MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **63** (2010) 319-332

ISSN 0340 - 6253

A simple method for the estimation of entropy differences

Marcus Weber, Karsten Andrae
Zuse-Institute Berlin, Takustraße 7, D-14195 Berlin, Germany

(Received May 7, 2009)

Abstract

We present a simple method to calculate conformational entropy differences between different polymer structures from a thermostated molecular dynamics simulation. The mathematical concept is based on a special Monte Carlo quadrature approach combined with density estimation. First, this approach will be derived for a general integrant f. Then, we will apply the entropy estimator in order to investigate different protein resistant surfaces based on derivates of polyethylenglycol peg. For peg, the highest entropy value is found, which is an indicator for an optimal protein resistant function of peg. Such a preference is also seen in thermostated molecular dynamics simulations of larger chain lengths of peg.

1. Introduction

In this article, we will investigate entropy differences of molecular systems from a computational statistical thermodynamics point of view. The basis for our considerations is the *canonical ensemble*, i.e., we will analyze systems in which the number of particles, the volume, and the *temperature* T are kept constant. In a canonical ensemble, different *molecular states* $q \in \Omega$ are observed with different probability depending on the *potential energy* $V(q), V : \Omega \to \mathbb{R}$. In this case, Ω is a high-dimensional state space. The probability to observe a state q is given by a *probability density function*, $p : \Omega \to \mathbb{R}$, with

$$p(q) = \frac{\exp(-\beta V(q))}{\int_\Omega \exp(-\beta V(\bar{q}))d\bar{q}}. \tag{1}$$

In this definition, $\beta = (kT)^{-1}$ and k is the *Boltzmann constant*. The *entropy* S of a molecular system is given by the differential entropy [1]:

$$S := -k\langle \ln \circ p \rangle = -k \int_\Omega \ln(p(q))\, p(q)\, dq. \tag{2}$$

An ordered system has a low entropy value, whereas, a disordered system has a high entropy value. Monte Carlo quadrature methods are widely used for the numerical approximation of integrals of the form

$$\langle h \rangle = \int_\Omega h(q)\, p(q)\, dq. \tag{3}$$

These methods are useful because Ω is a high-dimensional space in general and deterministic quadrature formulas suffer from the curse of dimensionality. Unfortunately, standard Monte Carlo methods are not suitable for the computation of integrals like (3) in statistical thermodynamics because most of the states $q \in \Omega$ have a very low probability density value $p(q)$. Additionally, the states with a significant p-value which contribute to the integral (3) are not restricted to a certain subdomain of Ω, they are "spread around" in Ω. Via very sophisticated molecular simulation methods or special Markov chain Monte Carlo algorithms it is possible to compute a set of states $q_1, \ldots, q_N \in \Omega$ such that the states q_i are distributed according to the probability density function p. In this situation, a quadrature problem of the form (3) might be estimated by

$$\langle h \rangle \approx \frac{1}{N} \sum_{i=1}^{N} h(q_i). \tag{4}$$

The corresponding approach to the integral (2) would be given by

$$S \approx -\frac{k}{N} \sum_{i=1}^{N} \ln\left(p(q_i)\right). \tag{5}$$

In (5), we have to compute the function $\ln(p(q_i))$ for every sampling point q_i. This is not possible if p is only known except for a normalization constant (i.e. the denominator in (1)). In the molecular context, this constant is not robustly available by any Monte Carlo quadrature method known so far. Computing this constant is equivalent to computing free energy differences [3]. Hence, we will present a different approach to the computation of entropy differences based on a Boltzmann sampling of the state space.

