

RESEARCH ARTICLE

Distance-Based Configurational Entropy of Proteins from Molecular Dynamics Simulations

Federico Fogolari^{1,2*}, Alessandra Corazza^{1,2}, Sara Fortuna¹, Miguel Angel Soler¹, Bryan VanSchouwen³, Giorgia Brancolini⁴, Stefano Corni⁴, Giuseppe Melacini³, Gennaro Esposito^{1,2}

1 Dipartimento di Scienze Mediche e Biologiche, Universita' di Udine, Piazzale Kolbe 4, 33100 Udine, Italy, **2** Istituto Nazionale Biostrutture e Biosistemi, Viale medaglie d'Oro 305, 00136 Roma, Italy, **3** Department of Chemistry and Chemical Biology, McMaster University, 1280 Main St. W. Hamilton, ON L8S 4M1, Canada, **4** Center S3, CNR Institute Nanoscience, Via Campi 213/A, 41125 Modena, Italy

* federico.fogolari@uniud.it



OPEN ACCESS

Citation: Fogolari F, Corazza A, Fortuna S, Soler MA, VanSchouwen B, Brancolini G, et al. (2015) Distance-Based Configurational Entropy of Proteins from Molecular Dynamics Simulations. PLoS ONE 10 (7): e0132356. doi:10.1371/journal.pone.0132356

Editor: Giorgio Colombo, Consiglio Nazionale delle Ricerche, ITALY

Received: March 14, 2015

Accepted: June 13, 2015

Published: July 15, 2015

Copyright: © 2015 Fogolari et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper or in the references provided.

Funding: This work was partly supported by Ministero dell'Istruzione, dell'Universita' e della Ricerca (PRIN 2012A7LMS3 001). Oak Ridge National Laboratory by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy is acknowledged for the supercomputing project CNMS2013-064. Facilities of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, are also

Abstract

Estimation of configurational entropy from molecular dynamics trajectories is a difficult task which is often performed using quasi-harmonic or histogram analysis. An entirely different approach, proposed recently, estimates local density distribution around each conformational sample by measuring the distance from its nearest neighbors. In this work we show this theoretically well grounded the method can be easily applied to estimate the entropy from conformational sampling. We consider a set of systems that are representative of important biomolecular processes.

In particular:

- i. reference entropies for amino acids in unfolded proteins are obtained from a database of residues not participating in secondary structure elements;
- ii. the conformational entropy of folding of β 2-microglobulin is computed from molecular dynamics simulations using reference entropies for the unfolded state;
- iii. backbone conformational entropy is computed from molecular dynamics simulations of four different states of the EPAC protein and compared with order parameters (often used as a measure of entropy);
- iv. the conformational and rototranslational entropy of binding is computed from simulations of 20 tripeptides bound to the peptide binding protein OppA and of β 2-microglobulin bound to a citrate coated gold surface.

This work shows the potential of the method in the most representative biological processes involving proteins, and provides a valuable alternative, principally in the shown cases, where other approaches are problematic.

acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Entropy, weighted by the system temperature and counterbalanced by enthalpy is key in predicting the outcome of natural processes through the well-known Gibbs equation. At the microscopic level it determines the behavior of cell's complex machinery by regulating proteins' biophysical processes, such as folding and binding. These depend on several thermodynamic contributions to the free energy so that when the stabilization of structures causes large decreases of the entropic contribution, its understanding becomes essential to streamlining and controlling such processes.

Notwithstanding the efforts in this field the experimentally available information about the enthalpy and entropy of processes involving proteins or peptides is still not completely understood because it arises from several different contributions (electrostatic, hydrophobic, vibrational and from more extended molecular flexibility) which are difficult to disentangle.

In the following the configurational entropy will be addressed and the entropy arising from the different conformations of a molecule will be referred to as "conformational entropy" as usual in protein thermodynamics literature.

NMR has given a first idea of how relevant can conformational entropy be in protein ligand association by linking entropy changes to changes in measured order parameters (from protein relaxation data) or more generally protein dynamics [1, 2], but the quantification suffers from modeling and from the atomic probes available which are typically backbone HN internuclear vectors. More recently an NMR experiment able to provide entropy variations corresponding to global or partial unfolding has been proposed [3, 4], but still, although local processes can be addressed, it is impossible to disentangle conformational from solvation entropies.

Theoretical methods have been recently reviewed [5–7] in the context of protein-ligand complex formation and an in-depth discussion of the thermodynamics of binding, including a discussion of all entropic contributions, has been given by Gilson, McCammon and coworkers [8, 9].

In this respect molecular dynamics simulations have been seen as a way to compute conformational entropic contributions from the analysis of sufficiently long trajectories. Unfortunately the comparison with experimental data in this field is not straightforward because as the temperature is varied (as done in experiments aiming at measuring entropy) hydrophobic and electrostatic solvation energies are significantly changed. Also, at higher temperature protein structure becomes more loose and more conformational space may become available beyond what expected e.g. for an ensemble of oscillators.

Methodologies developed or at least fully exploited more or less recently could change the situation and make it possible to assess the various contribution to the folding or binding entropy from molecular dynamics simulations. In particular:

1. solvation terms could be obtained by the accurate implicit solvent representation based on the Onufriev Bashford and Case approximation to volume integrals [10], which allows also to speed up significantly simulations;
2. advanced sampling methods, reviewed by Adcock and McCammon [11] which allow on one hand a better sampling of the conformational space of the simulated systems and on the other hand, with some methodologies, to assess the effect of temperature on the conformational space of the molecule under study [12];
3. and lastly, the k^{th} nearest-neighbor estimate of entropy for a general probability distribution, proposed by Demchuk and collaborators [13] and developed by Hnizdo, Gilson and others [14–16], that allows an efficient estimation of entropy for multivariate probability

distributions. The latter method has been applied mainly by the developers and few other groups [14–23].

In this work we explore the application of the nearest-neighbor method to estimate the entropy for a few systems that are representative of important biomolecular processes like folding, allosteric transitions and ligand binding. Although results cannot be directly compared with experimental data, they agree with previous theoretical estimates and in all cases are in the expected range.

The present work adds to the few works that make use of the approach of Demchuk and coworkers, showing its potential in most common applications and develops to some extent the subject by addressing the entropy of folding and the rototranslational entropy loss upon binding.

The paper is organized as follows:

- i. methods are first reviewed and linked to the formalism used here;
- ii. reference entropies for amino acids in unfolded proteins are obtained from a database of residues not participating in secondary structure elements;
- iii. the conformational entropy of folding of $\beta 2$ -microglobulin [24–26] is computed from molecular dynamics simulations after subtraction of the reference entropies for the unfolded state;
- iv. backbone conformational entropy is computed for molecular dynamics simulations of four different states of the EPAC protein [2, 27, 28] and compared with order parameters (often used as a measure of entropy) and overall conformational entropy;
- v. the conformational and rototranslational entropy of binding is computed from simulations of 20 tripeptides bound to the peptide binding protein OppA [29, 30] and of $\beta 2$ -microglobulin bound to a citrate coated gold surface.
- vi. finally the results and the comparison with experimental data are discussed.

The main conclusion of this work is that for a set of systems representative of important biomolecular processes, where other approaches to the computation of entropies are problematic, the method based on the nearest neighbor can be readily applied to molecular dynamics trajectory as a valid alternative to other widely used methods.

Materials and Methods

The thermodynamic potential of a solvated molecule

In this subsection we recall the basic equations for the thermodynamic potential of a solvated molecule, following the book of McQuarrie [31] and the review of Gilson et al. [8] and Roux and Simonson [32]. We will use in the following molecular instead of molar quantities.

The standard chemical potential is expressed in terms of configurational integrals:

$$\mu_A^0 = -k_B T \log \left(\frac{8\pi^2}{C^0} \prod \left(\frac{2\pi m_i^A k_B T}{h^2} \right)^{\frac{3}{2}} \int \exp(-\beta(U(\vec{r}_A) + \Delta G_{\text{solv}}(\vec{r}_A, T))) d\vec{r}_A \right) + \frac{P^0 \bar{V}_A}{N_{Av}} \quad (1)$$

with $\beta = \frac{1}{k_B T}$, where k_B is the Boltzmann constant and T is the temperature, h is Planck's constant, m_i^A the mass of atom i of the solute and solute (and solvent, in the following) coordinates are indicated by \vec{r}_A (and \vec{r}_S , respectively). P^0 is the standard pressure, \bar{V}_A the partial molar volume of the solute, C^0 is the standard 1 M concentration and N_{Av} is Avogadro's number.

The solvation potential of mean force $\Delta G_{solv}(\vec{r}_A, T)$ is defined by the configuration integral:

$$\exp(-\beta \Delta G_{solv}(\vec{r}_A, T)) = \frac{\int \exp(-\beta(U_{AS}(\vec{r}_A, \vec{r}_s) + U_s(\vec{r}_s))) d\vec{r}_s}{\int \exp(-\beta U_s(\vec{r}_s)) d\vec{r}_s} \quad (2)$$

In what follows we will not consider the kinetic energy terms in the partition functions because they will cancel in processes involving isolated proteins, and the term $P^0 \bar{V}_A$ will be neglected because it is typically small for proteins and its variations upon conformational changes are negligible. For instance upon folding protein volume changes by less than 0.5 percent at standard pressure [33], which amounts to a mean free energy variation of just about 20 J/mol for a protein of 50 kDa. The picture would obviously change completely at high pressures.

