6 Numerical Experiments

First, we compare estimations derived from UC with analytically known expectation values for *n-butane* in Sect. 6.1. Next, we present a detailed analysis of a hierarchical decomposition for the simple *n-pentane* molecule in Sect. 6.2. Finally, as a step towards "real-world" problems we apply in Sect. 6.3 UC to *epigallocatechin*, a constituent of green tea which has potential use in drug design.

All simulations are carried out with the algoritmic parts and settings as summarized in Sect. 5.4. To illustrate convergence properties of UC, we focus on approximated 2nd eigenvalues of (restricted) Markov operators(see Sect. 3.3). Throughout this section, hierarchical annealing is done at fixed temperature levels $1, \ldots, L$, which correspond to inverse temperatures $\beta_1 < \ldots < \beta_L := \beta_*$. We denote by $f(\beta)$ the corresponding canonical density, and by $P(\beta)$ the associated Markov operator of a HMC Markov chain that draws samples from $f(\beta)$. Hence, in UC the initial sampling draws samples from $f(\beta_1)$ and the target density f_* is given by $f(\beta_L)$.

6.1 *n*-Butane

With the united-atom representation as force field for n-butane we only have to deal with one torsion angle and a state space $\Omega \subset \mathbb{R}^{12}$ (see Sect. 2.2 and Fig. 4). From the two temperature levels, β_1 and β_*, corresponding to 300 K and 100 K, respectively, we already analyzed the initial sampling $\mathbf{x}_1$ at 300 K in detail in Sect. 5.2.1. The then identified three metastable sets are now used to set up bridge densities between β_1 and β_* and draw respective samples $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ therein. From the samples $\mathbf{x}_1, \ldots, \mathbf{x}_4$ and the weighting factors $\mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4$ we estimate the 4×4 coupling matrix

$$\hat{C} = \begin{pmatrix} 0.9939 & 0.0001 & 0.0058 & 0.0001 \\ 0.2500 & 0.7500 & 0 & 0 \\ 0.2500 & 0 & 0.7500 & 0 \\ 0.2500 & 0 & 0 & 0.7500 \end{pmatrix},$$

with stationary distribution

$$\hat{\pi} = (0.97637, 0.00054, 0.02254, 0.00056)^T,$$

which implies

$$(\hat{\pi}_1^*, \hat{\pi}_2^*, \hat{\pi}_3^*) = (0.0228, 0.9535, 0.0237).$$

Since for n-butane the torsional potential $\mathcal{V}_\omega$ is decoupled from the rest of the force field, the density in the torsion angle ω is solely determined by $\mathcal{V}_\omega$ and can be computed analytically [40]. This property makes n-butane a valuable test system allowing direct comparisons of UC with exact results.

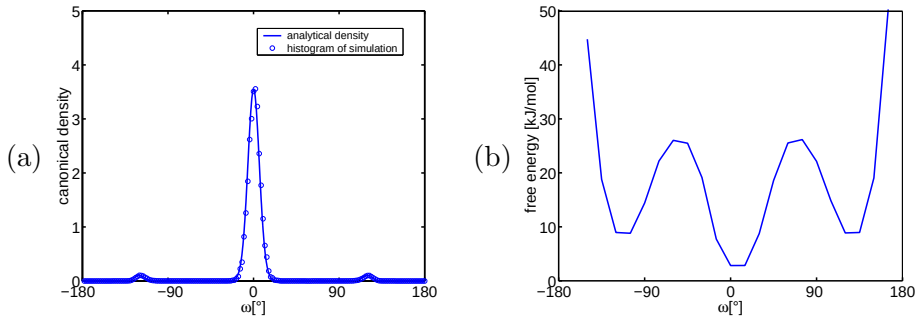

Figure 24: (a) Density f_* corresponding to 100 K with respect to the torsion angle ω. A simple reweighting of the samples according to (84) (histogram indicated by small circles) coincides with the analytically computed density (solid line). (b) Mixture reweighting via (87) results in a detailed free energy profile at $T = 100\,\mathrm{K}$.

Figure 24 (a) demonstrates the good accordance of the reweighted samples with the target density f_*.

Estimates of the statistical weights $p_j = \mathbb{E}_{f_*}(\mathbf{1}_{A_j})$ of the sets A_j result in

$$\hat{p} = (0.0228, 0.09535, 0.0237)\ ,$$

which we have to compare with the exact values

$$p = (0.0240, 0.9520, 0.0240)\ ,$$

which can be derived from the analytical density. We see that the uncoupling-coupling scheme in fact reproduces correct results as predicted by theory when applied to this simple molecular test system. Furthermore, Fig. 24 (b) shows a detailed free energy profile in ω computed from (87) by mixture reweighting.

6.2 *n*-Pentane

In comparison to the simple n-butane united-atom representation, we now use an all-atom Merck molecular force field (MMFF) [62] (see Sect. 2.2) to set up a Hamiltonian for n-pentane. In this representation, n-pentane consists of 5 C-atoms and 12 H-atoms, thus has a state space of $\Omega \subset \mathbb{R}^{51}$; its chemical structure is given by

$$\mathrm{CH_3} - \mathrm{CH_2} - \mathrm{CH_2} - \mathrm{CH_2} - \mathrm{CH_3}.$$

Figure 25 shows three typical configurations corresponding to three different metastable conformations. For decompositions of Ω, we use the self-organizing box map algorithm which is restricted to cluster in the two torsion angles that are defined by C-atoms only (which are known to be the two most essential degrees of freedom).