2. Concept

The mathematical concept for the estimation of entropy differences is based on a special Monte Carlo quadrature approach of an *integrand* $f : \Omega \to \mathbb{R}$. This approach can be used if a sampling of states $x_i \in \Omega, i = 1, \ldots, N$ according to f is available. When speaking of "distribution of states according to a function f", this is only possible if f is nonnegative. For general functions f, the following transformation can be used:

$$I = \int_\Omega f(x)\,dx = \Big(\int_\Omega |f(\bar{x})|\,d\bar{x} \Big) \cdot \int_\Omega \operatorname{sign}(f(x)) \frac{|f(x)|}{\int_\Omega |f(\bar{x})|\,d\bar{x}}\,dx. \tag{6}$$

Equation (6) indicates that the new concept depends on some mathematical assumptions. In the following, f is an absolutely integrable function and Ω is a compact space with volume $\operatorname{vol}(\Omega)$. The concept for the computation of integrals of nonnegative functions $|f|$ is based on the following idea: Although, the normalization constant (this is the desired result) for a density function $p \propto |f|$ is unknown, we can sample from this distribution via Markov chain Monte Carlo methods. The smaller the normalization constant the denser the sampling will be – compared to the equal distribution of states. The quantification of the "denseness" of the sampling can be estimated by a ratio of two numbers. For a given state $\widetilde{x} \in \Omega$, $U_1(\widetilde{x})$ is the number of states in an environment of $\widetilde{x}$ determined for a given sampling of the equal distribution, and $U_f(\widetilde{x})$ is the analogue quantity for a p-distribution. This concept for the computation of $\int_\Omega |f(\bar{x})|\,d\bar{x}$ in (6) can be expressed mathematically:

$$\begin{aligned} \frac{\operatorname{vol}(\Omega)\,|f(\widetilde{x})|}{\int_\Omega |f(\bar{x})|\,d\bar{x}} &= \frac{\big(\int_\Omega d\bar{x}\big)\,|f(\widetilde{x})|}{\int_\Omega |f(\bar{x})|\,d\bar{x}} \\ &= \frac{\int_\Omega \delta(x,\widetilde{x}) \frac{|f(x)|}{\int_\Omega |f(\bar{x})|\,d\bar{x}}\,dx}{\int_\Omega \delta(x,\widetilde{x}) \frac{1}{\int_\Omega d\bar{x}}\,dx} \\ &\approx \frac{U_f(\widetilde{x})}{U_1(\widetilde{x})}. \end{aligned} \tag{7}$$

In (7), the function $\delta(x,\widetilde{x})$ is the Dirac delta function. Thus, the integrals in (7) measure the density of states (equally distributed vs. $|f|$-distributed states) at a certain *representative* $\widetilde{x} \in \Omega$. The ratio of these local densities is estimated by a

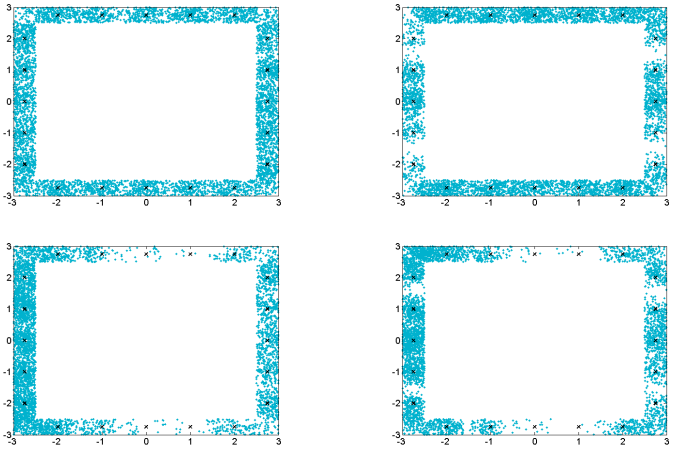

Figure 1: Sampling points (light dots) according to $|f|$-distribution for four different integrands $f_1, \ldots, f_4$ (from top left to bottom right). Representatives for the mean value computation (dark crosses).