We may use the probability distribution

$$p(\vec{r}_A, T) = \frac{\exp(-\beta(U(\vec{r}_A) + \Delta G_{solv}(\vec{r}_A, T)))}{\int \exp(-\beta(U(\vec{r}_A) + \Delta G_{solv}(\vec{r}_A, T))) d\vec{r}_A} \quad (3)$$

to express the entropy as:

$$S = -k_B \int p(\vec{r}_A, T) \log(p(\vec{r}_A, T)) d\vec{r}_A - \int \frac{\partial \Delta G_{solv}(\vec{r}_A, T)}{\partial T} p(\vec{r}_A, T) d\vec{r}_A \quad (4)$$

The first term is a classical conformational entropy term for the solute.

The second term is the entropy due to the solute-solvent interactions.

Note that the latter term cannot be simply estimated from the known or modeled dependence of $\Delta G_{solv}(\vec{r}_A, T)$ on the temperature because it depends also on the conformational space available which depends in turn on the temperature.

In the present work we are interested in the characterization of the first term, i.e. the entropy term arising from the conformational freedom of the protein.

Calculation of entropy from samples in conformational space

We follow here the approach of Singh et al. [13] developed and applied by Hnizdo, Gilson and coworkers [14, 15, 34, 35]. The approach considers n configurational samples for a random s -dimensional variable distributed according to the probability density $p(x_1, x_2, \dots, x_s)$ (or $p(\vec{x})$) for short). Given an s -dimensional sphere of radius r centered at \vec{x} a reasonable estimate for the local probability density ($\hat{p}(\vec{x})$) may be expressed by counting how many samples are found inside the volume of the sphere V_r . If this number is k , then $\hat{p}(\vec{x})$ is reasonably defined by the equation:

$$\hat{p}(\vec{x}) V_r = \frac{k}{n} \quad (5)$$

Note that in general V_r depends on the metric defined on the space which is considered. For euclidean distances the volume is given by:

$$\hat{V}_r = \frac{\pi^{\frac{s}{2}} r^s}{\Gamma(\frac{s}{2} + 1)} \quad (6)$$

where $\Gamma()$ is the Γ function.

With an estimate for the density the entropy can be estimated (we omit here the k_B factor for generality) as the average of $\log p(\vec{x})$:

$$S = - \int p(\vec{x}) \log p(\vec{x}) \approx - \frac{1}{n} \sum_i \log(\hat{p}(\vec{x}_i)) \quad (7)$$

For each sample i we take a sphere of radius r_i and count the number of samples k_i found in the sphere. By substituting Eqs 5 and 6 into the above equation we can reasonably approximate entropy by:

$$S \approx \frac{1}{n} \sum_i \log \left(\frac{n\pi^{\frac{s}{2}} r_i^s}{k_i \Gamma(\frac{s}{2} + 1)} \right) = \frac{s}{n} \sum_i \log(r_i) + \log \left(\frac{n\pi^{\frac{s}{2}}}{\Gamma(\frac{s}{2} + 1)} \right) - \frac{1}{n} \sum_i \log(k_i) \quad (8)$$

In the approach described by Singh et al. [13] the radius of the sphere centered about each conformational sample is taken as the distance to the k -th nearest neighbor and k_i is therefore the same for all samples. The above heuristic equations are corrected as to provide an unbiased estimator of the entropy for the probability distribution $p(\vec{x})$:

$$\hat{S}_k^{(n)} = \frac{s}{n} \sum_{i=1, n} \log R_{i,k} + \log \frac{n\pi^{\frac{s}{2}}}{\Gamma(\frac{1}{2}s + 1)} - L_{k-1} + \gamma \quad (9)$$

In the above equation $R_{i,k}$ is the distance of the i -th sample from its k -th nearest neighbor, $\Gamma()$ is, as before, the Γ function, γ is Euler's constant (0.5772...) and L_{k-1} is defined as: $L_0 = 0$, $L_j = \sum_{i=1, j} \frac{1}{i}$.

It is easy to recognize the similarity of the heuristic and exact equation.

Note that $R_{i,k}$ has a unit of measure and therefore the numerical value of $\hat{S}^{(n)}$ depends on the latter unit. This is consistent with the fact that in the equation for the entropy $S = -k_B \int p \log(p)$ the probability density p has units of measure.

To illustrate the symbols in the above equation consider a long molecular dynamics trajectory for a small molecule possessing only two rotatable bonds. The i -th (out of 1000) sample of the conformational space could be represented by the 2 torsional angles describing its conformation (assuming rigid bonds and angles) taken from the i -th snapshot. In this case the dimensionality of the system s would be 2 and the number of samples n would be 1000.

The distance d of each sample from all other samples would be computed according to the metric chosen (for torsional angles this is described below), and ranked, e.g. for the i -th snapshot $R_{i,k}$ would be then the k -th shortest computed distance. For each choice k we could use Eq (9) to estimate the configurational entropy of the molecule. In the following, unless otherwise stated, we will use $k = 1$, i.e. for each sample we consider the distance from its first nearest neighbor.

Limitations in the sampling capabilities of molecular dynamics simulations

A caution word is due about the possibility of sampling the conformational space of a molecule by molecular dynamics simulations. Indeed typical simulation times extend up to a hundred nanoseconds. It is likely that on this time scale sampling of the conformational space will not be complete. This consideration will apply to the approach used in this work as well as to other approaches found in the literature.

When comparing the entropy of different macrostates of the same molecule the effect of neglected accessible conformational space could be greatly reduced if the missing portion of sampled conformations is similar in the two simulations.

We will not address the issue here in detail as it affects similarly all methods that estimate conformational entropy from molecular dynamics simulations snapshots. For the shortest simulations we will only check convergence by considering the first and the last half of the

simulations and compare the results obtained from the two subsets of snapshots. Even such test does not guarantee that other conformations will not be accessible in longer simulation times.

The main interest here however is a proof of principle and to show that even from standard simulations it is possible to get valuable information about conformational entropy.

Calculation of the conformational entropy from molecular dynamics simulations

A widespread approach to compute the conformational entropy of a protein uses semi-classical quasi-harmonic analysis [36–41] or, in a refined approach, fitting probability distributions (see e.g. [42, 43]). Here we assume that the largest changes in the conformational entropy of a protein in most interesting biological processes arise from energetic restraints in the torsional freedom of backbone and sidechain groups. For this reason we consider torsional angles as variables and we neglect bond lengths and bond angles. The latter approximation amounts to assuming that the torsional degrees of freedom are decoupled from bond and angle degrees of freedom, which is mostly a good approximation for proteins, and that the vibrational partition function of bonds and angles is not significantly changed by conformational changes [38].

The latter observation might be system dependent. To give an estimate of the quality of this approximation, we considered the particular case of the residue undergoing the largest changes in conformational freedom between the holo/active and apo/inactive EPAC simulations described later. We computed the entropy using the method of Schlitter [40] as implemented in the software Carma [44] from molecular dynamics trajectories fixing or not the bond lengths and angles to their equilibrium values. In this way we were able to assess the bond and angle relative contributions to the entropy. The change in entropy between holo/active and apo/inactive, after fitting the residue to the first snapshot, was 0.264 kJ/(mol K) for free covalent geometry and 0.279 kJ/(mol K) for fixed covalent geometry. The effect appears therefore rather limited, compared to the overall computed entropy or the effect of the fitting reference chosen. This is in line with the observed changes in bond length and angle variances (about 6% with different signs). Note that the computed entropy change appears artifactually very large, possibly as a consequence of the adopted harmonic oscillator model.

We consider only the torsional degrees of freedom involving heavy atoms or polar hydrogens of serine, threonine, tyrosine, whereas proline, aromatic rings, amide and guanidinium moieties are taken as rigid. Other terminal hydrogens that could define torsional angles, but possess ternary symmetry, i.e. methyl and amino groups are considered free to rotate (or more generally we assume that their conformational distributions are not affected by the processes considered here) and therefore the contribution to the entropy of any conformational change will be zero for these moieties. For torsional angles distances we use the distance d_1 (d for simplicity here) studied in the work of Hnizdo and coworkers [20], which amounts to the circular distance in the torsional angle space. For two angular vectors $\vec{\theta}_1 = (\theta_{11}, \theta_{12}, \dots, \theta_{1n})$ and $\vec{\theta}_2 = (\theta_{21}, \theta_{22}, \dots, \theta_{2n})$ of n angles each, and with periodicities Θ_i the distance is defined as:

$$d = \sqrt{\sum_{i=1,n} \left(\frac{\Theta_i}{2} - \left| \frac{\Theta_i}{2} - |\theta_{1i} - \theta_{2i}| \right| \right)^2} \quad (10)$$

with θ_{1i} and θ_{2i} reported in the interval $[-\frac{\Theta_i}{2}, \frac{\Theta_i}{2}]$. It must be noted that on changing from cartesian to BAT (bond, angle, torsion) coordinates the Jacobian of the transformation enters the configurational integral [8, 45]. Under the assumption of rigid bond and angle geometry this extra term cancels out between reference (or initial) and final state of the system. The proof of

independence of the Jacobian from the configuration has been given by Go and Scheraga [46] cited in ref. [45] for a linear chain. The proof can be extended to a branched chain and for an arbitrary atom taken as the reference point for global translation. For this reason it will not be further considered here.