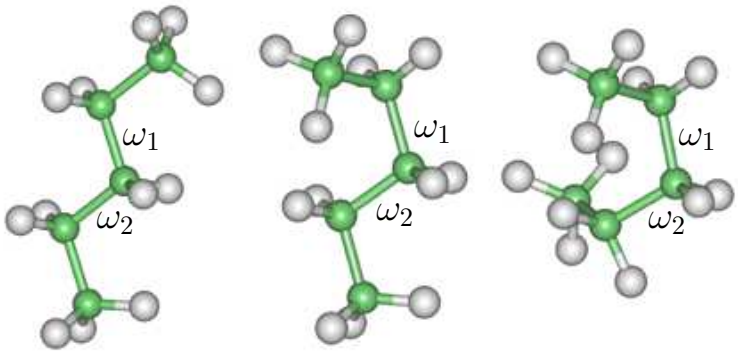

Figure 25: Configurations of *n*-pentane with an all-atom representation as it is used in the Merck molecular force field. Like the single torsion angle for *n*-butane (see Fig. 4), the two torsion angles ω_1 and ω_2 are the essential degrees of freedom for *n*-pentane. From the left to the right: trans-trans, trans-gauche, and gauche-gauche configuration.

To obtain an overview about convergence rates at different temperatures we perform simulation runs at 800 K, 600 K, 400 K, 200 K and 100 K. Each simulation run has a length of 2×10^5 steps where the trajectory length is set to $\tau = 80$ fs with internal stepsize of 1 fs. The corresponding spectra of $P(\beta)$ in Table 1 clearly indicate the movement of some eigenvalues towards the Perron root $\lambda_1 = 1$ for decreasing temperature; these eigenvalues form a well-separated cluster of eigenvalues in the vicinity of 1. At 800 K, 600 K, and 400 K nine eigenvalues belong to that cluster. Yet, for 200 K and 100 K the trapping problem causes the Markov chains to visit only a reduced number of metastable sets, which is indicated by a reduced number of eigenvalues in the vicinity of 1.

An explanation of the existence of nine metastable sets in the high temperature simulations is shown in Figure 26. One torsion angle potential part consists of three local minima located at $-120°$, $0°$, and $120°$, which correspond to the $-$gauche, trans, and $+$gauche-orientation; the product space of two torsion angle potentials consists of nine local minima. Although the minima gets slightly disturbed by the influence of Lennard-Jones interactions, the dominant role of the two torsion angle potentials in $\mathcal{V}$ guarantees that the minima landscape structure is preserved on the whole. Symmetry considerations of $\mathcal{V}$ reveal even more: we obtain symmetric sets by discretizing each torsion angle into three boxes of $120°$. If we denote by s_{ij} the statistical weight of the corresponding $120° \times 120°$-box, $S = (s_{ij})$ possesses the following symmetric structure:

$$S = \begin{pmatrix} d & b & c \\ b & a & b \\ c & b & d \end{pmatrix}. \tag{89}$$

Table 1: Eigenvalues of the discretized Markov operator $P(\beta)$ at different temperatures for n-pentane. The reduced number of eigenvalues in the vicinity of 1 at 200 K ($\lambda_7 = 0.2425$) already indicates the effect of insufficient sampling due to the trapping problem; only 6 instead of 9 metastable sets as in the higher temperature simulations are indicated by its spectrum. This get even worse at 100 K, where only two metastable sets have been visited in the simulation.

	800 K	600 K	400 K	200 K	100 K
λ_1	1.0000	1.0000	1.0000	1.0000	1.0000
λ_2	0.9530	0.9768	0.9935	0.9998	0.9999
λ_3	0.9386	0.9701	0.9925	0.9997	-0.2455
λ_4	0.9380	0.9691	0.9914	0.9997	-0.2206
λ_5	0.9367	0.9670	0.9913	0.9994	0.2161
λ_6	0.9002	0.9485	0.9859	0.9992	-0.1878
λ_7	0.8975	0.9470	0.9855	0.2425	0.1766
λ_8	0.7981	0.8673	0.9254	0.1919	0.1716
λ_9	0.7921	0.8401	0.9009	0.1541	0.1670
λ_{10}	0.3872	0.5188	0.3380	-0.1494	0.1499

Compared with the configurations from Fig. 25, the values a, b, c of the matrix elements correspond to the probability to be in a trans-trans, trans-gauche, and gauche-gauche configuration, respectively (d corresponds to a mirrored gauche-gauche configuration). In general, convergence assessment is a difficult task; for n-pentane the matrix S will serve us as a helpful tool to assess convergence of UC and HMC simulations.

We apply UC to n-pentane in two different scenarios. In the first UC simulation (Sect. 6.2.1), we build up a complete hierarchy starting from the HMC simulation at 400 K. To obtain better convergence properties, we start a second UC simulation at 800 K (Sect. 6.2.2), where we make use of an essential hierarchy as introduced in Sect. 5.2.3. In both UC simulations the target density $f_* := f(\beta_*)$ is set to 100 K.

6.2.1 Complete Hierarchy

As initial sampling for UC we take the HMC simulation at 400 K. Our annealing strategy is a three level hierarchy at temperatures 400 K, 200 K and 100 K corresponding to inverse temperatures β_1, β_2 and $\beta_3 = \beta_*$, respectively. As we can see from the discretized eigenvalues of $P(\beta_1)$ in Table 1 and from Fig. 26 (a), the simulation run visited all metastable sets at 400 K. A geometrical preclustering of the data which serves us as a discretization of the state space $A_1 = \Omega$ is shown in Fig. 26 (b).

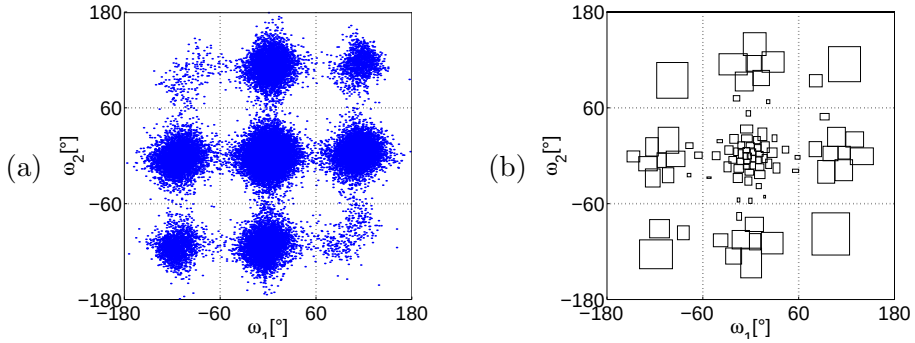

Figure 26: (a) Sample of the HMC simulation at 400 K. Each sample point is the projection of a three-dimensional structure of n-pentane onto the two torsion angles. Dotted lines indicate the $120^\circ \times 120^\circ$-boxes corresponding to (89). (b) Resulting box representation from preclustering the sample with self-organizing box maps.