corresponding ratio $U_f(\widetilde{x})/U_1(\widetilde{x})$ for a discretized histogram. The more sampling points are generated the more accurate is the histogram. By inserting (7) into (6), we arrive at

$$I \approx |f(\widetilde{x})| \, \mathrm{vol}(\Omega) \, \frac{U_1(\widetilde{x})}{U_f(\widetilde{x})} \, \frac{1}{N} \sum_{i=1}^{N} \mathrm{sign}\left(f(x_i)\right), \tag{8}$$

where the mean value of the signum function in (6) is estimated by taking a mean of this function over a finite set of $|f|$-distributed states $x_i, i = 1, \ldots, N$, in Ω, see (3) and (4). In equation (8), the choice of the representative $\widetilde{x}$ is free. In order to improve the estimation of the integral, the mean over different representatives $\widetilde{x}_j, j = 1, \ldots, M$, can be taken. Thus,

$$I \approx \mathrm{vol}(\Omega) \left(\frac{1}{M} \sum_{j=1}^{M} |f(\widetilde{x}_j)| \, \frac{U_1(\widetilde{x}_j)}{U_f(\widetilde{x}_j)} \right) \left(\frac{1}{N} \sum_{i=1}^{N} \mathrm{sign}(f(x_i)) \right). \tag{9}$$

3. Numerical Tests

For an investigation of the quadrature method (9), we simulate different integrands over a two-dimensional domain $\Omega = [-3, 3] \times [-3, 3]$ with $\mathrm{vol}(\Omega) = 36$. Four different integrands $f_1, \ldots, f_4 : \Omega \to \mathbb{R}$ are tested. The integrands vanish, $f_i(x, y) = 0$,

integrand	exact integral	method (9)	standard MC
f_1	11	11.22 ± 0.45	11.01 ± 0.15
f_2	-4.29	-4.52 ± 0.45	-4.27 ± 0.15
f_3	55.92	57.86 ± 2.44	55.80 ± 1.18
f_4	-7.39	-7.79 ± 2.42	-7.51 ± 0.91

Table 1: The mean values and standard deviations evaluated for 100 Monte Carlo computations of the integrands $f_1, \ldots, f_4$.

for $(x, y) \in [-2.5, 2.5] \times [-2.5, 2.5]$, whereas, for (x, y) outside this box we define: $f_1(x, y) = 1, f_2(x, y) = \cos(y), f_3(x, y) = x(x-1), f_4(x, y) = x(x-1)\cos(y)$. Using a Markov chain Monte Carlo method, the four densities $|f_i|$ are sampled. A set of $N = 10000$ sampling points is generated for every integrand, see Figure 1. Since some of the densities have a disconnected support, we apply sampling methods with jumps [4]. For the estimation of the integrals $I_1, \ldots, I_4$, according to (9), we select $M = 20$ representatives (crosses in Figure 1). However, four representatives are not included in the calculation of the integrals of f_3 and f_4 because the corresponding function values $|f_i|, i = 3, 4$, are zero. In order to give one example: We estimate the integral I_2 of f_2. The mean value of the signum function for the 10000 sampling points is -0.5050. This is an estimation for the second factor in (6). The correct value for this factor is -0.5176. For the representative $\widetilde{x} = (2.75, 0)$ we counted 306 sampling points in an environment of $\widetilde{x}$ (a 0.5×0.5-box). If we had sampled 10000 points according to the equal distribution in Ω, we would expect a number of $625/9$ points in an 0.5×0.5-box. Thus, according to (8), the integral I is estimated as

$$I_2 \approx |\cos(0)| \cdot 36 \cdot \frac{625}{9 \cdot 306} \cdot (-0.5050) = -4.125.$$

Computing the mean value for all $M = 20$ representatives with this method, the approximation is -4.23, the correct value is $I_2 = -4.29$. We have applied (9) 100 times for each integrand with $M = 20$ and $N = 10000$ and compared this method with standard Monte Carlo quadrature in the domain Ω, see Table 1. The results of the new Monte Carlo approach is correct inside the interval of the standard deviation. Obviously, method (9) has a lot of drawbacks, e.g., difficulties to sample exactly from a distribution $|f|$ having a disconnected support. Additionally, the density-based method has a worse standard deviation than the standard Monte Carlo quadrature for

the above examples and there is a systematic discretization error. The discretization of the local density estimation must be coarse enough to yield a statistically relevant number of sampling points inside the environment U_f. However, in some cases in which standard Monte Carlo methods cannot be applied due to spasely occupied high dimensional spaces, the new method may be advantageous if the function f is almost zero in the major part of its domain. For such functions f, it may be advantageous (or even mandatory) to generate sampling points only in the "important" parts of its domain. The Boltzmann distribution in statistical thermodynamics is an example for this class of functions and will be investigated in the next sections.