Nearest neighbor estimate of entropy is superior to classical harmonic analysis

It might be questioned whether the approach used here is needed at all when other methods are widely used. Most widely used methods are based on the quasi-harmonic approximation. The limitations and remedies to the quasi-harmonic approximation have been thoroughly described by McCammon and coworkers [42, 43] and Numata et al. [18]. We will not deal with this subject here but just provide a simple example demonstrating the issue.

If a system freely moving in one-dimension is considered and the range of movement is restricted, the quasi-harmonic approximation will give the correct change of entropy upon restriction. Indeed the square root of the variance for a uniform distribution is proportional to the range of the distribution itself. However it is sufficient to consider a toy model where we have a single freely rotatable bond (torsion angle θ) with three atoms fixed and the end atom moving, to see that the quasi-harmonic approximation breaks down.

We consider the unrestricted rotation as the reference state (probability density $\frac{1}{2\pi}$) and a restricted rotational macrostate where the rotational angle is freely varying between θ_1 and θ_2 with $\Delta\theta = \theta_2 - \theta_1$. The entropies of the two probability distributions are:

$$S_f = -k_B \int p_f \log(p_f) = k_B \log(2\pi)$$

and

$$S_r = -k_B \int p_r \log(p_r) = k_B \log(\Delta\theta)$$

with the difference in entropy from the free to the restricted macrostate being

$$\Delta S = -k_B \log\left(\frac{2\pi}{\Delta\theta}\right)$$

When we compute (classical) entropies using the harmonic approximation we must compute the variance-covariance matrix for two coordinates. The analysis will result in two harmonic oscillators whose classical entropy change from free to restricted macrostate is overestimating the true entropy for all choices of $\Delta\theta$. In particular for a reasonable restriction to an interval of $\Delta\theta = \pi/3$ the entropy is overestimated roughly by a factor 2, whereas it is estimated with an average root mean square error of 5 percent by the nearest neighbor method with just one hundred samples.

Calculation of unfolded proteins conformational entropy from database structures

Crystallographic structures of proteins are generally the most representative conformations adopted by proteins in their range of stability, and for this reason are considered a good representation of the native protein ensemble. However, the conformational preferences of amino acids observed experimentally in the folded structure are not excluded from the conformational diversity of the set of partially folded and unfolded conformations that proteins can adopt. Indeed it is reasonable to assume that a reversible process like folding (at least for small

proteins) simply freezes a set of amino acid conformations which can be adopted in the unfolded protein ensemble.

This principle has driven the refinement of available forcefields [47–49] and received an important validation from long time molecular dynamics simulations where the superiority of forcefield refined against experimental torsion angles distributions, and further refined afterwards, were able not only to fold small protein domains, but also to reproduce their folding rates [49, 50].

Based on the above considerations, we collected all possible conformations of all residues not directly involved in secondary structures of proteins, thus avoiding artifacts due to correlations, in the culled pdb dataset (3600 proteins) [51], as they provide samples of the conformational space available to aminoacids in their unfolded state, averaged over sequence details. This procedure assumes the presence of an identical interaction between the neighboring residues (i.e. the rest of the chain) and the unfolded and/or the database samples.

Therefore additional correlations between aminoacids in the unfolded state are not considered.

It should be noted that obtaining the same information from molecular dynamics simulations would not be easy because of the presumably largely asymmetric simulation box.

Obviously the accessible conformational space depends on the temperature, but it is reasonable to expect that this dependence is limited for stable proteins in their stability range. Indeed we can consider that for each torsional angle there is a number of preferred values (rotamers) and that the range of values explored about the most frequent ones increases with temperature. This situation is very similar to that of the harmonic oscillator. For a harmonic oscillator with frequency ν the vibrational partition function q_{vib} , the corresponding entropy S_{vib} and its derivative with respect to temperature, in the high temperature limit are given by:

$$q_{vib} = \frac{k_B T}{h\nu} \tag{11}$$

$$S_{vib} = k_B + k_B \log\left(\frac{k_B T}{h\nu}\right) \tag{12}$$

$$\frac{\partial S_{vib}}{\partial T} = \frac{k_B}{T} \tag{13}$$

If we approximate the effect of temperature on the conformational landscape of unfolded proteins with that of a collection of N harmonic oscillators, where N is the number of torsional degrees of freedom we obtain:

$$\frac{\partial S}{\partial T} \approx \frac{Nk_B}{T} \tag{14}$$

We may therefore expect that the dependence of the computed entropy on the temperature would be rather mild. We consider a typical range of temperatures of 30 K around 300 K, where most proteins are stable. To avoid misunderstanding we remark that here we are considering the effect of temperature on the variability of each single rotamer, i.e. about a well defined torsional energy minimum, and not different rotamers which are already sampled in the dataset.

The expected change in the entropy (per oscillator) corresponding to changes in distribution about the minimum energy rotamers due to temperature are limited by $k_B \frac{30}{300}$ which appears negligible compared to the overall entropy associated with the loss of conformational freedom.

All database proteins have been processed adding hydrogens by the program Reduce [52]. Torsional angles have been computed using the program Molmol [53]. Only rotatable torsional angles involving four heavy atoms or three heavy atoms and a polar hydrogen have been considered.

For polar hydrogens only residues involved in hydrogen bonds have been considered in order to avoid large artifactual entropy contributions from the preference in trans conformation for the program Reduce in the absence of favorable interactions. This choice was adopted for tyrosines, serines, threonines and cysteines. Cysteines involved or not in disulfide bridges were considered separately for analysis.

Distances in the torsional angles space are computed by taking euclidean distances in the periodic boundary n -dimensional angular space. In order to reduce the large conformational space, torsion angles involving different aminoacids have been considered as independent and therefore entropies are computed on a residue basis. The entropy corresponding to each aminoacid type, whose conformation is defined by the n -dimensional vector of torsion angles θ , is calculated with reference to the uniform distribution in the angular space, i.e.:

$$P_{ref}(\vec{\theta}) = \prod_{i=1,n} \frac{1}{\Theta_i} \quad (15)$$

$$S_{ref} = k_B \sum_{i=1,n} \log(\Theta_i) \quad (16)$$

where Θ_i is the periodicity of the i -th torsional angle, typically 2π .

Note that the volume of a n -dimensional sphere enters Eq (9) [34] and therefore suitable corrections to the equation should be applied when considering distances larger than half of the minimum periodicity of the torsional angles. This was not the case for our analysis because of a sufficiently large number of samples, leading to short nearest neighbor distances. Exact formulae have been reported by Hnizdo and coworkers [20].

The conformational entropy of folding

Wildtype $\beta 2$ -microglobulin was simulated for 30 ns, after equilibration at 320 K as previously described [54].

The entropy for each set of conformations was computed using Eq (9) with reference to the uniform torsion angle distribution exactly as in the previous section, taking 300 snapshots from the 30 ns molecular dynamics simulation. Then, the reference entropy computed from residues in the database representing unfolded amino acids was subtracted from the computed entropy, and the difference was taken as the conformational entropy of folding.

The conformational entropy of binding

The calculation of the binding free energy of association of biomolecules is very complex due to the large number of interacting bodies. In order to understand and possibly design novel interactions the free energy of binding is often approximated by the sum of different contributions, under the assumption that they can be decoupled. A derivation of this approach and an exhaustive discussion of its limitations has been given by Gilson et al. [8]. Typically the attention is focused on solute degrees of freedom and solvent degrees of freedom are taken into account using the potential of mean force [32]. Evaluation of the entropy loss upon interaction is not straightforward because the potential of mean force depends implicitly on the temperature.

We take the derivative of the Gibbs' free energy:

$$S_{A+B \rightarrow AB}^0 = - \frac{\partial \Delta G_{A+B \rightarrow AB}^0}{\partial T} \tag{17}$$

where:

$$\Delta G_{A+B \rightarrow AB}^0 = \mu_{AB}^0 - \mu_A^0 - \mu_B^0 \tag{18}$$

The choice of coordinates must allow easy calculation of the probability distribution $p()$. In this respect, following Gilson et al. [8], we choose the cartesian coordinates \vec{r}_A and \vec{r}_B of a reference atom on molecules A and B, respectively and other two atoms, assumed to be rigidly positioned with respect to the formers, on each molecule, to define the overall rotation state of the molecule.

We use as a criterion that the choice of the atoms must be the most insensitive with respect to dynamics in the bound state. This is done in order to minimize the correlation of these coordinates with other degrees of freedom, and therefore to decouple them from internal degrees of freedom.

For the complex we consider the coordinate system defined for molecule A as the reference one and we define the position of molecule B relative to \vec{r}_A , i. e. the position of molecule B is defined by $\vec{r}'_B = \vec{r}_B - \vec{r}_A$. The entropic contribution due to loss in rototranslational freedom is computed considering that all positions and orientations are still possible for the complex whereas the position and the orientation of molecule B is restricted, with respect to the position and orientation of the reference system of coordinates of molecule A.

Let us now denote by ξ_A and ξ_B the overall rotational degrees of freedom and by \vec{x}_A and \vec{x}_B the internal degrees of freedom.