Uncoupling-Coupling. Figure 27 shows bridge samplings between 400 K and 200 K in the seven identified metastable sets $A_2, \ldots, A_8$. These sets were obtained from the identification algorithm [30], with a detected spectral gap between $\lambda_7 = 0.9855$ and $\lambda_8 = 0.9254$. Note that the eigenvalues $\lambda_8 = 0.9254$ and $\lambda_9 = 0.9009$ together with the extremely large spectral gap beyond λ_9 already indicates the emergence of two further metastable sets. Actually, these two metastable sets got uncoupled on the next level of the hierarchy, which together with the other not further decomposed sets resulted in the sets $A_9, \ldots, A_{17}$. Then, restarted bridge samplings between 200 K and 100 K in the sets $A_9, \ldots, A_{17}$, eventually reached the target density $f(\beta_*)$ (see Fig. 28). Again, the 2nd eigenvalues indicate fast samplings in all metastable sets. The relatively high 2nd eigenvalues around 0.95 for the sets A_{13} and A_{16} are not due to numerical problems but have a natural explanation: Fig. 29 shows a detailed view of the samples of A_{16} and A_{17}. While in the set A_{17} the sample seems to be distributed around one single local minimum of $\mathcal{V}$ the distribution of the sample in the set A_{16} indicates the existence of two local minima (the same observation applies to the sets A_{13} and A_{12}, respectively).

Having completed the uncoupling step, we set up an estimation $\hat{C}$ of the 17×17 coupling matrix C by data analysis, which gave us the invariant distribution $\hat{\pi}$. Exemplarily for the coupling part of the algorithm, Fig. 30 shows reweightings of the bridge density $f_{1,5}$ to the densities f_1 and f_5. Equation (84) together with the restriction $\hat{\pi}^* = (\hat{\pi}_9, \ldots, \hat{\pi}_{17})$ then enables us to compute any desired expectation values with respect to $f(\beta_*)$.

Data Analysis. To check the weighted samples for $f(\beta_*)$, we compute an estimation $\hat{S}$ of the matrix S as given in (89). The nine entries in $\hat{S}$ almost entirely correspond to the nine metastable sets $A_9, \ldots, A_{17}$ of the

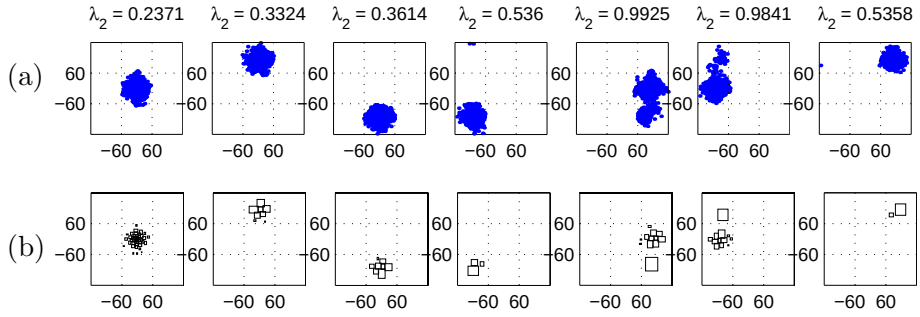

Figure 27: (a) Initial uncoupling of A_1 into seven metastable sets $A_2, \ldots, A_8$ as a projection of the two torsion angles. Shown are the samples from the restricted Markov chains corresponding to the bridge densities between 400 K and 200 K, whereas the number above each set denotes the 2nd eigenvalue of the associated Markov operator. The sets A_6 and A_7 which possess 2nd eigenvalues close to one (their 3rd eigenvalue is 0.3591 and 0.3395, respectively) got further uncoupled for the next hierarchical level into two sets each. (b) Box representations of the seven clusters. In fact, this decomposition is derived from the dynamical cluster algorithm applied to the boxes shown in Fig. 26 (b).

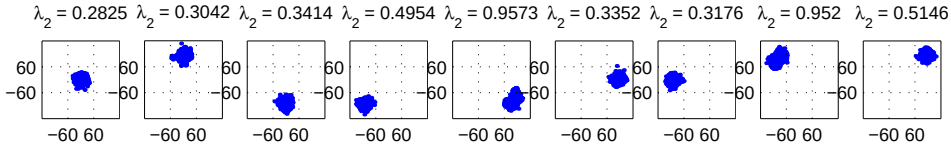

Figure 28: The nine sets $A_9, \ldots, A_{17}$ on the second level that result after uncoupling the sets A_6 and A_7 into two sets each. Shown are the samples from the bridge samplings between 200 K and 100 K together with the 2nd eigenvalues of the corresponding Markov operators.

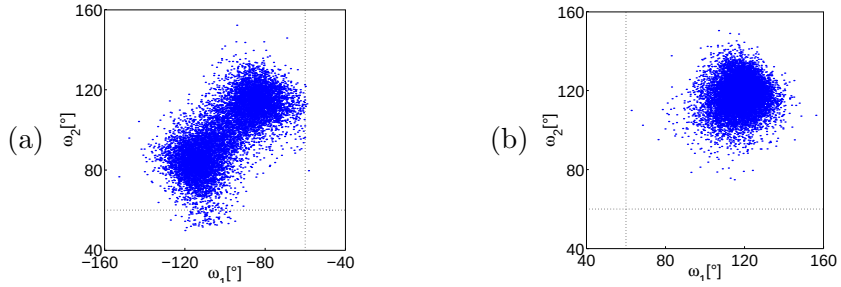

Figure 29: Detailed view of the samples from sets A_{16} and A_{17}. (a) The emergence of a splitting due to an energy barrier in A_{16} explains the high eigenvalue $\lambda_2 = 0.9520$ (whereas $\lambda_3 = 0.5446$). (b) Fast mixing for the set A_{17} with $\lambda_2 = 0.5146$.

