conveniently estimated via the sorted eigenvalues $\lambda_1 = 1 > \lambda_2 \geq \cdots \geq \lambda_N$ of P as

$$(12) \qquad \frac{1}{N\,|1 - \lambda_2|} \leq \kappa_{ME} \leq \frac{2(N-1)}{\prod_{i=2}^{N}(1 - \lambda_i)}.$$

In view of (10), we thus obtain that, in the presence of at least two conformations,

$$\kappa \sim \mathcal{O}(1/\varepsilon) \gg 1.$$

Summarizing, we are now ready to understand the *computational difficulties* arising in the existing conformation dynamics algorithms:

(i) In the presence of conformations, when there exists a Perron cluster of eigenvalues of P near the Perron root $\lambda_1 = 1$, i.e., when $\varepsilon \ll 1$, the computation of the global stationary density $\mathbf{w}$ is ill-conditioned.

(ii) If the second eigenvalue λ_2 of P is sufficiently bounded away from 1, i.e., if ε is not small, then the computation of the stationary density is well-conditioned, but PCCA+ will not identify any metastable conformations.

The theoretical results so far can also be summarized in a slightly different wording as follows. While the *absolute error* $\|P - P_{tr}\|_\infty$ can be controlled by the stopping criterion within ZIBgridfree, the elementwise *relative error*

$$|P(i,j) - P_{tr}(i,j)|/|P(i,j)|$$

is important for the determination of the stationary density $\mathbf{w}$ [25]. If the transition probability between two states is small, which is the case in the presence of meta-stabilities, then the relative error may be very large although the absolute error is moderate and the sampling converged. In this case, κ in (11) will necessarily be large.

Well-conditioned local density computation. Assume there is a molecular system with n_C pronounced metastable conformations. Then there exists a Perron cluster of eigenvalues $\lambda_1, \ldots, \lambda_{n_C}$ and an eigenvalue λ_{n_C+1} sufficiently bounded away from $\lambda_{n_C} < 1$. Via PCCA+ we can identify the corresponding hidden block structure of P. Assume we have identified disjoint index subsets $I_1, \ldots, I_{n_C}$ for each of these hidden blocks and reordered the states accordingly (which we do not do explicitly!). Then we can construct a real block-diagonal matrix $\hat{P}$ similar to P by applying the Simon–Ando disaggregation technique [34]. That means we take P, set all outer-block elements equal to 0, which is a perturbation of order ϵ, and change the diagonal elements such that $\hat{P}$ is again stochastic. This matrix $\hat{P}$ is reducible and can be split into n_C stochastic submatrices $\{P_i\}$. An easy calculation shows that each of these submatrices defines a reversible Markov chain with a stationary density equal to the corresponding part $\{\mathbf{w}_{I_i}\}_{i=1}^{n_C}$ of $\mathbf{w}$ apart from a scale factor. Details can be found in the appendix. Due to reversibility, DP and $D\hat{P}$ are symmetric. A standard eigenvalue estimate [16] then yields that λ_{n_C+1} is a measure for the deviation of the second largest eigenvalue of each submatrix from 1. Since, by assumption, λ_{n_C+1} is well bounded away from 1, the upper bound in (12) yields a condition number $\hat{\kappa}_i$ for each of the submatrices such that

$$\hat{\kappa}_i \ll \kappa;$$

i.e., local density computation is well-conditioned. This result strongly indicates that some aggregation/disaggregation technique might be useful. This idea will be worked out in section 2.

1.3. An illustrative example. For an illustration of the above stability analysis, consider a dynamical system given by Hamiltonian differential equations:

$$\dot{q} = p, \qquad \dot{p} = -\nabla V(q).$$

Let a characteristic potential energy $V : \mathbb{R}^2 \to \mathbb{R}$ be given in the form

$$V(q) = \min(f(q), g(q), h(q)), \qquad q \in \mathbb{R}^2,$$

$$f(q) = 3(q_1 - 3)(q_1 - 3) - 5(q_1 - 3)(q_2 + 4) + 3(q_2 + 4)(q_2 + 4) + 0.25,$$
$$g(q) = 3(4 - q_1)(4 - q_1) - 5(4 - q_1)(q_2 - 3) + 3(q_2 - 3)(q_2 - 3) + 0.25,$$
$$h(q) = 3q_1q_1 + 3q_2q_2.$$

This potential has three well-separated local minima. The central minimum has the deepest potential energy, but at the same time it has the *smallest* cluster weight such that it is impossible to derive the cluster weights from the energy values of the

local minima. The potential gives rise to a Boltzmann density with well-separated partial densities; see Figure 2(a). Two of the conformations are symmetric, but their basins of attraction have different orientations, which is disadvantageous for the original ConfJump method [38]. The weights of the three conformations can be computed analytically (rounded to four decimal digits) as

$$\text{(13)} \qquad \{\sigma_i^{\text{exact}}\} = \{0.3690, 0.2619, 0.3690\}.$$

Application of ZIBgridfree. Since the single peaks are completely separated and transitions from one cluster to another are rare events, we cannot expect ZIBgridfree to deliver the correct stationary density.

We used $N = 11$ basis functions $\{\phi_k\}_{k=1}^N$. The two symmetric minima were covered by different numbers of basis functions (three and four) so that the discretization does not inherit the symmetry. Moreover, the minima directly coincide with the center of one of the basis functions. The central minimum is resolved by four symmetrically located basis functions. As a result of our algorithm, these four basis functions should have the same statistical weights.

For each modified potential $\{V_i\}_{i=1}^N$ as defined in (8), 2000 points were generated by HMC sampling. These points were propagated by Hamiltonian dynamics applying 60 steps of the velocity-Verlet algorithm [14] with time step $\Delta\tau = 0.013$ fs. We applied PCCA+ to the 11×11 transition probability matrix P, which resulted in a decomposition into three clusters.