With this notation the entropy of complex formation may be then written as:

$$S_{A+B \rightarrow AB} = -k_B \int_{bound} p(\vec{r}'_B, \xi_B, T) \log(p(\vec{r}'_B, \xi_B, T) \times 8\pi^2 V_0) d\vec{r}'_B d\xi_B \tag{19}$$

$$-k_B \int_{bound, internal} p(\vec{x}_A, \vec{x}_B, T) \log(p(\vec{x}_A, \vec{x}_B)) d\vec{x}_A d\vec{x}_B \tag{20}$$

$$+k_B \int_{free, internal} p(\vec{x}_A, T) \log(p(\vec{x}_A)) d\vec{x}_A \tag{21}$$

$$+k_B \int_{free, internal} p(\vec{x}_B, T) \log(p(\vec{x}_B)) d\vec{x}_B \tag{22}$$

$$- \int_{bound} \frac{\partial \Delta G_{solv}(\vec{x}_A, \vec{x}_B, T)}{\partial T} p(x_A, \vec{x}_B, T) \tag{23}$$

$$+ \int_{free} \frac{\partial \Delta G_{solv}(\vec{x}_A, T)}{\partial T} p(x_A, T) d\vec{x}_A \tag{24}$$

$$+ \int_{free} \frac{\partial \Delta G_{solv}(\vec{x}_B, T)}{\partial T} p(x_B, T) d\vec{x}_B \tag{25}$$

where the integrals over the bound state require a definition of the bound state, which is typically easy to attain if the bound state has a deep energy minimum. Otherwise it may be chosen as to adhere to experimental signal distinguishing bound from free molecules.

The first term of the above equation is difficult to treat and it has been traditionally estimated assuming that rotational and translational degrees of freedom are independent. The

latter assumption seems however poorly grounded and mostly motivated by the difficulty of treating the two degrees of freedom together. We will follow however the same approach here for simplicity.

Note that however, in order to use a distance based estimation of the entropy, a distance between rotation states must be defined. Rotation states are denoted equivalently by quaternions or 3×3 rotation matrices. There are several metrics which can be defined in the space of rotations [55]. It must be noted that Eq (9) assumes an euclidean distance in a cartesian space and therefore the number of points of randomly distributed n -dimensional variables within an n -dimensional distance d increases as $\frac{d^n}{\Gamma(\frac{n}{2}+1)}$.

Once a metric is defined we should compute the volume in rotation space corresponding to each distance and substitute it in the original Eq (9).

If we consider the representation of rotations which specifies the polar and azimuthal angles ϕ, ψ of the rotation axis and the rotation angle θ , the probability measure for a uniform distribution in rotation space is that with the axis of rotation uniformly distributed over the solid angle 4π and the rotation angle θ distributed in the range $[0, \pi]$ with probability density [56]:

$$\frac{2}{\pi} \sin^2\left(\frac{\theta}{2}\right) = \frac{1}{\pi} (1 - \cos(\theta)) \tag{26}$$

The metric we have used defines the distance d between two rotations described by 3×3 matrices R_1 and R_2 as:

$$d = \arccos \frac{\text{Tr}(R_2^{-1}R_1) - 1}{2} = \bar{\theta} \tag{27}$$

where Tr is the trace operator, $\bar{\theta}$ is the angle of rotation about the rotation axis for the composite rotation $R_2^{-1}R_1$.

With this metric the volume in rotation space is:

$$\frac{4\pi^2}{\pi} \int_0^{\bar{\theta}} (1 - \cos(\theta')) d\theta' = 4\pi(\bar{\theta} - \sin \bar{\theta}) \tag{28}$$

Note that here we considered rotations in a range of π and not 2π to avoid double counting of rotations, because a rotation by an angle θ about the vector \vec{v} is the same as the rotation by an angle $-\theta$ about the vector $-\vec{v}$.

Comparison with other methods

Entropies were also calculated according to other methods to provide comparison. A simple histogram method was applied where the range of the variables is divided in bins and the probability distribution is approximated as a piece-wise constant function on the bins. The constant value of the approximated probability distribution \hat{p}_i within bin i is the ratio of the number of counts n_i in the bin over the total number of counts n_c i. e. $\hat{p}_i = \frac{n_i}{n_c}$. The entropy is computed as $\hat{S} = -\sum_i \hat{p}_i \log \hat{p}_i$ with reference to the uniform distribution of the same number of counts on a specified interval.

For translational entropies the set of vectors was histogrammed in 3D and the entropy computed with respect to the standard 1 M reference state. The width of the 3D bins was $0.133 \text{ \AA} \times 0.133 \text{ \AA} \times 0.133 \text{ \AA}$ for the results reported here.

For rotational entropies the rotation matrices were converted into a set of three angles: two (ϕ, ψ) to specify the rotation axis, and one (θ) to specify the rotation around that axis. θ was allowed to vary only between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Since the random distribution of rotation is not

uniform over the three angular ranges, the range was divided in such a way that the integral over each bin was uniform. With this prescription the three angular ranges were divided in $10 \times 10 \times 100$ bins, respectively. The reference state was the uniform distribution of rotations.

As mentioned above, the method due to Schlitter [40] estimating entropies from the covariance matrix was used as implemented in the software Carma [44]. The source code was modified as to plot the contributions of each eigenvector to the entropy.

Molecular dynamics simulations

β 2-microglobulin. Molecular dynamics simulations of β 2-microglobulin were performed essentially as previously described [54]. Protons were added to the starting molecular structure (PDB code: 3HLA, chain B) [25] using the program `pdb2gmx` in the GROMACS software package [57]. Forcefield parameters (CHARMM v.27 [58] with the CMAP correction [47]) were assigned using the `psfgen` utility of the NAMD simulation software [59]. Ions were added as previously described [54] and TIP3P water [60] was added using the `solvate` module of the program VMD [61]. The simulation box was ca. 290000 \AA^3 and the number of atoms was 29638.

The temperature was set to 320 K, well below the β 2-microglobulin melting temperature (ca. 330 K), and controlled using Langevin Dynamics with a relaxation rate of 1 ps^{-1} . The pressure was set to 1.01325 bar and controlled using the Langevin Piston method [62, 63] with an oscillation time and a decay time of 100 fs. The timestep was 1 fs for bonded interactions, 2 fs for nonbonded interactions and 4 fs for long range electrostatic interactions.

10 ns equilibration followed by 30 ns production molecular dynamics simulations were run using the program NAMD v. 2.9b3.

Contact analysis was performed using home-written routines defining a contact whenever two atoms are closer than 1 \AA plus the sum of their van der Waals radii [64].

EPAC. The molecular dynamics simulations of the Exchange Protein directly Activated by Cyclic AMP (EPAC), which have been analysed in the present work, have been previously described by Melacini and coworkers [28]. The simulated portion of the protein entails residues 280 to 612 and 643 to 990, i.e. all the catalytic region except for the flexible, unresolved, loop 613–642 connecting the REM and the RA domains. Four EPAC states have been simulated: two equilibrium simulations for the cAMP bound (holo-) active state and apo-inactive state, and the two non-equilibrium apo-active and holo-inactive states. The simulations were run for 60 ns and the analysis was conducted on the last 50 ns. All the details about the simulations can be found in ref. [28].

OppA-tripeptide complexes. Molecular dynamics (MD) simulations were run on the free OppA protein (pdb code: 1RKM) and its complexes with tripeptides KLK (1B9J), KIK (1B3G), KDK (1B4Z), KNK (1B5I), KVK (1QKB), KAK (1JET), KFK (1B40), KMK (1B32), KGK (1B3L), KSK (1B51), KQK (1B5J), KKK (2OLB), KEK (1JEU), KTK (1B52), KHK (1B3F), KPK (1B46), KYK (1B58), KRK (1QKA), KCK (1B05), KWK (1JEV).

Crystallization water was removed, and the OppA histidines 55, 117, 142, 161, 405, 440 were protonated in all the systems. The protonation state was calculated using the program `H+` (available at URL <http://biophysics.cs.vt.edu/index.php>) [65] at pH 7 and ionic strength 0.05M. The system, neutralised with NaCl was set up in a cubic box with 0.6 nm buffer around the protein and 0.05M ionic concentration. We used `amber99` force field with `tip4p` water [48]. The cutoff for all the energies is 1.2nm. The system energy is minimized with the steepest descent algorithm as implemented in Gromacs. MD simulations were run for 45 nanoseconds (timestep of 0.002ps) for the free protein, its complexes, and the free peptides. We then sampled for the analysis 2,000 snapshots over the last 20ns.

β 2-microglobulin with citrate coated gold nanoparticles. The β 2-microglobulin structure was taken from the NMR solution structure (PDB id: 1JNJ). All titratable protein side chains, were assigned their standard protonation state at pH 7.7 with H⁺⁺ [65] corresponding to the experimental pH. Preliminary rigid-body docking simulations were carried out using Brownian dynamics (BD) techniques with the ProMetCS continuum solvent model [66] implemented in the SDA code (available at URL <http://projects.villa-bosch.de/mcmssoft/sda/6.00/>), from which the preferred binding site of the protein towards the citrate coated gold nanoparticles (cit-AuNPs) were extracted.