4. Estimation of Entropy Differences

For a Boltzmann distribution of states, the differential entropy (2) can be expressed in terms of thermodynamic state functions, the *inner energy* U and the *free energy* A. The entropy difference between two systems is given by

$$\Delta S = \frac{\Delta U - \Delta A}{T}, \tag{10}$$

where T is the given temperature (in Kelvin) of the system. The probability for the occurrence of a certain state $q \in \Omega$ is given by the Boltzmann distribution w.r.t. the canonical ensemble. Via Markov chain Monte Carlo simulation we can sample from this distribution. If the system has a low entropy value (if it is very orderly) the sampled conformations will probably look very similar. If the entropy value is high, the Boltzmann distribution will be less focussed and the sampled conformations will appear very dissimilar. The denseness of the sampling is a measure for the entropy of the system. The denser the sampling, the lower the entropy value. In this section we will formulate this intuition more precisely. The first part of equation (10) is the computation of the inner energy difference ΔU. The inner energy of a system is defined as the mean energy value with regard to the distribution of states. In our case, it is the mean potential energy function $V_1 : \Omega_1 \to \mathbb{R}$ and $V_2 : \Omega_2 \to \mathbb{R}$ respectively, i.e.,

$$\begin{aligned} \Delta U &= \int_{\Omega_2} V_2(q) \frac{\exp(-\beta V_2(q))}{\int_{\Omega_2} \exp(-\beta V_2(\bar{q}))\, d\bar{q}}\, dq - \int_{\Omega_1} V_1(q) \frac{\exp(-\beta V_1(q))}{\int_{\Omega_1} \exp(-\beta V_1(\bar{q}))\, d\bar{q}}\, dq \\ &= \langle V_2 \rangle - \langle V_1 \rangle. \end{aligned} \tag{11}$$

The value of ΔU can be estimated by taking the mean potential energy value over a sampling of position states q_i according to the Boltzmann distribution. Again, this is a Monte Carlo approach with $h = V$, see (3) and (4). The next step in (10) is the computation of the free energy difference ΔA. We now apply the new quadrature method (8) for its estimation. The free energy A is defined as logarithm of the partition function, $A := -\beta^{-1} \ln(\int_\Omega \exp(-\beta V(q))\, dq)$. The integrand of the partition function is the Boltzmann expression $f(q) = \exp(-\beta V(q))$. This function is positive, i.e., the mean value computation of the signum function in (8) can be neglected.

$$\begin{aligned} \Delta A &= -\beta^{-1} \ln \left(\frac{\int_{\Omega_2} \exp(-\beta V_2(q))\, dq}{\int_{\Omega_1} \exp(-\beta V_1(q))\, dq} \right) \\ &\approx -\beta^{-1} \ln \left(\frac{\exp(-\beta V_2(q_2)) \operatorname{vol}(\Omega_2) \frac{U_{12}(q_2)}{U_{V2}(q_2)}}{\exp(-\beta V_1(q_1)) \operatorname{vol}(\Omega_1) \frac{U_{11}(q_1)}{U_{V1}(q_1)}} \right) \\ &\approx -\beta^{-1} \ln \left(\frac{U_{V1}(q_1)}{U_{V2}(q_2)} \right) + V_2(q_2) - V_1(q_1). \end{aligned} \tag{12}$$