Ill-conditioned global density computation. The stationary density was calculated by solving the eigenvector problem of $P^\top$ according to the eigenvalue $\lambda = 1$. The eigenvector was computed by the MATLAB routine `eigs` [21] based on the implicitly restarted Arnoldi method. The Meyer condition number for this calculation appeared as

$$\kappa = 2 \cdot 10^{11}.$$

Well-conditioned local density computation. On the basis of the clustering, we approximated the transition matrix P by a block-diagonal matrix $\hat{P}$ in that we added the entries outside the blocks rowwise to the diagonal, a process described in section 1.2 above and in the appendix. Next, we calculated the stationary densities of the blocks. This time, the corresponding Meyer condition numbers arose as

$$\{\hat{\kappa}_i\} = \{2.15, 4.10, 4.69\}.$$

As one can see, the local condition numbers are extremely small compared with the global condition number $\kappa = 2 \cdot 10^{11}$.

Computation of cluster weights. We decomposed the density according to the proposed clustering into the partial densities. We know that these partial densities are reliable because our samplings were rapidly mixing within the clusters. But we do not know the ratios or proportionality factors exactly. The cluster weights computed according to (3) came out wrong as

$$\{\sigma_i\} = \{0.4303, 0.3453, 0.2244\}.$$

Given the statistical weights $\{w_i\}_{i=1}^{n_C}$ of the basis functions, we computed a histogram of the Boltzmann density. For this purpose, we decomposed Ω into 70×70 boxes. Then we counted the number of sampling points within each of these boxes. In this
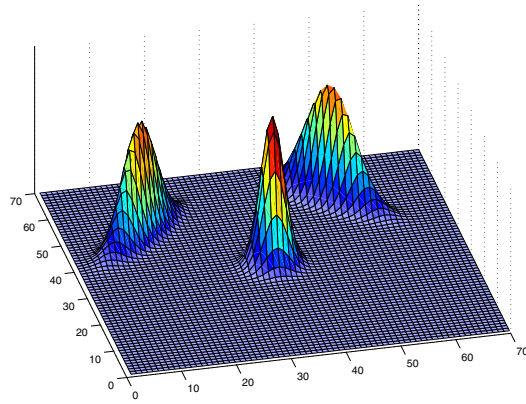
(a) Correct density.
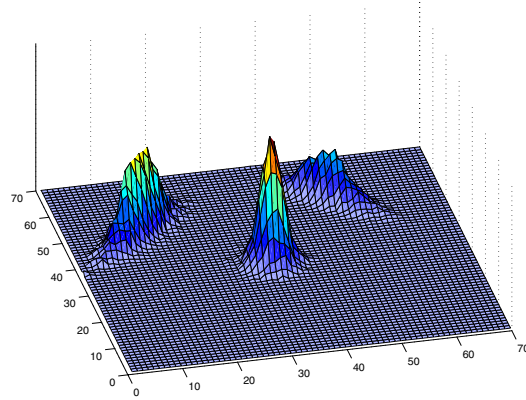
(b) Approximate density computed as left eigenvector of the transition probability matrix.
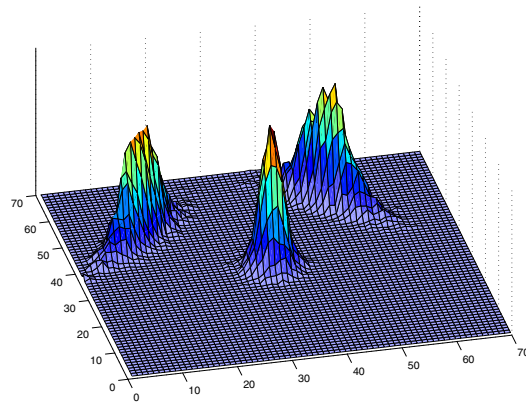
(c) Approximate density computed by ConfJump.

FIG. 2. *Artificial potential: Spatial Boltzmann density computed in three different ways.*

counting process, every point q was scaled by a factor

$$\mu(q) = w_i(q)/M_i,$$

where $w_i(q)$ is the weight of the basis function i from which q was generated, and M_i is the total number of points that was generated to sample the Boltzmann distribution according to basis function i. The histogram according to our example is illustrated in Figure 2(b).

The computed weights are expected to reflect the symmetry of the potential energy surface. Obviously, this is not true in Figure 2(b). Hence, these weights must be incorrect—in agreement with the analysis in section 1.2 above.

2. Sampling-based aggregation/disaggregation. Recall from section 1.2 that the stationary density $\mathbf{w}$ can, in principle, be computed as left eigenvector of the stochastic matrix P to the Perron eigenvalue $\lambda = 1$. In the presence of at least two metastable conformations, this eigenvector is very close to further eigenvalues in the Perron cluster. The computation of the degenerate Perron cluster eigenspace as a whole is well-conditioned, but the computation of any eigenvector in this space is ill-conditioned. This is the situation when the matrix P, which has been computed by truncated sampling, is given as the input and the stationary density $\mathbf{w}$ to be computed is the output. The result of this insight is that the computation of the matrix as such introduces some algorithmic instability.

In order to circumvent this instability, we will step back behind the computation of P; i.e., we will modify the underlying sampling technique. Therefore, in the present section, we will replace the *global* sampling, which gives rise to the instability introduced via the ill-conditioned eigenvalue problem for the matrix P, by a set of independent *local* samplings corresponding to submatrices $\{P_i\}$ associated with well-conditioned eigenvalue problems for the computation of sufficiently accurate partial (or local) densities.

2.1. Basic idea. The key to overcoming the above described computational difficulties is to transfer the well-known matrix-based aggregation/disaggregation approach due to Simon and Ando [34] to some domain decomposition realized in the frame of the sampling technique.