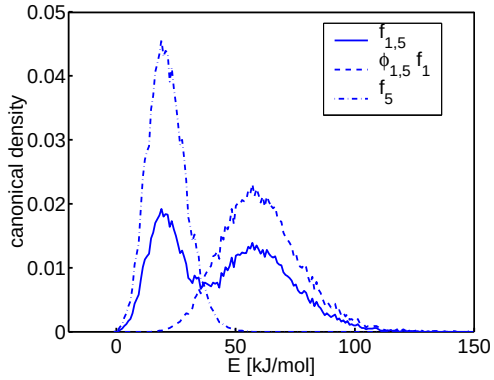

Figure 30: Reweighting of the bridge density $f_{1,5}$ to the densities $\phi_{1,5}f_1$ and f_5. The overlap between $f_{1,5}$ with each of the reweighted densities indicates a statistical reliable computation of the ratio of normalizing constant $Z_{\phi_{5,1}h_5}/Z_{\phi_{1,5}h_1}$.

UC simulation. Using (84) we obtain

$$\hat{S}_{\mathrm{UC}}(\beta_*) = \begin{pmatrix} 3.97 \times 10^{-7} & 0.0123 & 4.09 \times 10^{-4} \\ 0.0118 & 0.9542 & 0.0117 \\ 2.50 \times 10^{-4} & 0.0095 & 1.57 \times 10^{-7} \end{pmatrix}, \tag{90}$$

which is in good agreement with the symmetry of S (see also Table 2). In comparison, the direct HMC simulations at 200 K and 100 K from Table 1 result in

$$\hat{S}_{\mathrm{HMC}}(\beta_2) = \begin{pmatrix} 0 & 0.098 & 0 \\ 0.168 & 0.558 & 0.104 \\ 8.17 \times 10^{-3} & 0.065 & 0 \end{pmatrix}$$

and

$$\hat{S}_{\mathrm{HMC}}(\beta_*) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.601 & 0 \\ 0 & 0.399 & 0 \end{pmatrix},$$

respectively. The matrix $\hat{S}_{\mathrm{HMC}}(\beta_2)$ indicates that three of the four low weighted metastable sets have not been visited at 200 K; even worse, at 100 K the matrix $\hat{S}_{\mathrm{HMC}}(\beta_*)$ indicates how trapping in metastable sets leads to totally wrong results (we already observed this insufficient sampling in the spectra from Table 1).

Furthermore, as outlined in Sect. 5.3.3, reweighting from a mixture distribution allows us to incorporate all samples. We computed a free energy profile in the two torsion angles as shown in Fig. 31. Due to the inclusion of samples from the first two levels of the hierarchy, transition regions between local minima are resolved; we have once more illustrated (this time in terms of the free energy) the existence of nine metastable conformations.

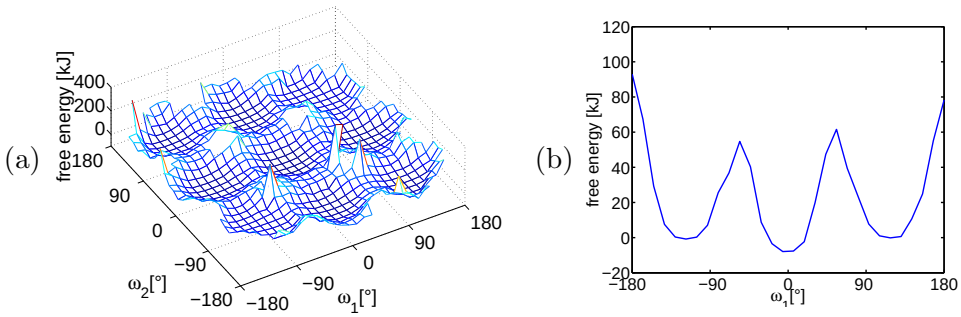

Figure 31: (a) Free energy profile in terms of the two torsion angles. Reweighting all samples from the mixture distribution (87) gives a detailed picture of the free energy landscape even in higher energy regions. The local minima in the free energy landscape coincide with the identified metastable conformations (see Fig. 28). (b) Free energy profile in terms of ω_1 for $\omega_2 = 0$.

Regarding the small statistical weights of the metastable sets $\hat{s}_{11}$, $\hat{s}_{33}$, $\hat{s}_{13}$ and $\hat{s}_{31}$ of $\hat{S}_{\mathrm{UC}}(\beta_*)$, an appropriate strategy to concentrate on *essential metastable sets* would be to discard identified metastable sets during annealing whenever the weight such a set is below a given threshold. This idea leads us to a second UC simulation of n-pentane.

6.2.2 Essential Hierarchy

If we start the initial sampling for n-pentane at a higher temperature, we can make use of a much improved convergence rate due to a lower 2nd eigenvalue of the associated Markov operator. For example, we can see from Table 1 that λ_2 is reduced from 0.9935 at 400 K to 0.9530 at 800 K. Yet, initial sampling at a higher temperature also increases the number of levels in UC and therefore the overall number k of metastable sets $A_1, \ldots, A_k$; metastable sets at high temperature may be identified which on the lowest level have an almost vanishing weight with respect to the target distribution. Furthermore, uncoupling via eigenvectors of eigenvalues that are bounded farer away from 1 are more likely to produce bad decompositions due to numerical problems from simulation or discretization, which could lead to artificial decompositions on subsequent levels of the hierarchy.

To cope with these two problems we already introduced the concept of an essential hierarchy and its realization in UC in Sects. 5.4 and 5.3.1. Its key property is that discarding low weighted metastable sets does not introduce errors in the relative weighting factors for the remaining metastable sets. Applying this scheme to n-pentane, we start with an initial sampling at 800 K and discard identified sets, if their weight with respect to the canonical distribution of their temperature level is less than 2 per cent.