Temperature Replica-Exchange (T-REMD) simulations were started from the most representative and populated complexes resulted from rigid-body BD docking. More in details, 20 ns of unrestrained T-REMD of 32 replicas covering the temperature range between 290 and 320 K [67] were run yielding an aggregated simulation time of 640 ns. All simulations were based on the GoIP [68] force field with the SPC/E water model as implemented in the GRO-MACS package [69]. The lengths of bonds were constrained with the LINCS algorithm. Surface gold atoms and bulk gold atoms were frozen during all simulations but gold dipole charges were left free. Periodic boundary conditions and the Particle-Mesh-Ewald algorithm were used. A 2 fs integration time step was used. Analysis of the trajectories was performed over the last 10 ns of simulations. More details on the method and the model for the cit-AuNPs can be found in Brancolini et al. [70].

Results

Conformational entropy of unfolded proteins

In order to compute conformational entropy changes upon folding it is necessary to have a reference for the unfolded state. An explicit simulation of the unfolded state of a polypeptide chain would be unpractical because of both equilibrium constant, dictating the relative population, and kinetic constants coupled with free energy landscape, dictating the length of the simulation time.

Reference entropy values computed here may be used for computing conformational entropy differences from simulations of folded proteins, assuming, as usual, that aminoacid conformations in the unfolded state are essentially the same for all unfolded proteins. The assumption is in line with the typical lack of any secondary chemical shift observed in NMR spectra of unfolded proteins [71]. Conformational entropies for unfolded proteins (with reference to uniform torsional angle distribution) have been estimated from the conformation of aminoacids in public structural databases, as motivated in the Materials and methods section. Residues from the culled pdb dataset [51] not involved in secondary structures were taken as representative of unfolded conformations. The number of samples for the different aminoacids within protein sequence is reported in Table 1, together with entropies estimated from the nearest neighbor distances. Backbone angles ϕ or ψ were removed from the set of torsional angles to simulate residues at N- and C-termini, respectively. The entropy computed taking different k -nearest neighbors was reasonably consistent with the adopted choice of 1st-nearest neighbors. We report in Fig 1 the computed entropy for the alanine residue (only two torsional angles, ϕ and ψ , both with 2π periodicity) and those of valine, isoleucine, tryptophane and lysine (3, 4, 5 and 6 torsional angles, respectively) versus the average distance of the nearest-neighbors which increases as k increases. The entropy versus distance is reported instead of the entropy versus k because the average distance of the k -nearest-neighbors represents the resolution of the corresponding histogram approach to probability and entropy computation. As expected with increasing average distances details of the distribution are lost and the entropy increases. As the average distance approaches half of the minimum periodicity of the torsional

Table 1. Entropy and average nearest neighbor distance of unfolded aminoacids in k_B units. The row CYS (ϕ, ψ, χ_1) refers to cysteines involved in disulfide bridges. For N- and C-terminal residues see text.

aa	Entropy	av. dist.	N-terminal		C-terminal		counts
			Entropy	av. dist.	Entropy	av. dist.	
ALA	-2.4	0.5	-0.6	0.0	-1.0	0.0	23859
CYS	-4.7	15.0	-3.1	9.2	-3.3	8.5	845
CYS (ϕ, ψ, χ_1)	-3.6	4.7	-2.1	1.5	-2.3	1.3	3250
ASP	-4.5	5.5	-2.7	2.6	-2.9	2.4	27673
GLU	-5.2	11.3	-3.5	7.4	-4.0	6.7	19087
PHE	-4.9	7.0	-3.2	3.6	-3.7	3.1	11603
GLY	-1.9	0.5	-0.4	0.0	-0.5	0.0	43795
HIS	-4.4	8.5	-2.7	4.6	-3.1	4.0	8991
ILE	-6.6	4.8	-4.9	2.3	-4.9	2.2	12908
LYS	-7.5	16.8	-6.0	12.5	-6.3	11.6	18930
LEU	-6.3	4.5	-4.6	2.1	-5.2	1.8	22763
MET	-6.1	14.7	-4.5	10.1	-4.9	9.1	4848
ASN	-4.7	6.6	-3.0	3.3	-3.1	3.1	20987
PRO	-0.8	0.0	-0.8	0.0	0.0	0.0	27620
GLN	-5.5	13.7	-3.9	9.3	-4.4	8.3	11567
ARG	-6.9	18.2	-5.3	13.6	-5.7	12.7	15776
SER	-4.6	8.0	-3.0	4.2	-3.2	3.9	10053
THR	-5.1	7.2	-3.4	3.7	-3.6	3.5	8829
VAL	-4.6	2.2	-3.0	0.5	-3.0	0.5	17243
TRP	-4.8	9.2	-3.1	5.1	-4.1	23.5	4332
TYR	-5.4	17.2	-3.9	12.0	-4.2	10.9	3550

doi:10.1371/journal.pone.0132356.t001

angles, corrections should be made to the reference hyperspherical volume implied in Eq (9). In the absence of corrections artifactual curve behaviour is apparent. In the figure the onset of such effects takes place at 180 degrees. The entropy computed by the present method cannot be compared straightforwardly with the many estimates of sidechain entropy reported in the literature [72] because here the distribution of all residue torsional angles are considered together. In an approximate analysis we subtracted the entropy corresponding to alanine (as a model for all aminoacids backbone) from all other entropies and compared the resulting values with the values calculated by Koehl and Delarue [73]. The results correlate well ($r = 0.86$), although the estimated entropy is on average 1.8 times larger, most likely because of finer resolution due to larger database and different method. The trend of computed entropy with average nearest neighbor distance is observed in Fig 1.

Conformational entropy of a folded protein: β 2-microglobulin

The conformational entropy change upon folding is obtained in this section as the difference in the conformational entropy computed from molecular dynamics simulation between the folded state of the protein and the reference entropies for the unfolded state computed from database samples in the previous section.

The system chosen here is β 2-microglobulin which is a small, well characterized protein.

300 snapshots from 30 ns molecular dynamics simulations of β 2-microglobulin have been converted in torsional angles and entropy was computed, residue by residue. It should be noted

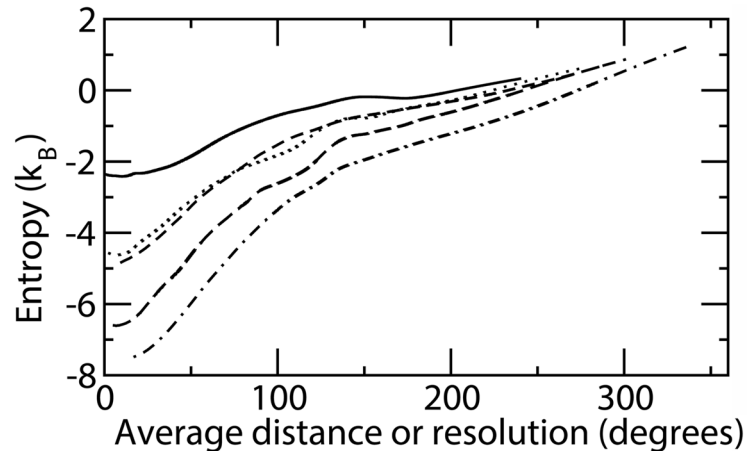


Fig 1. Entropy for residues in unfolded proteins calculated according to the k^{th} nearest neighbor vs. the average distance of the k^{th} nearest neighbor. Distances are increasing with k . The curves are relative to alanine (continuous line), valine (dotted line), isoleucine (short dashed line), tryptophane (long dashed line) and lysine (dot dashed line).

doi:10.1371/journal.pone.0132356.g001

that in the approach only correlations among degrees of freedom of the same residues are considered, whereas those among torsional angles of different residues would lead to very high dimensional space. This issue will be addressed in this subsection.

The entropies computed for the model residues in the unfolded protein (see previous subsection) have been subtracted from the computed entropies. The plot of computed entropy vs. residue number is sensitive to the nature of each residue, however, it is worth noting that the first and last residue display small but positive differential entropy. It is likely that this is due to the fact that these are really terminal residues, with more conformational freedom than the model ones, which were taken from residues in the sequence with the terminal torsional angles ignored. For all other residues the individual entropies range from $-6.1 k_B$ for R81 to -0.4 and $-0.7 k_B$ for R89 and K15 in the mobile FG and AB loop, respectively. It is reassuring that most mobile residues show entropies close to those of the unfolded residues.

The residues displaying the largest differential entropy are mostly located in stable regions of the protein. For instance, R81 ($-6.1 k_B$) is close to C80 involved in a disulfide bridge in the most stable region of the protein. R81 is partly exposed to the solvent but is involved in a salt bridge with residue D38. The picture is however somehow obscured by the individual nature of residues. In order to overcome this difficulty we consider the sum of entropies of residues clustering together. To this end we list for each residue the contacting residues and sum their computed entropies as representative of the entropy corresponding to a local opening reaction involving that residue. By local opening we mean the process of partial or global unfolding, where an amide hydrogen becomes exposed to the solvent, whose global entropy is measured e.g. by the Bluu-Tramp hydrogen-exchange NMR experiment [3, 4].