Matrix-based aggregation/disaggregation. The computation of stationary densities of Markov chains has been the topic of intensive research for many years [27, 35, 7, 2]. Iterative aggregation/disaggregation (IAD) methods as introduced in [20, 4, 19, 36] turned out to be efficient algorithms for nearly completely decomposable Markov chains. IAD methods are matrix-based domain decomposition methods. Here, domain decomposition refers to the sampling space, and the states of the Markov chain are defined by the partitioning $\{\Omega_i\}$. First, in the disaggregation step, the states Ω_i are grouped to n_C subsystems (in the following called clusters), which are solved separately. Then, in the aggregation step, the global solution is constructed from the partial solutions. In other words, the unknown overall vector $\mathbf{w}$ is divided into partial vectors $\{\mathbf{w}_i\}_{i=1}^{n_C}$ which may well be of different dimensions:

$$\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_{n_C}), \qquad ||\mathbf{w}||_1 = 1.$$

The solutions of the disaggregation step are the partial densities

$$\widetilde{\mathbf{w}}_i = \frac{\mathbf{w}_i}{||\mathbf{w}_i||_1},$$

where the overall normalization constants

$$\widetilde{\sigma}_i := ||\mathbf{w}_i||_1, \qquad i = 1, \ldots, n_C,$$

are unknown. These normalization constants denote the discrete cluster weights. Note that they are special instances of $\{\sigma_i\}$ defined in (3) in the case that $\{\chi_i\}$ are characteristic functions. From a coupling matrix one can estimate these missing cluster weights which are then used to construct the overall density in the aggregation step:

$$\mathbf{w} = (\widetilde{\sigma}_i \widetilde{\mathbf{w}}_i, \ldots, \widetilde{\sigma}_{n_C} \widetilde{\mathbf{w}}_{n_C}).$$

The method benefits from the fact that the computation of the partial vectors $\mathbf{w}_i$ is well-conditioned. An iterative process, which alternately estimates $\{\widetilde{\sigma}_i\}$ and $\mathbf{w}$, accounts for the interactions between the subsystems. These interactions are given by transition probabilities between states which belong to different clusters. If these values were correct, the IAD algorithm would compute the stationary density efficiently. However, our problem is different in that we cannot rely on these values. That means we cannot apply the method directly to our transition matrix because the aggregation step would fail. Consequently, we had to find a method for the computation of the cluster weights which does not rely on the matrix.

Transfer to sampling techniques. There exist several sampling methods which are based on aggregation/disaggregation techniques. One of them is the hierarchical uncoupling-coupling Monte Carlo method (UCMC) by Fischer et al. [13] and Fischer [12]. To approximate ratios of normalizing constants of different densities, UCMC uses temperature-based bridge sampling. However, that technique suffers from the fact that the densities must be sampled correctly on every hierarchy level, which is very expensive. In our case, here configuration-based bridge sampling [22, 15] would also fail due to missing data in the overlap of sampling regions and due to different shapes of these regions.

This situation motivated the development of a new approach which does not suffer from drawbacks as described above. The partitioning of the molecular configuration space into N different sampling regions as in ZIBgridfree can be considered as the disaggregation step. It also includes the clustering of states into metastable conformations. This partitioning is based on the eigenvector structure of the transition matrix P which is analyzed via PCCA+ [10]. With the result of the clustering we obtain *normalized* partial densities

$$\widetilde{\pi}_i(q) := \frac{\chi_i(q)\pi(q)}{\sigma_i}, \qquad i = 1, \ldots, n_C.$$

This definition is comparable to the definition of local densities in (7), but here partial densities correspond to clusters whereas local densities correspond to the discretization. In order to construct the global density

$$\pi(q) = \sum_{i=1}^{n_C} \sigma_i \widetilde{\pi}_i(q)$$

one has to estimate the normalizing constants $\{\sigma_i\}_{i=1}^{n_C}$ which equal the weights of the conformations. This is part of the sampling-based aggregation step which will be described next.
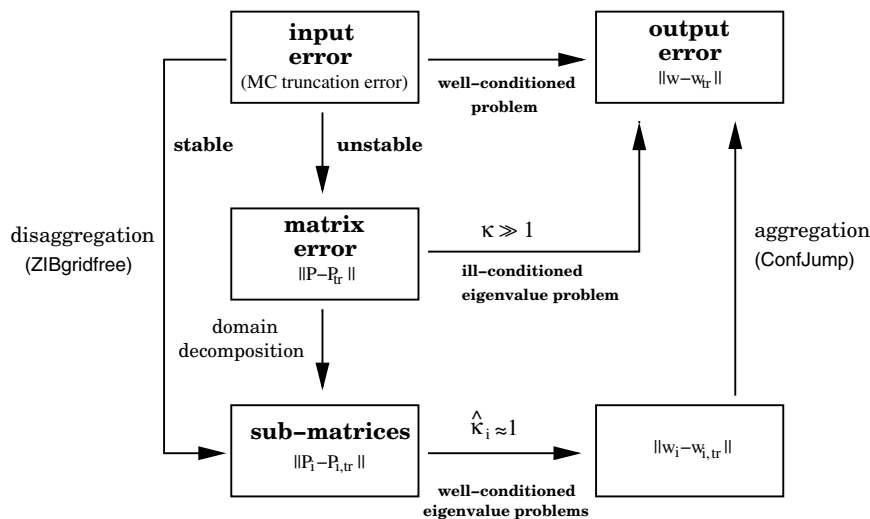
FIG. 3. *Stability diagram corresponding to the basic aggregation/disaggregation algorithm.*

Algorithmic structure. The relation of the different parts of the algorithm together with the stability aspects is schematically presented in Figure 3. The sampling-based aggregation/disaggregation algorithm carefully observes the condition of the global versus the local eigenvalue problems as analyzed in section 1.2 above. Note that in passing we thus have assured that the overall problem as such is well-conditioned. The questions of how to realize this basic algorithm in detail will be discussed in the forthcoming sections.

2.2. Global aggregation. Assume we are given the correct partial densities $\{\mathbf{w}_{I_k}\}_{k=1}^{n_C}$ apart from the scaling factors. There are two possibilities to rescale them correctly. On one hand, one can rescale every partial density individually. If we knew the exact density in a single point q and the actual (wrong) value, we could use this ratio to compute the rescaling factor. That idea is illustrated in Figure 5 and builds the basis for the algorithm presented in the next section.

On the other hand, one can rescale the partial densities w.r.t. each other. For this purpose, one has to compute ratios w_i/w_j for representative indices $i \in I_{k_1}$, $j \in I_{k_2}$, $k_1 \neq k_2$. After renumbering the indices, denote these statistical weights, computed with ZIBgridfree, by $w_1, \ldots, w_{n_C}$ and the corresponding basis functions by $\phi_1, \ldots, \phi_{n_C}$. The idea of the algorithm presented in this section is to sample the Boltzmann distribution directly according to these basis functions $\{\phi_i\}_{i=1}^{n_C}$ by one long trajectory. The number of sampling points per basis function will be proportional to the statistical weights and can thus be used to compute the correct ratios w_i/w_j. The selection of representative basis functions, in the following called patches, avoids a complete resampling within conformations. In our algorithm, we select the basis function with maximum weight within each cluster. From the original N basis functions, we thus obtain n_C patches at the beginning of the sampling algorithm.