For the last step of equation (12), it is important that the two position spaces Ω_1 and Ω_2 have a comparable structure. More precisely, the definition of an environment U_{11} in Ω_1 and of U_{12} in Ω_2 has to be comparable, i.e., the product of the volume $\operatorname{vol}(\Omega_i)$ of a position space $(i = 1, 2)$ multiplied with the expected number of equally distributed states that can be found inside the environments U_{1i} should be equal for Ω_1 and Ω_2. Together with the estimation of ΔU in (11) and the definition of ΔS in (10), we arrive at

$$\Delta S \approx \frac{[\langle V_2 \rangle - V_2(q_2)] - [\langle V_1 \rangle - V_1(q_1)]}{T} + k \ln \left(\frac{U_{V1}(q_1)}{U_{V2}(q_2)} \right). \tag{13}$$

In practise, the potential energy functions V_1 and V_2 can be modelled except for an unknown additional constant $c_i, i = 1, 2$. Fortunately, these constants cancel out in (13). This is not true for ΔA in (12) or ΔU in (11). Free energy differences or inner energy differences cannot be calculated with the available modelling of potential energies if c_1 and c_2 are different and unknown constants. Note, that the Boltzmann distribution of states, however, does not depend on additional constants of the potential energy function. Entropy is a measure for the disorder of this distribution. In equation (13), the choice of the representatives is free. If one selects two representatives q_1 and q_2 having a mean potential energy value, i.e., $V_1(q_1) = \langle V_1 \rangle$ and

$V_2(q_2) = \langle V_2 \rangle$ respectively, equation (13) is simplified to

$$\Delta S \approx k \ln \left(\frac{U_{V1}(q_1)}{U_{V2}(q_2)} \right). \tag{14}$$

Thermodynamical entropy is a mystical quantity for many people. One reason may be that seemingly there are different definitions for the term "entropy". But the common insight that we gain from entropy is that a system does not like to be orderly. Equation (14) clearly expresses this relation between entropy and the density of sampling points. Equation (14) is valid if the representatives have a mean potential energy value and the estimation of local densities are comparable for Ω_1 and Ω_2. The choice of the representatives q_1 and q_2 in (12) is free. In equation (9), this fact has been used for an estimation of the integral I by taking a mean over different representatives. In the case of ΔS, the corresponding formula is

$$\Delta S \approx \frac{\langle V_2 \rangle - \langle V_1 \rangle}{T} + k \ln \left(\frac{M_1 \sum\limits_{j=1}^{M_2} \frac{\exp(-\beta V_2(q_j^{(2)}))}{U_{V2}(q_j^{(2)})}}{M_2 \sum\limits_{j=1}^{M_1} \frac{\exp(-\beta V_1(q_j^{(1)}))}{U_{V1}(q_j^{(1)})}} \right). \tag{15}$$

A further simplification can be done analogously to equation (14): For a special choice of the representatives, i.e., representatives $q_j^{(1)}, j = 1, \ldots, M_1$, and $q_j^{(2)}, j = 1, \ldots, M_2$, with a mean potential energy value $V_1(q_j^{(1)}) = \langle V_1 \rangle$ and $V_2(q_j^{(2)}) = \langle V_2 \rangle$ respectively the estimation (15) can be simplified to

$$\Delta S \approx k \ln \left(\frac{M_1 \sum_{j=1}^{M_2} [U_{V2}(q_j^{(2)})]^{-1}}{M_2 \sum_{j=1}^{M_1} [U_{V1}(q_j^{(1)})]^{-1}} \right). \tag{16}$$