The results are reported in Fig 2 where the single residue and the cluster (i.e. the sum over neighboring residues) entropies are reported, together with experimental results. The latter include not only the conformational but also solvation entropy, and serve only the purpose to show the scale of the phenomena under consideration. It is interesting to note that the experimental entropy values which are most largely different from the conformational ones are those for the only three leucines for which experimental information is available, L39, L40 and L87

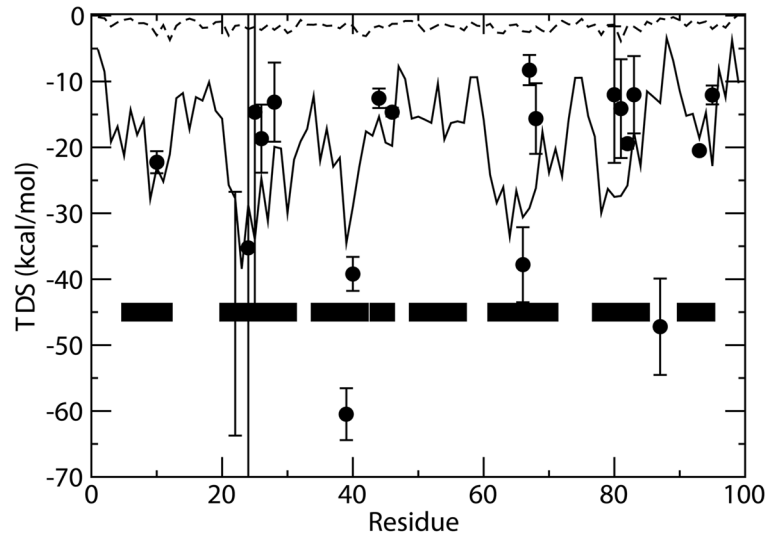


Fig 2. Entropy versus residue number from 30 ns β 2-microglobulin molecular dynamics simulation. Residue computed entropy (dashed line), summed entropy corresponding to residue neighborhood (continuous line). In the low part of the figure residues with defined secondary structure are indicated by black boxes. Black dots with error bars represent experimentally determined entropy from Bluu-Tramp hydrogen-deuterium exchange NMR experiments [3]. The latter experiments measure the enthalpy and entropy of the process (partial or global unfolding) which exposes protein amide hydrogens to solvent.

doi:10.1371/journal.pone.0132356.g002

and the two aromatic residues F22 and Y66. For these residues hydrophobic solvation contributions to entropy are expected to be relevant. For other hydrophobic or aromatic residues data (Y10, Y26, Y67, V82, V93, I46, W95) the same large difference is however not observed, pointing out the importance of specific details of local opening rather than a general hydrophobic effect, which are not captured by our simple model for local openings.

It should be noted that the approach followed here neglects correlations between degrees of freedom belonging to different amino acids. For a complete treatment of such correlations, approaches have been developed by Gilson and coworkers [15, 74]. The correlation within the same amino acid are expected to be large and actually many studies of protein conformations have demonstrated the correlation between secondary structure and side chain torsional angles [75]. It is however true that backbone angles of adjacent residues show often correlated motions which preserve the overall direction of the chain [76]. Such correlations have been observed very early in molecular dynamics simulations [77]. In order to assess the magnitude of this effect we computed the entropies S_{ϕ_i} and $S_{\psi_{i-1}}$ corresponding to the 1-dimensional distribution of angles ϕ_i and ψ_{i-1} separated by the peptide bond and the entropy $S_{\phi_i, \psi_{i-1}}$ corresponding to the 2-dimensional joint distribution. The difference ($S_{\phi_i} + S_{\psi_{i-1}} - S_{\phi_i, \psi_{i-1}}$), i.e. the mutual information of ϕ_i and ψ_{i-1} [74] was found to be positive, as expected, and on average of small entity ($0.13 \pm 0.05 k_B$). An extreme test for contacting side chains is provided by the degrees of freedom of cysteines involved in a disulfide bridge, where a strong correlation is expected. Also here the explicit calculation of the entropy corresponding to the two cysteines compared to the two isolated residues gave a positive difference ($1.5 k_B$) which however amounts to ca. 10% of the overall entropy of the two residues, before subtraction of free residues one, and 30% of the entropy after subtraction. We expect that in the absence of restraints on chain direction in the unfolded state, the correlation should be rather small. A similar test was performed for all torsional angles of R81 with all other torsional angles of β 2-microglobulin. The average mutual

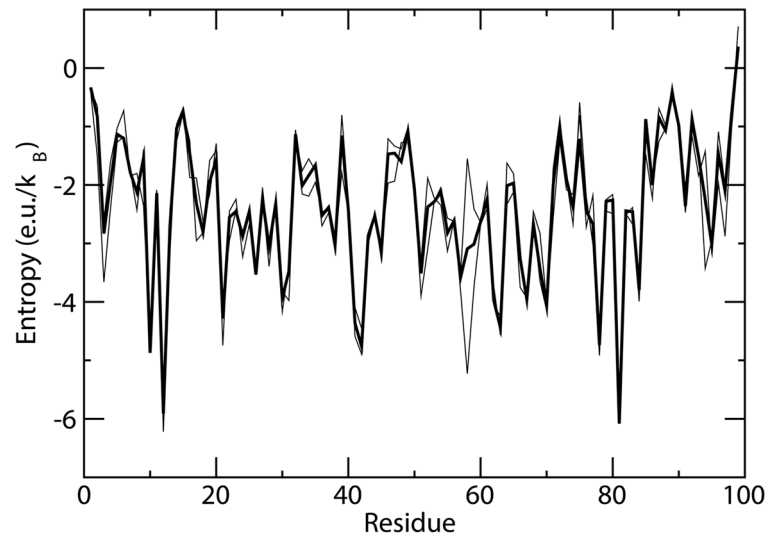


Fig 3. Entropy versus residue number from 30 ns β 2-microglobulin molecular dynamics simulation (thick line) and from the first and second half of the simulation (thin lines).

doi:10.1371/journal.pone.0132356.g003

information is about $0.06 k_B$. Overall these data suggest that correlations, other than those within the same aminoacids may give corrections to the conformational entropies which could be in the range of one order of magnitude smaller than the single amino acids conformational entropies of folding.

For this rather short simulation we checked also convergence. The trajectory was divided in two and the results for the separated halves were plotted together with the results for the total simulation time. As it can be seen in Fig 3 the results are rather stable with few exceptions. The largest difference is observed for K58 which loses close interaction with D59 in the second half of the simulation becoming more flexible.

Conformational entropy and order parameters

In this section we address the relationship between conformational entropies and order parameters, which are measured by NMR typically on HN vectors and are taken as a measure of conformational dynamics of each aminoacid.

The system chosen for this purpose is the exchange protein directly activated by cAMP (EPAC) which is a receptor undergoing functionally critical conformational and dynamics changes upon binding cAMP.

Changes in dynamics have been monitored by changes in HN order parameters, as measured by NMR, used, in turn, to measure changes in conformational entropy.

Four simulations, starting from the active and inactive conformations in the apo and holo forms, were previously performed, analysed and compared with experimental data by Melacini and coworkers [28]. We use the distance-based estimates of entropies for the four trajectories, which compared well with experimental data, and use this example to address the following issues:

1. how well the order parameters S^2 , typically measured on HN vectors using NMR, correlate with the entropy associated with backbone degrees of freedom;
2. how well the the entropy associated with backbone degrees of freedom correlates with the entropy of all torsional degrees of freedom.

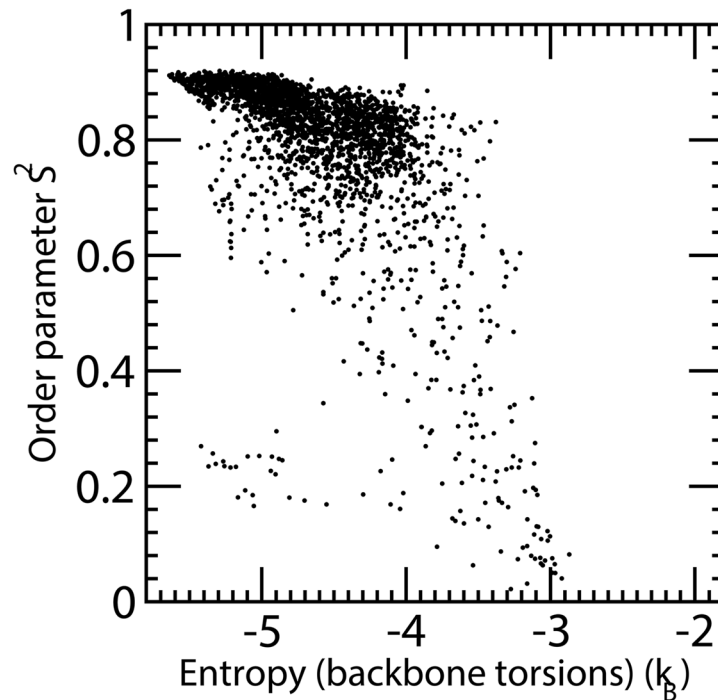


Fig 4. HN vector order parameter versus the computed backbone entropy on all four EPAC MD trajectories. The single simulations show some differences, but deviations from a sigmoidal behaviour is found for all four simulations.

doi:10.1371/journal.pone.0132356.g004

For the first question, a straightforward correlation of the order parameters of the HN vector with the conformational freedom of backbone torsion angles is not expected because the HN vector could be kept in place e.g. by a hydrogen bond even with floppy flanking peptide moieties. Rather, an agreement between order parameters and entropies averaged over chain segments should be expected. Indeed the raw correlation between backbone entropy and the order parameter computed on MD trajectories is poor, besides a necessary transformation between unbounded entropy and bounded order parameter. However, when the backbone entropy is smoothed using the lowess algorithm [78] implemented in the software R [79], with the parameter f set to 0.02, the (anti)correlation increases to -0.61 (Fig 4). The plot of order parameter versus the computed backbone entropy shows a semiquantitative relationship between the two. Notwithstanding the very different assumptions underlying the present calculation and the harmonic oscillation model used by Wand and coworkers [80], the points are scattered about a sigmoidal curve, thus confirming the insight provided by the original link between order parameters and associated entropy.