From time to time, one has to perform a "jump step" between the corresponding "domains" Ω_i, see Figure 4, because transitions between patches belonging to different conformations are rare events. For this purpose, we use the previously developed algorithm ConfJump [38]. Originally, ConfJump was developed to sample the Boltzmann distribution in the whole configurational space in the presence of metasta-
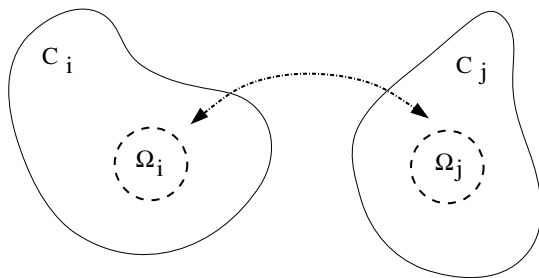
FIG. 4. *Application of* ConfJump *to selected "domains"* $\Omega_i \subset C_i$, $\Omega_j \subset C_j$ *corresponding to basis functions* ϕ_i, ϕ_j.

bilities, but it can also be applied to restricted domains as in our applications. The key idea of ConfJump is to combine the HMC method with intermediate jump steps that preserve the main properties of the sampling scheme: reversibility and convergence towards the equilibrium distribution. One of the main tools in ConfJump is a symmetric jump-proposition matrix A that contains the jump probabilities between the patches. To achieve higher HMC acceptance ratios, these probabilities are related to the potential energies of the patches. Details for the computation of A can be found in [38]. Once a target patch for the jump step has been selected, one has to choose a target configuration q within this patch. For this purpose, a representative configuration is assigned to each patch. Then the target configuration is selected relative to the representative configuration such that reversibility is ensured. To be more precise, a translational displacement is computed w.r.t. some internal coordinates x, here the values of dihedral angles, bond lengths, and bond angles, which are then retransformed into configurational coordinates q; see [38]. In the following, the representative configurations, characterized by their internal coordinates, are denoted as m_i. This transformation from Cartesian coordinates in configuration space to internal coordinates in some lower-dimensional space is essential for achieving reasonable acceptance ratios in the HMC steps [38]. For general dynamical systems, such transformations can possibly be omitted. In our framework, one step of the adapted ConfJump method reads as follows:

1. Let q be the initial configuration (in Cartesian coordinates) w.r.t. basis function ϕ_i.
2. Pick a random number z in $(0, 1)$.
3. If $z > $ jumpRate, perform an HMC step within basis function ϕ_i (i.e., with the modified potential V_i).
4. Else, perform a jump step:
   (a) Select another basis function ϕ_j with probability A_{ij}.
   (b) Compute x, the internal coordinates of q.
   (c) Compute $\tilde{x} = x + (m_j - m_i)$, and transform $\tilde{x}$ to Cartesian coordinates to get the new point $\tilde{q}$.
   (d) Accept the new point with probability $\min\left(1, e^{-\beta(V_j(\tilde{q}) - V_i(q))}\right)$ (Metropolis acceptance criterion with modified potentials).

The numbers of sampled points within each of the patches divided by the number of all ConfJump sampling points give the new weights $\{w_i^{\text{new}}\}_{i=1}^{n_C}$. Thus we get correction factors

$$f_j = \frac{w_j^{\text{new}}}{w_j}, \qquad j = 1, \ldots, n_C.$$

Based on the assumption that the ratios w_i/w_k are correct if basis functions i and k belong to the same cluster, we consider the factors f_j as correction factors for all statistical weights of basis functions belonging to conformation j. This assumption is correct if the local weight computation is well-conditioned. Now we need to multiply the N original statistical weights by the correction factor corresponding to their conformation. For this purpose, the soft membership vectors χ are transformed into a crisp clustering $\overline{\chi}$. The entry $\overline{\chi}(i,j)$ is set to 1 if cluster j is the one to which basis function i belongs with largest probability, and 0 otherwise. Given the crisp membership vectors $\overline{\chi}$, the new statistical weights $\{w_i^{\text{new}}\}_{i=1}^N$ of the basis functions can be computed as

$$(14) \qquad w_i^{\text{new}} = \frac{1}{Z} \sum_{j=1}^{n_C} \overline{\chi}(i,j) w_i f_j, \qquad i = 1, \ldots, N,$$

where Z is the appropriate normalization constant. The corrected statistical weights of the basis functions are used to compute the corrected cluster weights of the conformations via (3).

Note that, theoretically, the correction factors can be determined as accurately as needed if the ConfJump sampling has run long enough. As an advantage, the success of ConfJump does not depend on any local regularity of the potential energy function, which is a key feature of Monte Carlo-based algorithms. Although ConfJump also works without a decomposition method like ZIBgridfree, it is applicable only for larger molecules in combination with ZIBgridfree. This is due to the fact that larger molecules have basins of attraction with very different shapes and energies, which leads to low acceptance ratios in the jump algorithm [38, 28]. Only jumps near the reference configurations m_i have a higher probability of being accepted. Therefore, ConfJump benefits from the restriction to smaller regions via ZIBgridfree.