5. Entropy differences in practise

Due to the simplifications that have been done so far, we can only compare surface polymers which are structurally very similar. We want to explain these simplifications mathematically in order to get hints for the correct definition of environments U_{V1} and U_{V2} in equation (16). These environments must include a statistically relevant number of sampling points. In this context, a projection $\pi : \Omega \to \mathbb{R}^d$ from the state space Ω to a low-dimensional space $\mathbb{R}^d$ is necessary. But how should this projection look like? A major simplification has been done in equation (22), where we assumed

that the definitions of environments in Ω_1 and Ω_2 are comparable. This can be achieved by projecting the full-dimensional conformation space Ω to the cartesian coordinates $\mathbb{R}^d$ of a subset of (essential) atoms. In order to get a comparable definition for U_{V1} and U_{V2}, we take the same number of atoms in case of Ω_1 and Ω_2. A second question is, how to measure distances in $\mathbb{R}^d$? Since rotation and translation are not essential for the structure and potential energy of the molecules, we can, e.g., compute the root mean square distance RMSD between the representative and the sampling points (RMSD only restricted to the selected set of atoms). The last question is, how to select essential atoms? In equation (7), we replace the computation of the ratio of densities at a certain representative $\widetilde{x}$ by the computation of this ratio for a histogram, i.e., (theoretically) for a decomposition of Ω into *bins* $B_1, \ldots, B_L$. The correct computation of entropy (2) for such a decomposition of Ω would be

$$S = -k \sum_{i=1}^{L} \int_{B_i} \ln(p(q))\, p(q)\, dq = \sum_{i=1}^{L} S_i, \tag{17}$$

where $S_i = -k \int_{B_i} \ln(p(q))\, p(q)\, dq$. A histogram-based (quantized) evaluation of entropy reads

$$S \approx -k \sum_{i=1}^{L} w_i \, \ln(w_i), \tag{18}$$

where $w_i = \int_{B_i} p(q)\, dq$ is the statistical weight of bin B_i. Note, that a quantized entropy (18) is always nonnegative, whereas a differential entropy (17) can have a negative value. By the quantization we neglected that each bin B_i has its own entropy value $\overline{S}_i$:

$$\begin{aligned} \overline{S}_i &= -k \int_{B_i} \ln\left(\frac{p(q)}{\int_{B_i} p(\bar{q})\, d\bar{q}}\right) \frac{p(q)}{\int_{B_i} p(\bar{q})\, d\bar{q}} dq \\ &= -\frac{k}{w_i} \int_{B_i} p(q)\, [\ln(p(q)) - \ln(w_i)]\, dq \\ &= -\frac{k}{w_i} \left(\int_{B_i} p(q)\, \ln(p(q))\, dq - w_i \, \ln(w_i) \right) \\ &= \frac{1}{w_i} S_i + k \, \ln(w_i). \end{aligned} \tag{19}$$

Thus, the difference between the correct (17) and the histogram-based view (18) is given by:

$$S - \big(-k \sum_{i=1}^{L} w_i \, \ln(w_i) \big) = \sum_{i=1}^{L} w_i \, \overline{S}_i. \tag{20}$$

We want to compute the entropy difference between two molecular systems, i.e., the right hand side of (20) should be identical for the two systems. We can organize the decomposition of Ω in such a way that the entropy $\overline{S}_i$ is comparable for all subsets B_i. For the computation of entropy differences this means that the entropy value $\overline{S}_i$ for the RMSD-environments of the representatives in (19) should be identical for the two molecular systems. The *selected* atoms are quasi-fixed because we only take a small RMSD-environment of the representative $\tilde{q}$ into account. In this case, we assume a Gaussian normal distribution for *neglected* atoms. The selected atoms are also quasi-fixed, if the projection π preserves all structural relevant information of the molecule (it only cancels out quasi-redundant degrees of freedom). The quality of the projection method can be investigated. Analogously to the RMSD computation, we can also eliminate rotational and translational degrees of freedom of the molecules for a parameter estimation of the assumed normal distribution of the neglected degrees of freedom (by applying an alignment algorithm). The entropy value of a multivariate normal distribution with covariance matrix Σ in an n-dimensional space is given by

$$\overline{S}_i = k \ln \left(\sqrt{(2\pi e)^n |\Sigma|} \right), \tag{21}$$

where $|\Sigma|$ is the determinant of Σ. If this quantity $\overline{S}_i$ is almost identical for the two molecular systems, the systems are comparable via entropy estimation (16).