The changes in the order parameter S^2 occurring upon conformational transitions may be associated, based on the concept that they reflect overall conformational freedom changes, to changes in conformational entropy. For this reason we examine how backbone entropy, which is semiquantitatively measured by the order parameter S^2 , is correlated with overall conformational entropy. In other words we wish to estimate the latter from the measured HN order parameter S^2 .

We computed the entropy associated with backbone and all torsional angles for all four simulations and fitted the overall conformational entropy as a linear combination of backbone

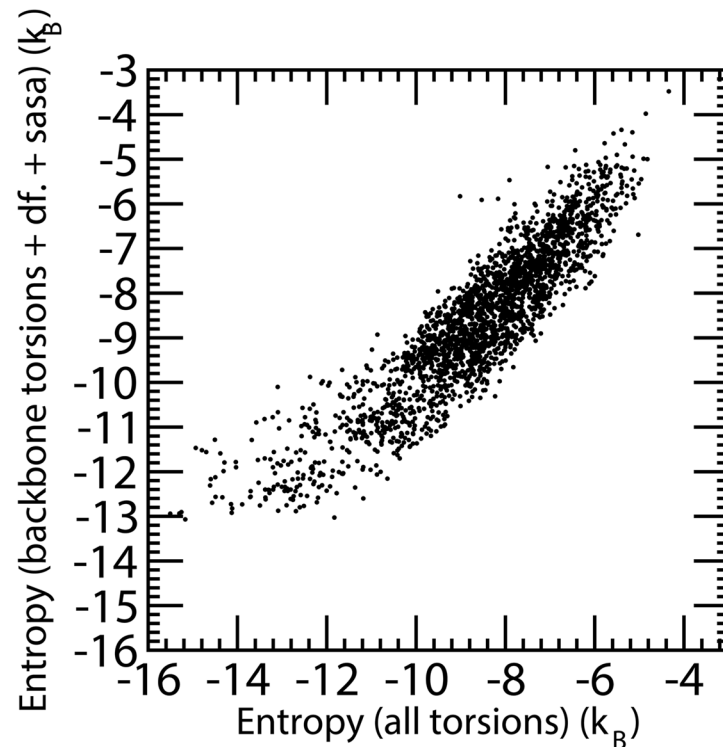


Fig 5. Residue entropy computed for all four EPAC MD simulations. On the x-axis the entropy is computed using all residue torsion angles, on the y-axis the entropy is fitted as a linear combination of backbone entropy, the number of sidechain degrees of freedom and the solvent accessible surface area of the residue. Glycine, Alanine and Proline residues that do not possess sidechain torsion angles, are not reported in the figure.

doi:10.1371/journal.pone.0132356.g005

conformational entropy (coefficient 1.120) and the number of sidechain rotational degrees of freedom (coefficient -1.425), assuming implicitly that each degree of freedom gives a similar entropic contribution. Alternatively the difference between the absolute entropy of all degrees of freedom and backbone degrees of freedom could be fit as linearly dependent on the number of sidechain degrees of freedom (coefficient -1.633) with similar results. Residues alanine, glycine and proline that do not possess torsions other than backbone ones were removed from the comparison. The fit shows a strong correlation with fitted data (correlation coefficient 0.859). This analysis does not take into account whether the residue is exposed or not. A slight modification of the approach considers also the exposed area of the residue. In this case the estimated absolute entropy of the residue is a weighted sum of the backbone entropy (weight: 1.08), of the number of degrees of freedom of the sidechain (weight: -1.78) and the exposed surface area in \AA^2 (weight: 0.012). With this simple linear model the correlation coefficient increases to 0.906 and the root mean square deviation of the predicted versus computed absolute entropy is $0.76 k_B$ (Fig 5).

Rototranslational entropy loss upon binding: OppA-tripeptide complexes and β 2-microglobulin association with a citrate coated gold surface

The free energy of binding of tripeptides with the E. coli peptide protein binding has been extensively studied by computational methods [30]. We address here the rototranslational

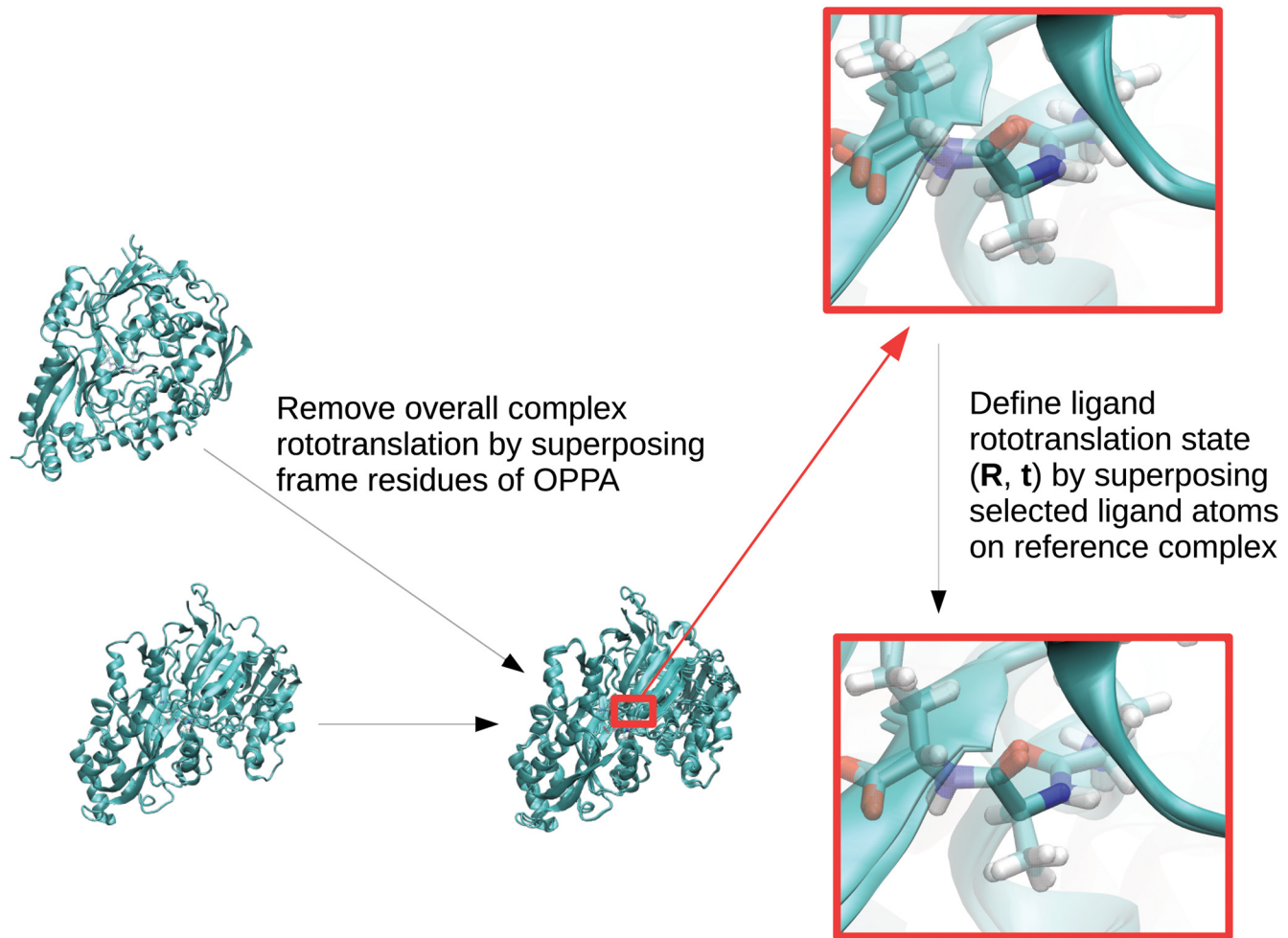


Fig 6. Graphical representation of how the rotation matrix and translation vector describing the rototranslational state of the tripeptide ligand with respect to OppA is obtained. First, overall rototranslation of the complex is removed by superposition of frame residues. Second, the central residue N, CA and C atoms are superimposed on the reference structure. The resulting rotation matrix and translation vector describe the rototranslational state of the tripeptide ligand.

doi:10.1371/journal.pone.0132356.g006

entropy loss upon binding. We first select an almost rigid three-atom fragment which is most decoupled from internal degrees of freedom of the peptide. To this end we consider the N-CA-C fragment of each of the three aminoacids and we compute the root mean square fluctuations of the three atoms upon superposition of the OppA regions involved in binding. All three aminoacids gave similar results, depending on the specific tripeptide, and therefore we kept the central aminoacid backbone atoms as the reference triad atoms.