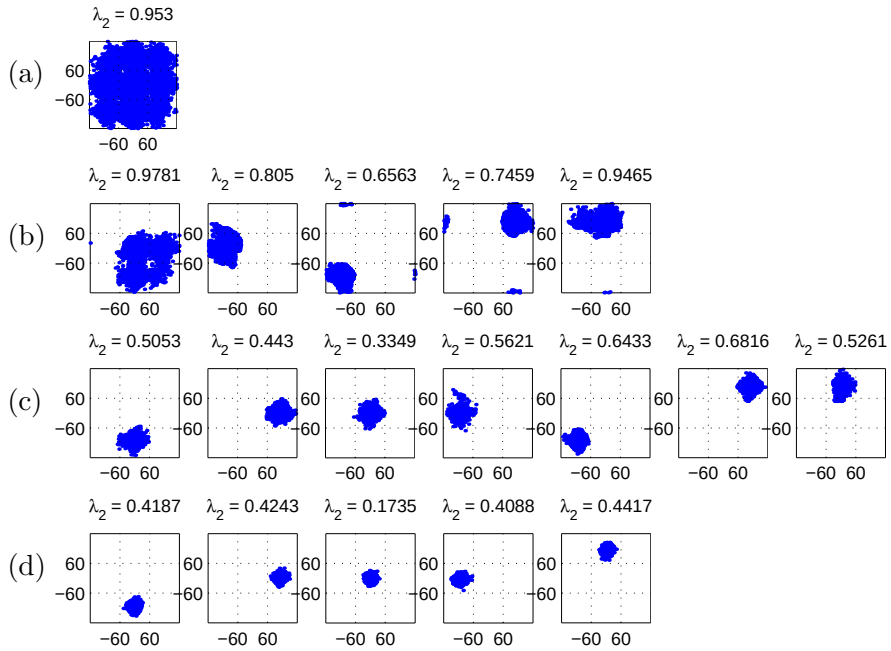

Figure 32: Metastable sets of the essential hierarchy together with the 2nd eigenvalues of the corresponding Markov operators. (a) Initial sampling at 800 K (set A_1). (b) five bridge samplings between 800 K and 400 K (sets $A_2, \ldots, A_6$). (c) seven bridge samplings between 400 K and 200 K (sets $A_7, \ldots, A_{13}$); two identified metastable sets (not assigned to set numbers A_k) got dropped due to their low weight according to our criterion (corresponding to the upper left and lower right corner). (d) five bridge samplings between 200 K and 100 K (sets $A_{14}, \ldots, A_{18}$); two further metastable sets got dropped (corresponding to the remaining lower left and upper right corner).

UC Simulation. Figure 32 shows the entire UC hierarchy of all samples. Our initial sample corresponds to the HMC simulation at 800 K from Table 1. The spectral gap between $\lambda_5 = 0.9367$ and $\lambda_6 = 0.9002$ leads to a decomposition into five metastable sets $A_2, \ldots, A_6$ (Fig. 32 (b)). Restricted sampling in these five sets reveals that A_2 has to be further decomposed into four ($\lambda_4 = 0.891$ and $\lambda_5 = 0.499$), and A_6 into two ($\lambda_2 = 0.947$ and $\lambda_3 = 0.647$) metastable sets. In each of these decompositions we ended up with a set of weight 0.46 per cent and 0.55 per cent for A_2 and A_6, respectively. Bridge sampling between 400 K and 200 K was performed in the remaining seven metastable sets $A_7, \ldots, A_{13}$ (see Fig. 32). The corresponding 2nd eigenvalues indicate that no further uncoupling is needed, yet due to the annealed temperature two of the seven sets are now below our cut-off criterion; the sets A_{11} and A_{12} with a weight of 1.05 and 0.96 per cent, respectively, were discarded. We are left with the five metastable sets $A_{14}, \ldots, A_{18}$, in which we performed a bridge sampling between 200 K and 100 K.

Table 2: Comparison of the two hierarchical decompositions of n-pentane. Shown are the matrices $\hat{S}_{\mathrm{UC}}$ for the complete and the essential hierarchy.

	Complete Hierarchy	Essential Hierarchy
800 K		$\begin{pmatrix} 0.027 & 0.130 & 0.068 \\ 0.142 & 0.256 & 0.147 \\ 0.064 & 0.137 & 0.030 \end{pmatrix}$
400 K	$\begin{pmatrix} 5.04 \times 10^{-3} & 0.114 & 0.029 \\ 0.123 & 0.415 & 0.140 \\ 0.042 & 0.124 & 7.48 \times 10^{-3} \end{pmatrix}$	$\begin{pmatrix} 5.20 \times 10^{-3} & 0.117 & 0.040 \\ 0.128 & 0.418 & 0.132 \\ 0.036 & 0.120 & 3.84 \times 10^{-3} \end{pmatrix}$
200 K	$\begin{pmatrix} 9.97 \times 10^{-5} & 0.065 & 7.28 \times 10^{-3} \\ 0.074 & 0.690 & 0.081 \\ 1.13 \times 10^{-2} & 0.071 & 9.69 \times 10^{-5} \end{pmatrix}$	$\begin{pmatrix} 1.74 \times 10^{-5} & 0.070 & 1.04 \times 10^{-2} \\ 0.077 & 0.692 & 0.075 \\ 9.56 \times 10^{-3} & 0.066 & 0 \end{pmatrix}$
100 K	$\begin{pmatrix} 3.97 \times 10^{-7} & 0.012 & 4.09 \times 10^{-4} \\ 0.012 & 0.954 & 0.012 \\ 2.50 \times 10^{-4} & 0.010 & 1.57 \times 10^{-7} \end{pmatrix}$	$\begin{pmatrix} 0 & 0.013 & 0 \\ 0.014 & 0.949 & 0.013 \\ 0 & 0.011 & 0 \end{pmatrix}$

Complete vs. Essential Hierarchy. Table 2 gives a detailed comparison of the two UC simulations from Sect. 6.2.1 and this section. The matrices $\hat{S}_{\mathrm{UC}}$ on the temperature levels at 400, 200 and 100 K match extremely well. In fact, if we compare the entries of the remaining 5 essential metastable sets in $\hat{S}_{\mathrm{UC}}(\beta_*)$ at 100 K, differences between the two UC simulations are more or less negligible; regarding the convergence of the involved Markov chains, these differences are within the usual range of fluctuations. From Table 2 we can also observe a tendency that low weighted sets reduce their weights when decreasing the temperature. Therefore, errors introduced by this procedure tends to get smeared out during annealing. For example, we dropped a total weight of 1.01 per cent at 400 K and 2.01 per cent at 200 K in the simulation of the essential hierarchy, but the total weight of the respective entries $\hat{s}_{11}$, $\hat{s}_{33}$, $\hat{s}_{13}$ and $\hat{s}_{31}$ of $\hat{S}_{\mathrm{UC}}(\beta_*)$ sums up to only 6.59×10^{-4} per cent in the complete hierarchy; letting aside the trapping problem a standard MCMC sampling would only rarely if at all sample in these parts of the state space.