2.3. Local aggregation via density estimation. Sometimes the acceptance ratio in the ConfJump algorithm is undesirably low, especially if the jump patches have large differences in shape and in the average energy. In such cases, an alternative aggregation method can be used, known as *density estimation*. This heuristic method does not use an iterative sampling strategy. Thus, there is no convergence against the correct reweighting factors. Density estimation constructs an estimate for the underlying continuous Boltzmann density function based on the observed sampling points. Given a sample $\{q_k\}_{k=1}^n$ of points in Ω, the numerically computed density at an arbitrary point q can be estimated by the *kernel density approximation* [26]

$$\hat{\pi}(q) = \frac{1}{n} \sum_{k=1}^n W(q - q_k).$$

As kernel function W, we usually choose an exponential function

$$(15) \qquad W(q) = \exp(-c \, ||q||), \qquad c \in \mathbb{R},$$

which is similar to the basis functions used within ZIBgridfree. In our case, the points $\{q_k\}$ were sampled according to N different basis functions. Let the points $\{q_k^{(l)}\}$ correspond to basis function ϕ_l. With the statistical weights $\{w_l\}_{l=1}^N$ of these basis functions computed by PCCA+, the formula for the density estimation must be modified in the following way:

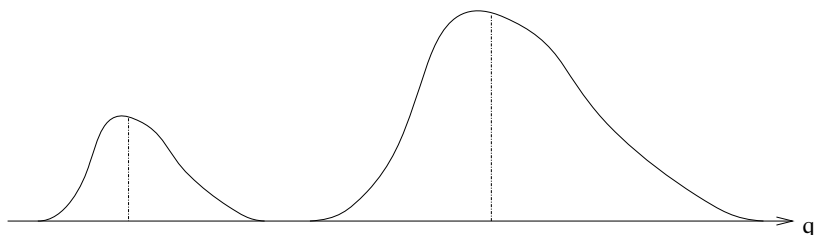$$(16) \qquad \hat{\pi}(q) = \frac{1}{N} \sum_{l=1}^N \frac{w_l}{M_l} \sum_{k=1}^{M_l} W(q - q_k^{(l)}),$$

FIG. 5. *Rescaling of partial densities. The value of the density in a single point determines the scaling factor as long as the density has been sampled correctly locally.*

where M_l denotes the number of sampling points corresponding to basis function ϕ_l.

Furthermore, we know the unnormalized value of the restricted Boltzmann density at an arbitrary point q:

$$\pi(q) \propto \exp(-\beta V(q)).$$

This allows us to compute correction factors according to the idea illustrated in Figure 5. As in the previous section, we first select the basis function with maximum weight within each cluster. Then we choose from the sampling points corresponding to this basis function the point with minimum energy. For these points, in the following denoted by $\{m_j\}_{j=1}^{n_C}$, we compute the estimated densities $\hat{\pi}(m_j)$ according to (16) and scale them such that $\sum_j \hat{\pi}(m_j) = 1$. Moreover, we calculate $\tilde{\pi}(m_j) = \exp(-\beta V(m_j))/\sum_j \exp(-\beta V(m_j))$. The ratios

$$f_j = \tilde{\pi}(m_j)/\hat{\pi}(m_j), \qquad j = 1, \ldots, n_C,$$

are the correction factors for the weights of cluster j. Now we multiply the N original statistical weights of the basis functions by the correction factor corresponding to their conformation, as shown in (14). Then we apply (3) and normalize to obtain the cluster weights.

Since this method does not need a new sampling, it is very fast and cheap. One might have the impression that the approach via local density estimation outperforms the ConfJump algorithm. However, one should be aware that local density estimation works well only in areas where the potential energy surface is sufficiently smooth. This is especially satisfied for small systems and near minima. As already mentioned, ConfJump does not depend on any local regularity assumption.

2.4. Numerical examples. The advantage of the aggregation/disaggregation methods will become clear in the following examples. We briefly revisit the artificial example from section 1.3 and then present two biochemical applications.

Artificial potential. Recall section 1.3, where the stationary density had been computed as eigenvector of the transition probability matrix, which meant that the cluster weights were incorrect.

This is the point where ConfJump comes into play. Within each cluster, we select the basis function with the largest weight and apply the jump method to obtain the scaling factors of the partial densities $\{\tilde{\pi}_i(a)\}$. Then we obtain the corrected cluster weights,

$$\{\sigma_i\} \approx \{0.3777, 0.2585, 0.3637\},$$

to be compared with $\{\sigma_i^{\text{exact}}\}$ in (13). The corresponding histogram is illustrated in Figure 2(c). Now it reflects the symmetry of the potential energy quite satisfactorily, and the approximated cluster weights agree with the analytical weights σ_i^{exact}. Alternatively, we corrected the height of the histogram by a local density approximation. We used the exponential kernel (15) with $c = 6.5$. Other values for c in the range $[4, 10]$ were tested as well, but the results were insensitive w.r.t. the choice of this constant. We obtained the following cluster weights:

$$\{\sigma_i\} \approx \{0.3798, 0.2611, 0.3591\}.$$

The weights agree quite well with the cluster weights obtained by ConfJump and the exact weights in (13).

  Bicyclononane (BCN). This molecule is a hydrocarbon which consists of two connected 6-rings. Each of these rings has two major conformations, the chair (C) and the boat (B) conformation. Thus, we expected four possible combinations (CC), (CB), (BC), and (BB), all separated by high energy barriers. A computation with ZIBgridfree at a temperature of 300K with eight basis functions $\phi_1, \ldots, \phi_8$ led to an 8×8-transition matrix P with the following spectrum:

$$\{\lambda_i\} = \{1.000, 0.999, 0.986, 0.972, 0.916, 0.848, 0.774, 0.692\}.$$

We have $\epsilon = 1 - \lambda_2 = 0.001$ and a significant gap after λ_4. Hence we apply PCCA+ with $n_C = 4$. The uncorrected cluster weights of the four conformations (CC), (CB), (BC), and (BB) came out to be

$$\{\sigma_i\} \approx \{0.943, 0.056, 0.001, 2 \cdot 10^{-14}\}.$$

Obviously, the expected symmetry of the (CB) and the (BC) conformation does not show up correctly. The reason for that is that due to

$$\kappa \sim 1/\epsilon \approx 1000,$$

the global weight computation is not reliable. Note that a correct weight computation does not depend only on κ but also on the expected truncation error $\|P - P_{tr}\|_\infty$. A variance-based estimation provides the truncation error $\|P - P_{tr}\| \approx 0.008$. Unfortunately, if we want to reduce this error by a factor of 10, we have to sample a factor of 100 times more points. As described in section 1.2, we computed the transition probability matrix $\hat{P}$ corresponding to a completely decoupled Markov chain. The maximum condition number $\hat{\kappa}$ for the submatrices of $\hat{P}$ is