6. Numerical results

Now, we want to present a possible application of an entropy estimator. In statistical thermodynamics, entropy can be related to other state functions like inner energy U or free energy A. From (10) we gain

$$\Delta A = \Delta U - T\Delta S. \tag{22}$$

The free energy difference $\Delta A = A_2 - A_1$ between two systems in equation (22) determines the preferred direction of the reversible transition $1 \leftrightarrow 2$. If $\Delta A > 0$, system 1 is preferred, whereas system 2 is prefered if $\Delta A < 0$. Consider the situation in Figure 2, a protein binds to polymer molecules which are immobilized on a surface [8]. System 1 is the "unbounded state", system 2 the "bounded state". Thus, ΔA is named

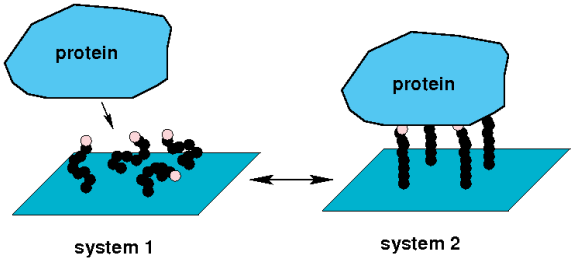

Figure 2: Surface polymers and protein in an unbounded state (system 1) and in a bounded state (system 2). The binding process reduces the entropy of the surface polymers.

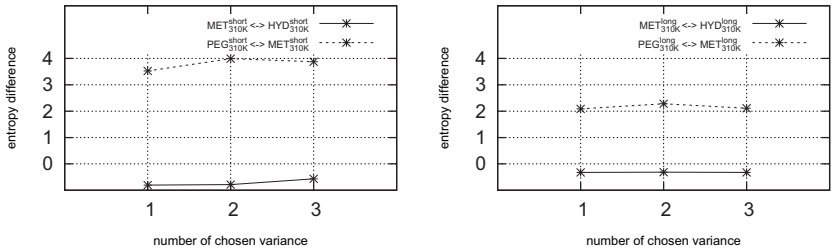

Figure 3: Left: Differential entropy estimation for polyethylenglycol ("peg") and its derivates "hyd" and "met" with a chain length of six subunits (short). Right: Plot for longer glycol chains with a chain length of nine subunits calculated again at a temperature of 310K (long).

binding affinity in the following. We want to compare the binding affinities for different polymer structures. For the sake of convenience, we will assume that the inner energy differences ΔU of the binding processes are comparable for all polymer structures because the head groups (where the protein binds to) of the different polymers are identical. Thus, the binding affinity varies due to an entropic effect. The binding process induces an ordering of the surface polymers – entropy decreases. A decrease of entropy $\Delta S = S_2 - S_1 < 0$ promotes system 1. The stronger the ordering effect the higher ΔA. If $\Delta A \gg 0$, the protein does not prefer the binding to the surface, i.e., the surface is protein resistant. Entropy differences between different immobilized polymer structures can be used to find out which one may be suitable to build up a protein resistant surface. In vacuum the arrangement of the polymer chains is

governed by the free energy of the polymer system. Their entropy as well as intra- and intermolecular hydrogen bonds will determine the arrangement of the polymer. In water the structural formation depends on the free energy of the polymer-aqueous system. Both subsystems, i.e. polymer chains and the surrounding water contribute with an energetic and entropic part to the total free energy. In our test cases we can assume that the inner energy differences ΔU_{water} and the entropy differences ΔS_{water} of the solvent are comparable for all three polymer systems because each one consists of nearly the same number of water molecules and the corresponding volume V_{water} stays constant. Especially this means that translational entropy changes of the solvent can be neglected [7].