Regarding the computational cost we can compare the total number of sets in which we simulated Markov chains. In the complete hierarchy there were 17 sets on three hierarchical levels with $1, 7$ and 9 sets, respectively. In the essential hierarchy we have 18 sets on four temperature levels with $1, 5, 7$

and 5 sets, respectively. Although these actual numbers are comparable, the much lower 2nd eigenvalues allow for a faster sampling due to a better rate of convergence.

Choice of Hierarchy. UC sampling with a complete hierarchy results in a detailed and correct picture of the canonical distribution and its metastable sets. The price to pay is that more and more computational cost is spend on "irrelevant" parts of the state space; for larger molecules this could lead to an explosion of metastable sets and make the method less practical.

Instead, UC with an essential hierarchy focuses on metastable sets of high probability and limits their overall number without introducing a noticeable deviation from the target distribution. Uncoupling is already performed at high temperatures, and sampling via an essential hierarchy proved to be quite robust. Furthermore, in case artificial sets in the hierarchy would emerge from uncoupling due to problems from the numerical simulation or the discretization in the low weighted boundary regions, they also would most probably be discarded.

6.3 Epigallocatechin Molecule

From a biochemical point of view, *epigallocatechin gallate* (EGCG) is a polyphonic constituent of green tea that possesses potent anti-tumor, anti-inflammatory and antioxidant properties. Figure 33 shows a typical configuration of EGCG. The structural components include a double ring consisting of a benzene ring and a 3-carbon ring, as well as two separated benzene rings. From chemistry it is known that benzene rings are planar, whereas the 3-carbon ring is non-planar. This behavior is modelled in the force field and clearly shows up in our simulations.

The molecule, which consists of 51 atoms, is a good test system for modest-sized biomolecules that have the potential of being developed into a drug. A fairly detailed description of its three-dimensional structure is given by a set of 35 torsion angles, which are sufficient to describe conformational changes. From these, the three torsion angles ϕ, ω, and ψ as given in Fig. 33 will become of special importance in our analysis.

Simulations are carried out by means of the MMFF force field (see Sect. 2.2). For each HMC step a number n is chosen randomly between 40 and 100 Leapfrog steps of 1.4 fs, which sums up to a total trajectory length between 56 and 140 fs. These settings lead to an acceptance ratio around 75 per cent for the initial HMC as well as for all subsequent PSHMC simulations.

Before we start with the UC simulation, we perform two standard HMC simulations (each of length 200000) at 1200 K and 900 K. Table 3 illustrates the temperature dependence of eigenvalues for the associated Markov operators. The "discretized" eigenvalues were obtained by some combined geo-

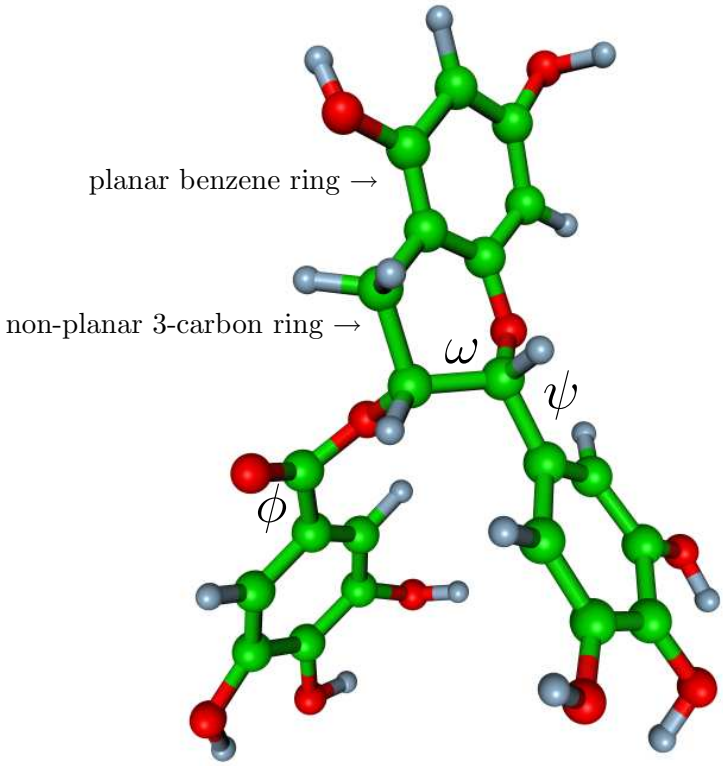

Figure 33: Configuration of epigallocatechin gallate (EGCG) chosen from a sample at 1200 K. The torsion angle ω is suitable to describe separation into the two most metastable sets. The angles ϕ and ω are good convergence indicators, since marginal distributions of the canonical distribution in these angles are periodic with a period of 180 degree.

Table 3: Eigenvalues of the discretized operators $P(\beta)$ at 1200 K and 900 K from standard HMC simulations. Severe trapping problems at lower temperatures prevent us from estimating corresponding eigenvalues at lower temperatures.