$$\hat{\kappa} = \max_i \hat{\kappa}_i = 6.17.$$

In order to scale the well-sampled local densities against each other, we applied the pointwise density estimation for a correction of the weights. Here, in contrast to the artificial example, $V(q)$ is invariant against rotation and translation of $q \in \Omega$. Therefore, we first aligned the trajectories to the points where we wanted to estimate the density via Kabsch's alignment algorithm [17, 18]. Via local density estimation (section 2.3) we computed the corrected cluster weights of the four conformations as

$$\{\sigma_i\} \approx \{0.9893, 0.0053, 0.0054, 3 \cdot 10^{-16}\}.$$

In both examples we used an *irregular* decomposition of the conformational space into basis functions $\phi_1, \ldots, \phi_N$. The discretization did not reflect the symmetry of

the molecule. This means that the reproduction of the molecular symmetry within the weight computation is not an artificial result but a generic property of the described algorithms and can be viewed as a validation.

Inhibitor BSI. The molecule BSI (short name of 2-(Biphenyl-4-sulfonyl)-1,2,3,4-tetrahydro-isoquinoline-3-carboxylate) is a possible inhibitor for human neutrophil collagenase (MMP-8), which degrades collagen.
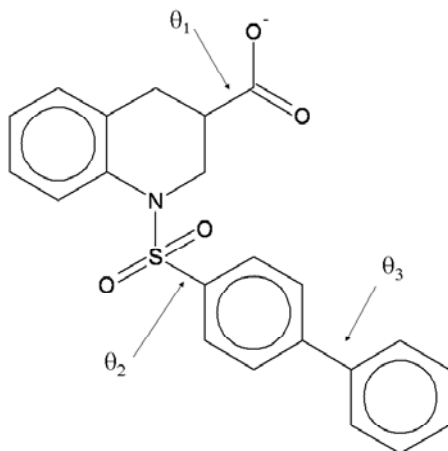

FIG. 6. *Possible MMP-8 inhibitor BSI with three torsion angles θ_1, θ_2, and θ_3 which induce symmetric conformational changes. The values of θ_1 are symmetrically distributed, because the negative charged COO-group has a delocated electron pair instead of a single and a double (C-O)-bond.*

Not only the cluster weights σ_i inherit the symmetry of the molecule: If a change in internal degrees of freedom of the molecule leads to chemically identical structures, all corresponding histograms for these degrees of freedom are symmetric, too. BSI is an example of a molecule for which many sampling steps are necessary in order to compute the correct probability densities. Figure 6 shows this molecule as a structural formula.

BSI has three rotationally symmetric bonds. We discretized the configuration space of BSI into 187 basis functions $\phi_1, \ldots, \phi_{187}$. We performed a sampling with 3000 up to 15000 steps per basis function (total number of steps: $\approx$ 1 million), depending on a convergence indicator. We propagated each sampling point by a molecular dynamics simulation with randomized initial momenta for a time span $\tau = 78$ fs. After computation of the 187×187-transition matrix P we computed the ill-conditioned, incorrect statistical weights $\{w_i\}_{i=1}^{187}$ of the basis functions.

In Figure 7(a) the results for the pure ZIBgridfree approach are plotted with dashed lines. The histograms do not reflect the correct symmetry of the three bonds. But after local density estimation as in section 2.3 we corrected the statistical weights and found that the histograms (dash-dotted line in Figure 7(b)) are now symmetric. However, note that for this special example the good results of the local density estimation are due to the symmetric choice of the center points q for the kernel density approximation. In the example of BSI, each of the 187 basis functions was used for the pointwise density estimation. If the reweighting process with ConfJump is based on each single basis function, this is comparable to performing the pure ConfJump method for the original potential. In the case of BSI, the pure ConfJump
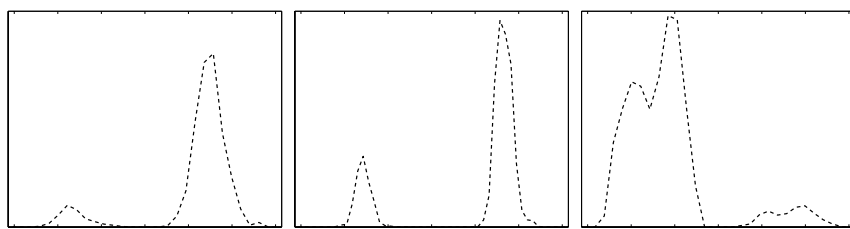
(a) ZIBgridfree without reweighting. The lack of symmetry clearly indicates that the statistical weights of the discretized states are incorrect.
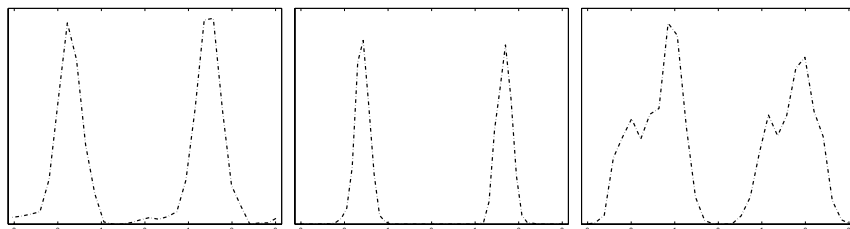
(b) Same sampling as in Figure 7(a) but reweighted by pointwise density estimation. Here, symmetry is nicely reproduced.
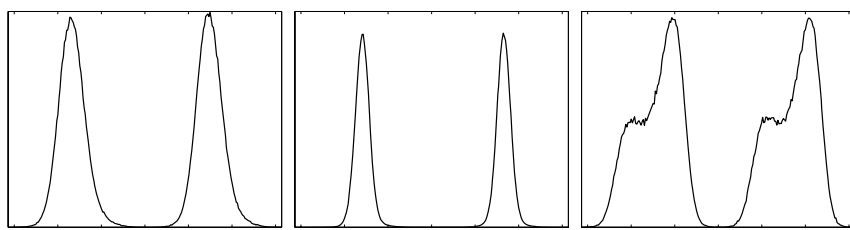
(c) ConfJump reweights the sampling points in such a way that they correctly represent the Boltzmann distribution. Here, histograms came out symmetric.