The aim is the determination of a surface polymer with maximal entropy value S_1 among all possible similar surface polymers. For this purpose, we need a set of sampling data according to Boltzmann distribution. A theoretically well founded and practically well tested [5] method for the theoretical investigation of MD-simulations is the program package GROMACS. It was used with the GROMOS96 force field and the simple point charge water model [6] with rectangular periodic boxes with a 0.7nm solute-wall minimum distance. In this subsection we are performing some thermostated MD-simulations with this program package. The two panels in Figure 3 show the estimated entropy differences according to formula (16) (without Boltzmann constant k) between the polymer polyethylenglycol peg and its methylated polymer met, and between the methylated met and hydroxylated polymer hyd. The representative points were obtained by fixing an energy interval around the mean potential energy value. The number of states in a certain RMSD-environment of the representatives depends on the Å-size of this environment. The finally used parameters are summarized in Table 2. The selected atoms were defined by the common peg-substructure of the molecules.

	[kJ/mol] difference to mean energy	[Å] RMSD difference to representative
1	0.22	0.17
2	0.22	0.22
3	0.25	0.22

Table 2: Parameters that were used for formula (16).

The first results in Figure 3 refers to the three polymer structures with a length of six subunits, which are shown in the left panel. The comparison of differential entropy among these polymer structures reveals that peg has the highest flexibility. The second result investigates the differential entropy of the same polymer structures but with a longer chain length of nine subunits. The differential entropy is again higher when going from polymer met via polymer hyd to structure peg. Again, this qualitative picture shows the marginal increase of flexibility of hyd versus met to be much less significant than comparing peg with met or peg with hyd. The experimental observation is, that for short chain lengths peg is a good protein resistant substance, whereas hyd forms a more resistant surface for longer chain lengths. Moreover in further simulations it can be seen that the differential entropy fairly depends on different temperatures, which has to be tested in experiments. Protein resistant surfaces are just an example for the application of an entropy estimator. In general molecular dynamics (MD) does not sample a sufficiently large part of conformational space. For a better estimation of conformational entropy a meshless Hybrid Monte Carlo (HMC) algorithm and a decomposition approach like ZIBgridfree [9, 10] will be used in the future. Nevertheless, the new method according to formula (16) allows to evaluate a comparative index of mobility i.e. flexibility in terms of differential entropy.

Acknowledgements

This work has been supported by the Konrad-Zuse-Institut. Authors furthermore acknowledge fruitful discussions with Rainer Haag, Marie Weinhart and Peter Deuflhard.

References

[1] A. Lazo and P. Rathie (1978). On the entropy of continuous probability distributions. 24(1):120-122 IEEE Transactions on Information Theory.

[2] J. M. Hammersley and D.C. Handscomb (1964). Monte Carlo Methods. Methuen ISBN 0416523404.

[3] Ch. Chipot and A. Pohorille, editors (2007). Free Energy Calculations. Springer series in chemical physics 86. Springer ISBN 9783540384472.

[4] H. Senderowitz and W. C. Still (1998). MC(JBW): Simple but smart monte carlo algorithm for free energy simulations of multiconformational molecules. 19(15):1736-1745 Journal of Computational Chemistry.

[5] E. Lindahl, B. Hess, and D. van der Spoel (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. 7:306-317 J. Mol. Model.

[6] H.J.C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans (1981). Interaction models for water in relation to protein hydration. In: Intermolecular Forces. B. Pullman (editor). Reidel Publishing Company, Dordrecht, The Netherlands. 331-342.

[7] Y. Harano and M. Kinoshita (2005). Translational-Entropy Gain of Solvent upon Protein Folding. 89:2701 Biophys. J.

[8] C. Siegers, M. Bisalski, and R. Haag (2004). Self-assembled monolayers of dendritic polyglycerol derivatives on gold that resist the adsorption of proteins. 10:2831-2838 Chem. Eur. J.

[9] H. Meyer, M. Weber, and A. Riemer. ZIBgridfree. Software package for HMC simulation and conformation analysis. Software owned by Zuse Institute Berlin.

[10] M. Weber (2006). Meshless Methods in Conformation Dynamics. Doctoral Thesis, Freie Universität Berlin, Verlag Dr. Hut, Munich, Germany.