	1200 K	900 K
λ_1	1.0000	1.0000
λ_2	0.9840	0.9958
λ_3	0.9614	0.9755
λ_4	0.9275	0.9481
λ_5	0.8766	0.9127
λ_6	0.8631	0.9053
λ_7	0.8552	0.8973
λ_8	0.8478	0.8831
λ_9	0.8403	0.8751
λ_{10}	0.7954	0.8512

metric and dynamic clustering of the state space as described in Sect. 3.3.4 (a detailed description can be found in [21]). In this approach dynamical clustering is applied hierarchically to the 35 torsion angles; the coordinates that proved to be most metastable are used to construct a Galerkin discretization of the whole state space. Similar to n-pentane (see Table 1) we observe that the eigenvalues move toward 1 when lowering the temperature. At 900 K a 2nd eigenvalue of $\lambda_2 = 0.9958$ already indicates the presence of strong metastabilities, which is the reason why below 900 K trapping problems make it impossible to obtain a reliable sampling within reasonable time.

UC Simulation. We carried out UC simulations over four temperature levels in steps of 300 K, starting at 1200 K and ending up at 300 K. For the initial HMC simulation, we use the sample at 1200 K with the eigenvalues given in Table 3. Application of dynamical clustering leads to three metastable sets, which are shown in Fig. 35. In terms of torsion angles the separation into these sets is given by the orientation of ω belonging to the non-planar 3-carbon ring (which correlates to other torsion angle rotations of this ring). A change in ω induces a flipping behavior for the two separated benzene rings (compare Fig. 35 (a) with Fig. 35 (b)). One of these subsets is further divided by ϕ, whereas in the other subset no metastability is induced from ϕ.

Proceeding further, restricted bridge samplings between 1200 K and 900 K were started in the three metastable sets shown in Fig. 35. At 900 K no further metastable sets were detected, so UC proceeded with three bridge simulations between 900 K and 600 K. Further uncoupling lead to seven bridge

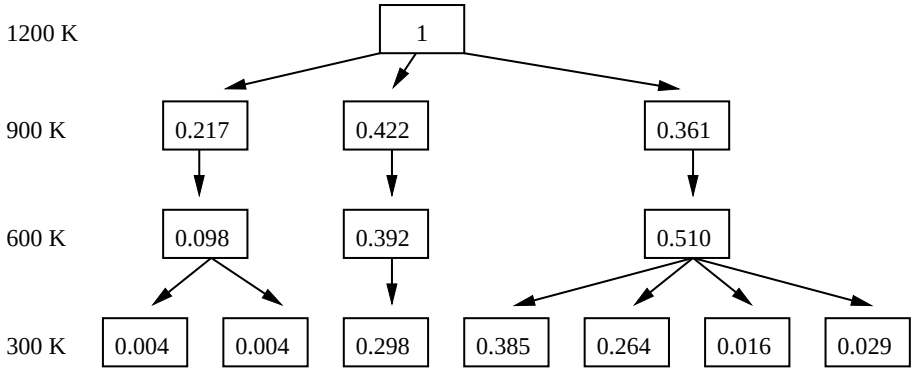

Figure 34: Hierarchical tree of UC simulation. Percentages indicate the statistical weight of the sets wrt. the temperature level. The root corresponds to the initial HMC sampling (cf. with Fig. 22), the three sets at 900 K with weights 0.217, 0.422, and 0.361 correspond to the metastable sets shown in Fig. 35 (a), (b), and (c), respectively.

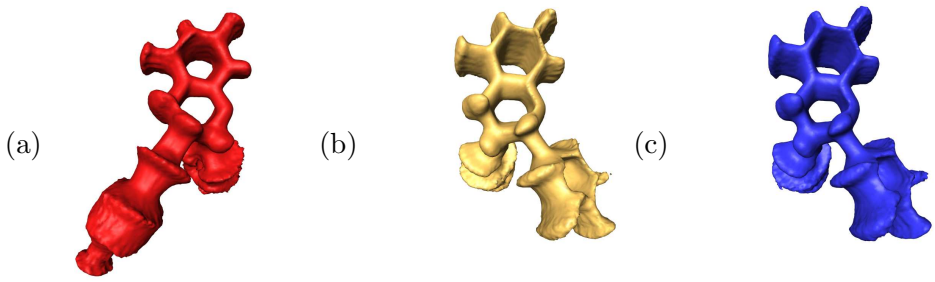

Figure 35: Isosurfaces of three metastable sets of epigallocatechin gallate (EGCG) identified at 1200 K (for details of visualization see [111]). At this temperature, all sets have approximately the same weight. Metastable set (a) with a weight of 30.47 per cent differs from (b) and (c) in the orientation of the two benzene rings relative to the double ring (caused from a different orientation of the torsion angle ω). Metastable sets (b) and (c) with weights 37.07 and 32.46 per cent, respectively, look visually almost identical. Yet, they differ from an 180 degree rotation around ψ, and some minor differences not visible in the isosurface plot.

simulations between 600 K and 300 K.

In Fig. 34 the UC hierarchy together with the statistical weights of its metastable sets are given. Although we used the concept of an essential hierarchy, all identified sets had enough weight to be above the threshold of 1 per cent, so none of them was discarded. If we would proceed to a lower temperature, however, two out of the seven sets at 300 K would be discarded.

All bridge simulations show a good sampling of their lower and upper temperature regions. The behavior is comparable to n-pentane (cf. Fig. 30), but due to the increased dimension of the state space sampling now takes place in a much broader energy range.