FIG. 7. *Inhibitor BSI: Equilibrium distribution of values for the dihedral angles θ_1, θ_2, and θ_3 (from left to right) obtained from different samplings of the Boltzmann density (from top to bottom). Symmetry of the histograms in the interval $[-\pi, \pi]$ w.r.t. zero indicates that the Boltzmann density was sampled correctly.*

approach with 1 million sampling points leads to very good dihedral histograms [28]; see Figure 7(c). It was possible for the first time to sample the Boltzmann density in such a way that it correctly reflects the inherent symmetry of the molecule. The jump acceptance ratio was about 32%, which is very high for this approach [38]. We repeated the ConfJump approach with different random number sequences and found out that it is very stable.

Knowing a good approximation of the stationary density, we can compute the coupling matrix $\mathcal{P}$ according to (4). We obtained two conformations with

$$\mathcal{P} = \left(\begin{array}{cc} 0.999987 & 0.000013 \\ 0.000002 & 0.999998 \end{array} \right).$$

They result from a rotation of the unsymmetric N-S-bond. With time span $\tau = 0.78$ fs, the lifetimes computed according to (5) are

$$\tau_1 \doteq 6 \text{ ns}, \qquad \tau_2 \doteq 39 \text{ ns}.$$

These numbers demonstrate how our model can be used to predict the long-term be-
havior of molecules from short-term molecular simulations. Thus we are able to bridge
the time gap between molecular simulations, which use time steps of femtoseconds,
and chemical experiments in the range of nanoseconds.

Conclusion. This paper suggests a stable aggregation/disaggregation algorithm
for the well-conditioned computation of stationary densities of metastable dynamical
systems. The algorithm as it stands now scales with the number of identified con-
formations and can be used as a postprocessor applied to traditional conformation
analysis. Future work will aim at telescoping the traditional tools with this postpro-
cessor.

Appendix. Simon and Ando [34] proposed a method which makes it possible to
compute correctly the partial densities according to subchains of a reversible, nearly
completely decomposable Markov chain. Assume we are given a decomposition of
the index set $\{1, \ldots, N\}$ into n_C subsets $\{I_k\}_{k=1}^{n_c}$ according to the partition into
metastable clusters. Indices i and j are related to each other, $i \sim j$, if there exists an
index k such that $i \in I_k \wedge j \in I_k$; otherwise, $i \not\sim j$. Then a matrix $\hat{P}$ is constructed
as follows:

$$
\hat{P}(i,j) = \begin{cases} 0 & \text{if } i \not\sim j, \\ P(i,j) & \text{if } i \sim j \wedge i \neq j, \\ P(i,i) + \sum_{j, j \not\sim i} P(i,j) & \text{if } i = j. \end{cases}
$$

If $\mathbf{w}$ denotes the stationary density of P ($\mathbf{w}^\top P = \mathbf{w}^\top$, $\sum_{i=1}^N w_i = 1$), then $\mathbf{w}$ is also
a stationary density of $\hat{P}$:

$$
\begin{aligned}
(\mathbf{w}^\top \hat{P})(j) = \sum_{i=1}^N w_i \hat{P}(i,j) &= \sum_{i, i \sim j} w_i P(i,j) + w_j \left(P(j,j) + \sum_{k, k \not\sim j} P(j,k) \right) \\
&= \sum_{i, i \sim j} w_i P(i,j) + w_j P(j,j) + \sum_{k, k \not\sim j} w_k P(k,j) \\
&= \sum_{i=1}^N w_i P(i,j) \\
&= w_j.
\end{aligned}
$$

As can be seen from the construction, the row sums are preserved. Thus, the sub-
matrices $P_{I_k} = \{P(i,j)\}_{i,j \in I_k}$ are stochastic. Moreover, since $\hat{P}(i,j) = P(i,j)$ for
$i \sim j$, we have $w_i \hat{P}(i,j) = w_i P(i,j) = w_j P(j,i) = w_j \hat{P}(j,i)$. In other words, the
submatrices P_{I_k} inherit the reversibility from the original matrix P.

Since $\hat{P}$ is block-diagonal and thus reducible, the stationary density is not unique.
In fact, the subspace spanned by the vectors $\mathbf{v}_{I_k}$ with

$$
\mathbf{v}_{I_k}(j) = \begin{cases} w_j & \text{if } j \in I_k, \\ 0 & \text{else} \end{cases}
$$

is the invariant subspace corresponding to the n_C-fold eigenvalue $\lambda = 1$. The nonzero
components of the vectors $\mathbf{v}_{I_k}$ are the partial density vectors $\mathbf{w}_{I_k}$. They are propor-
tional to the stationary densities of the submatrices P_{I_k}, but the scaling factors are
unknown.

REFERENCES

[1] I. Andricioaei, J. Straub, and A. Voter, *Smart darting Monte Carlo*, J. Chem. Phys., 114 (2001), pp. 6994–7000.

[2] M. Benzi and M. Tůma, *A parallel solver for large scale Markov chains*, Appl. Numer. Math., 41 (2002), pp. 135–153.

[3] B. Berne and J. Straub, *Novel methods of sampling phase space in the simulation of biological systems*, Curr. Opin. Struct. Biol., 7 (1997), pp. 181–189.

[4] W.-L. Cao and W. J. Stewart, *Iterative aggregation/disaggregation techniques for nearly uncoupled Markov chains*, J. Assoc. Comput. Mach., 32 (1985), pp. 702–719.

[5] P. J. Courtois, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, 1977.

[6] T. Dayar, *Permuting Markov Chains to Nearly Completely Decomposable Form*, Technical report BU-CEIS-9808, Department of Computer Science and Engineering, Bilkent University, Ankara, Turkey, 1998, http://citeseer.ist.psu.edu/70412.html.

[7] T. Dayar and W. J. Stewart, *Comparison of partitioning techniques for two-level iterative solvers on large, sparse Markov chains*, SIAM J. Sci. Comput., 21 (2000), pp. 1691–1705.

[8] P. Deuflhard, *From molecular dynamics to conformational dynamics in drug design*, in Trends in Nonlinear Analysis, M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, eds., Springer, Berlin, 2003, pp. 269–287.

[9] P. Deuflhard and C. Schütte, *Molecular conformation dynamics and computational drug design*, in Applied Mathemetics Entering the 21st Century, J. M. Hill and R. Moore, eds., SIAM, Philadelphia, 2004, pp. 91–119.

[10] P. Deuflhard and M. Weber, *Robust Perron cluster analysis in conformation dynamics*, Linear Algebra Appl., 398 (2005), pp. 161–184.

[11] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B, 195 (1987), pp. 216–222.

[12] A. Fischer, *An Uncoupling-Coupling Method for Markov Chain Monte Carlo Simulations with an Application to Biomolecules*, Doctoral thesis, Freie Universität Berlin, Berlin, Germany, 2003, http://www.diss.fu-berlin.de/2003/234/indexe.html.

[13] A. Fischer, Ch. Schütte, P. Deuflhard, and F. Cordes, *Hierarchical uncoupling–coupling of metastable conformations*, in Computational Methods for Macromolecules: Challenges and Applications, T. Schlick and H. H. Gan, eds., Springer, Berlin, Heidelberg, New York, 2002, pp. 235–259.

[14] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Comput. Sci. Ser. 1, Academic Press, San Diego, London, 2002.

[15] A. Gelman and X. Meng, *Simulating normalizing constants: From importance sampling to bridge sampling to path sampling*, Statist. Sci., 13 (1998), pp. 163–185.

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

[17] W. Kabsch, *A solution for the best rotation to relate two sets of vectors*, Acta Cryst. Sect. A, 32 (1976), pp. 922–923.

[18] W. Kabsch, *A discussion of the solution for the best rotation to relate two sets of vectors*, Acta Cryst. Sect. A, 34 (1978), pp. 827–828.

[19] H. D. Kafeety, C. D. Meyer, and W. J. Stewart, *A general framework for iterative aggregation/disaggregation methods*, in Proceedings of the Fourth Copper Mountain Conference on Iterative Methods, 1992.

[20] J. R. Koury, D. F. McAllister, and W. J. Stewart, *Iterative methods for computing stationary distributions of nearly completely decomposable Markov chains*, SIAM J. Algebraic Discrete Methods, 5 (1984), pp. 164–186.

[21] TheMathWorks Inc. Germany, *MATLAB* 6.5.0, 1994–2005.

[22] X.-L. Meng and W. H. Wong, *Simulating ratios of normalizing constants via a simple identity: A theoretical exploration*, Statist. Sinica, 6 (1996), pp. 831–860.

[23] C. D. Meyer, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Rev., 31 (1989), pp. 240–272.

[24] C. D. Meyer, *Sensitivity of the stationary distribution of a Markov chain*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 715–728.

[25] C. A. O'Cinneide, *Entrywise perturbation theory and error analysis for Markov chains*, Numer. Math., 65 (1993), pp. 109–120.

[26] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist., 33 (1962), pp. 1065–1076.

[27] B. Philippe, Y. Saad, and W. J. Stewart, *Numerical methods in Markov chain modeling*,

Oper. Res., 40 (1992), pp. 1156–1179.

[28] A. RIEMER, *Accuracy, Stability, Convergence of Rigorous Thermodynamic Sampling Methods*, Master's thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany, 2006, http://www.zib.de/bib/dipl/RiemerDiplom.pdf.

[29] J. RUST, *Using randomization to break the curse of dimensionality*, Econometrica, 65 (1997), pp. 487–516.

[30] CH. SCHÜTTE, *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*, Habilitation thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany, 1999, http://www.zib.de/bib/habil/index.de.html.

[31] CH. SCHÜTTE AND W. HUISINGA, *Biomolecular conformations can be identified as metastable sets of molecular dynamics*, in Handbook of Numerical Analysis, Vol. X, P. G. Ciarlet and C. Le Bris, eds., North–Holland, Amsterdam, 2003, pp. 699–744.

[32] H. SENDEROWITZ, F. GUARNIERI, AND W. C. STILL, *A smart Monte Carlo technique for free energy simulations of multiconformational molecules. Direct calculations of the conformational populations of organic molecules*, J. Am. Chem. Soc., 117 (1995), pp. 8211–8219.

[33] H. SENDEROWITZ AND W. C. STILL, *Simple but smart Monte Carlo algorithm for free energy simulations of multiconformational molecules*, J. Comput. Chem., 19 (1998), pp. 1736–1745.

[34] H. A. SIMON AND A. ANDO, *Aggregation of variables in dynamical systems*, Econometrica, 29 (1961), pp. 111–138.

[35] W. J. STEWART, *Numerical methods for computing stationary distributions of finite irreducible Markov chains*, in Computational Probability, W. Grassmann, ed., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997, pp. 81–112.

[36] W. J. STEWART AND W. WU, *Numerical experiments with iteration and aggregation for Markov chains*, ORSA J. Comput., 4 (1992), pp. 336–350.

[37] G. M. TORRIE AND J. P. VALLEAU, *Monte Carlo study of a phase-separating liquid mixture by umbrella sampling*, J. Chem. Phys., 66 (1977), pp. 1402–1408.

[38] L. WALTER AND M. WEBER, *ConfJump: A Fast Biomolecular Sampling Method Which Drills Tunnels through High Mountains*, ZIB-Report 06-26, Zuse Institute Berlin, Berlin, Germany, 2006, http://www.zib.de/bib/pub/pw/.

[39] M. WEBER, *Meshless Methods in Conformation Dynamics*, Verlag Dr. Hut, München, Germany, 2006; also available online from http://www.zib.de/bib/diss/index.de.html.

[40] M. WEBER AND T. GALLIAT, *Characterization of Transition States in Conformational Dynamics Using Fuzzy Sets*, ZIB-Report 02-12, Zuse Institute Berlin, Berlin, Germany, 2002, http://www.zib.de/bib/pub/pw/.

[41] M. WEBER AND H. MEYER, *ZIBgridfree—Adaptive Conformation Analysis with Qualified Support of Transition States and Thermodynamic Weights*, ZIB-Report 05-17, Zuse Institute Berlin, Berlin, Germany, 2005, http://www.zib.de/bib/pub/pw/.