The 2nd eigenvalues from the restricted Markov operators, which we

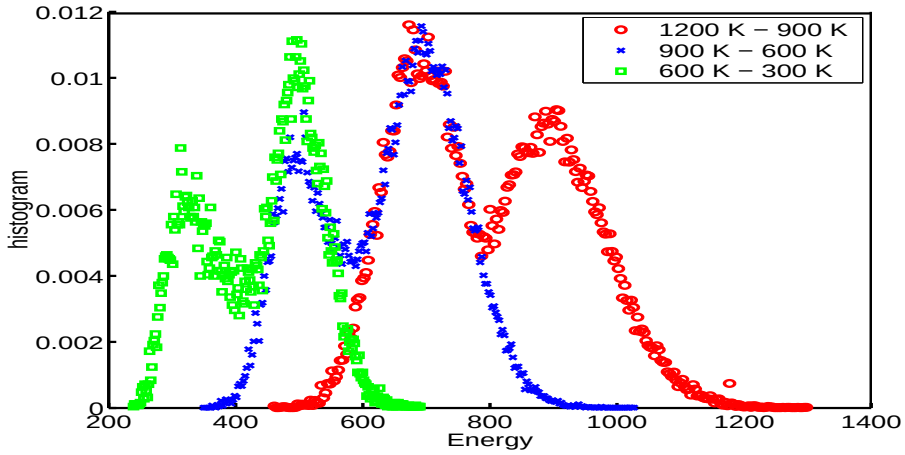

Figure 36: Three energy distributions from bridge samplings span up the whole energy range from 1200 K to 300 K (the three distributions belong to the middle branch of Fig. 34). The histograms nicely illustrates the overlap in energy of successive bridge distributions during annealing. Reweighting to fixed temperatures results in similar behavior as in Fig. 30, and is therefore omitted here.

used so far as indicators for the quality of the UC simulation, are not available any more. The reason is that although dynamical clustering is based on an eigenvalue analysis on each torsion angle, the actual eigenvalues of the Markov operator are (in contrast to the low-dimensional situations we faced so far) not explicitely computed as part of the UC simulation. Alternatively, to judge the quality of the outcome, we exploit a symmetry of $\mathcal{V}$: The torsion angles ϕ and ψ are periodic with a period of 180 degrees, which correspond to a 180 degree rotation of the two single benzene rings. Therefore, the distributions on ϕ or ψ need to be symmetric, hence providing useful spatial observables. Figure 37 shows histograms for ϕ, ω, and ψ on all four temperature levels. The histograms are computed from reweighted samples from the bridge distributions weighted with their coupling factors as given in Eq. (84). The angles ϕ and ψ reproduce more or less the analytical given periodicity, with stronger deviations at lower temperatures. The orientation of ω determines the initial separation between the set (a) and the sets (b) and (c) of Fig. 35. Interestingly, the weight of the metastable set from Fig. 35 (a), which corresponds to the interval $[-180, -60] \cap [120, 180]$ in ω, reduces from initially 30.5 per cent at 1200 K to 9.82 per cent at 600 K and almost vanishes at 300 K.

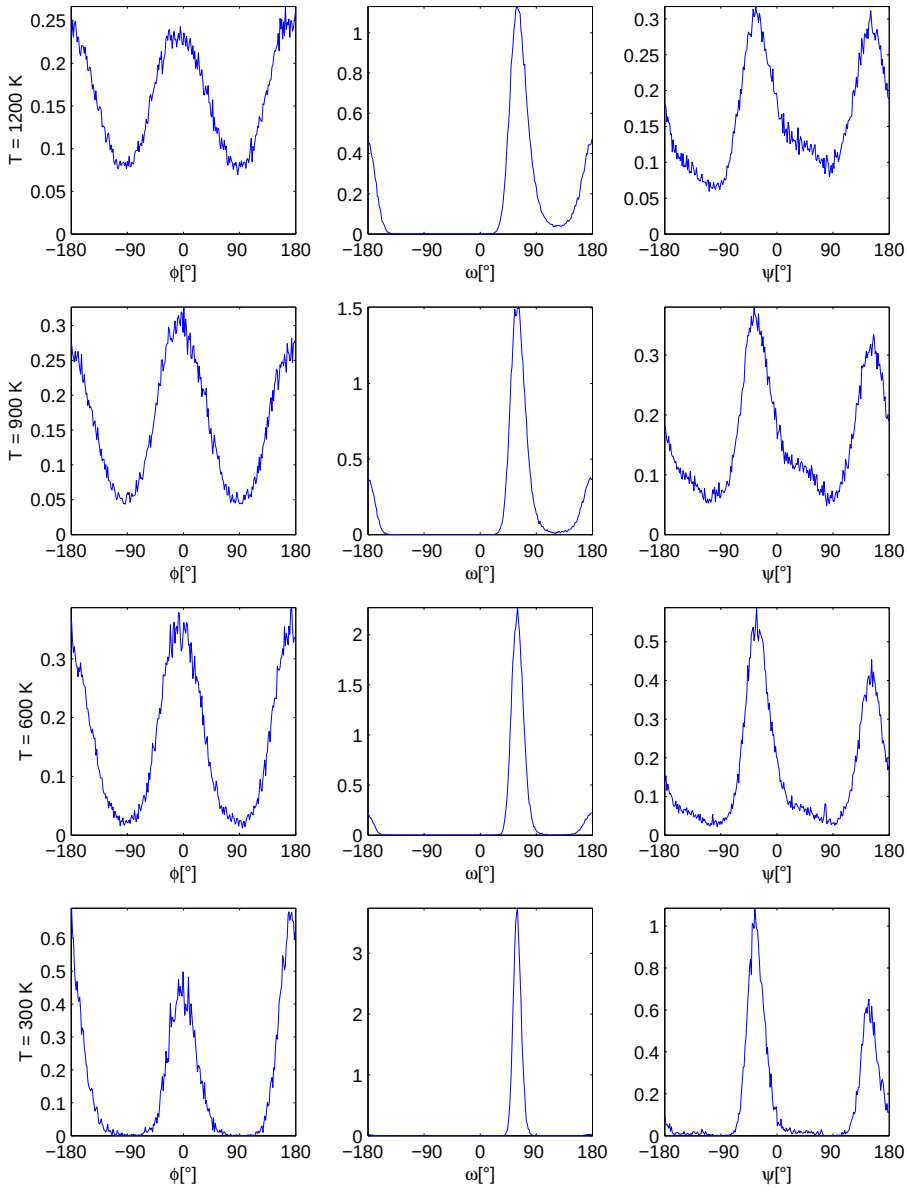

Figure 37: Histograms of ϕ, ω, and ψ wrt. the canonical distribution on each temperature level. The histograms are obtained according to Eq. (84). The spatial observables ϕ and ψ are known to be periodic with a shift of 180 degree. The torsion angle is not periodic, but describes the seperation of the highest metastability in EGCG (cf. Fig. 35).