Overall complex rototranslation and the movements internal to OppA protein were removed by superimposing backbone atoms of residues 32–37, 245–246, 269–271, 397–404, 414–419 of the OppA protein to the starting structure. The superimposed structure was then used for superposition of the three reference ligand backbone atoms and the resulting rotation matrix and translation vector were written. This procedure is illustrated in Fig 6.

The resulting computed entropies are reported in Table 2. The reference states are the randomly oriented rigid triad of backbone atoms ($S = -k_B \log(4\pi^2)$), for the rotational state, and

Table 2. Rotational and translational entropies in k_B units for the binding of 20 tripeptides to OppA protein. In parentheses values computed using the histogram method are reported. The average nearest-neighbor distance (in Å for translations and in rad for rotational distances) is reported in Av. dist. columns.

peptide	Rot. entropy	Av. dist.	Transl. entropy	av. dist.
KAK	-8.17 (-8.44)	0.017	-8.81 (-8.84)	0.04
KCK	-8.73 (-8.74)	0.014	-9.03 (-9.00)	0.04
KDK	-8.34 (-8.45)	0.016	-9.11 (-9.14)	0.03
KEK	-8.76 (-8.73)	0.014	-9.17 (-9.17)	0.03
KFK	-8.83 (-8.90)	0.014	-9.10 (-9.17)	0.03
KGK	-8.09 (-8.33)	0.018	-8.40 (-8.56)	0.04
KHK	-8.85 (-8.94)	0.013	-9.31 (-9.29)	0.03
KIK	-7.62 (-8.01)	0.020	-7.98 (-8.22)	0.05
KKK	-8.99 (-8.98)	0.013	-9.36 (-9.26)	0.03
KLK	-8.04 (-8.08)	0.018	-7.92 (-8.16)	0.05
KMK	-8.68 (-8.77)	0.014	-9.14 (-9.15)	0.03
KNK	-8.40 (-8.48)	0.016	-9.04 (-8.98)	0.04
KPK	-8.55 (-8.64)	0.015	-9.28 (-9.26)	0.03
KQK	-8.68 (-8.78)	0.015	-9.12 (-9.11)	0.03
KRK	-9.00 (-8.89)	0.013	-9.64 (-9.48)	0.03
KSK	-8.30 (-8.53)	0.016	-8.86 (-8.84)	0.04
KTK	-8.63 (-8.82)	0.015	-8.98 (-8.98)	0.04
KVK	-8.86 (-8.90)	0.014	-9.33 (-9.16)	0.03
KWK	-8.90 (-8.92)	0.013	-9.27 (-9.17)	0.03
KYK	-8.71 (-8.76)	0.014	-8.90 (-8.85)	0.04

doi:10.1371/journal.pone.0132356.t002

the 1M concentration state, for the translational state ($-k_B \log(1/1660)$), with angles measured in radians and distances in Å. The resulting rototranslational entropy of association at room temperature is ca. 10 kcal/mol. A similar value was predicted earlier based on semiquantitative considerations by others (e.g. [39, 81]).

Rotational and translational entropies have been computed using also the histogram method for the same set of 1000 snapshots. The results are dependent on the number of bins used. The latter was chosen as to limit the number of bins containing a single sample. With this choice the agreement is remarkable with correlation coefficients over 0.96 and mean square root difference of less than $0.2 k_B$ (Table 2).

In order to study a less tightly bound system we consider $\beta 2$ -microglobulin non-covalently bound to a gold nanoparticle coated with citrate. Here the protein is limited in its movements along the axis perpendicular to the surface plane and in its orientations. We have considered the distance of the center of geometry of the most structured regions from the plane through the oxygen atoms of the citrate bound to the gold surface. The reference rotation state is computed through superposition of the backbone atoms of the residues that contact most the surface, i.e. I1, K3, K58, W60 with the corresponding atoms in the first structure. After setting a common reference frame for the bound state, the rotation matrix is found by superposition of secondary structure elements. Because the bound state is defined irrespective of the orientation and position over the plane, the computed rotational entropy must be computed with a modified formula taking into account the reduction in dimensionality. The translational entropy is

computed by Eq (9) with dimension one and taking the distance between two sample points as the difference in the distance from the surface plane.

Similarly for the rotational entropy of the molecule in contact with the surface it should be considered that it should be corrected by an additional term $k_B \log(2\pi)$ to take into account any rotation about the axis perpendicular to the surface plane. With these corrections the overall translational and rotational entropy change from the 1M reference state to the surface bound molecule amounts to $-2.1 k_B$ and $-4.8 k_B$, respectively. These figures, which are considerably smaller than the corresponding ones for tight binding, could be underestimated because of the expected correlation between translational and rotational degrees of freedom for the bound state.

Discussion

In this work we have applied a recently developed method to estimate entropies based on the distances between conformational samples of proteins. We have considered a set of systems representative of important processes like protein folding and binding, where the computation of entropies may be problematic, e.g. when the distribution of conformational variables is multimodal and/or when the dimensionality and the number of samples prevents the use of histograms.

We have first derived reference entropies for single aminoacids in unfolded proteins. Conformational samples have been taken from the regions not involved in secondary structure elements in a non-redundant database of 3600 protein structures.

The set of reference entropies has been subtracted from the entropies computed from molecular dynamics simulation snapshots to obtain folding entropy contributions from each aminoacid of $\beta 2$ -microglobulin. The latter entropies (locally summed) have been compared with NMR measured entropies (which entail also solvation entropy) showing that the conformational entropy gives an important contribution to the measured data.

The effect of correlations between residues appears limited, at least as judged from the few examples reported in the Results section.

S^2 order parameters, which are often taken as a probe for local conformational entropy, have been compared with the entropy estimated for four molecular dynamics simulations of different EPAC states showing that:

- i. HN order parameter is correlated with the entropy of backbone degrees of freedom;
- ii. the relationship has the sigmoidal trend hypothesized by Wand and coworkers [80];
- iii. the entropy associated with backbone degrees of freedom is well correlated with the entropy of all degrees of freedom, when the number of sidechain degrees of freedom and solvent exposure is taken into account. These results support therefore the use of HN vector order parameter as a probe for overall entropy.

The rotational and translational entropy loss upon protein-ligand binding and upon protein-surface binding have been estimated, resulting in values compatible with those based on simpler models.

With all the limitations that can be envisaged, including lack of consideration of bond length and angle fluctuations, of correlations among degrees of freedom, poor sampling of conformational space, the estimate of entropy based on nearest neighbor distance provides in all the presented examples insight into the conformational and configurational entropy of the processes considered. The method can be readily applied to molecular dynamics snapshots and provides an entirely alternative route to estimate entropy where other more traditional

methods based on histograms or on variance-covariance matrix of fluctuations may prove difficult to apply or less accurate.

Acknowledgments

We thank prof. G. Scoles for useful discussion and suggestions. This work was partly supported by Ministero dell'Istruzione, dell'Universita' e della Ricerca (PRIN 2012A7LMS3_001). Oak Ridge National Laboratory by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy is acknowledged for the supercomputing project CNMS2013-064. Facilities of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, are also acknowledged.

Author Contributions

Conceived and designed the experiments: FF AC GE. Performed the experiments: FF AC GE SF MAS BV GB SC GM. Analyzed the data: FF AC GE SF MAS BV GB SC GM. Wrote the paper: FF AC GE SF MAS BV GB SC GM.

References

1. Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. *Nature*. 2007; 448:325–329. doi: [10.1038/nature05959](https://doi.org/10.1038/nature05959) PMID: [17637663](https://pubmed.ncbi.nlm.nih.gov/17637663/)
2. Das R, Mazhab-Jafari MT, Chowdhury S, SilDas S, Selvaratnam R, Melacini G. Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *J Biol Chem*. 2008; 283:19691–19703. doi: [10.1074/jbc.M802164200](https://doi.org/10.1074/jbc.M802164200) PMID: [18411261](https://pubmed.ncbi.nlm.nih.gov/18411261/)
3. Rennella E, Corazza A, Codutti L, Causero A, Bellotti V, Stoppini M, et al. Single-shot NMR measurement of protein unfolding landscapes. *Biochim Biophys Acta*. 2012; 1824:842–849. doi: [10.1016/j.bbapap.2012.04.002](https://doi.org/10.1016/j.bbapap.2012.04.002) PMID: [22522028](https://pubmed.ncbi.nlm.nih.gov/22522028/)
4. Rennella E, Corazza A, Codutti L, Bellotti V, Stoppini M, Viglino P, et al. Determining the energy landscape of proteins by a fast isotope exchange NMR approach. *J Am Chem Soc*. 2012; 134:4457–4460. doi: [10.1021/ja209004q](https://doi.org/10.1021/ja209004q) PMID: [22380530](https://pubmed.ncbi.nlm.nih.gov/22380530/)
5. Polyansky AA, Zubac R, Zagrovic B. Estimation of conformational entropy in protein-ligand interactions: a computational perspective. *Methods Mol Biol*. 2012; 819:327–353. doi: [10.1007/978-1-61779-465-0_21](https://doi.org/10.1007/978-1-61779-465-0_21) PMID: [22183546](https://pubmed.ncbi.nlm.nih.gov/22183546/)
6. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*. 2007; 36:21–42. doi: [10.1146/annurev.biophys.36.040306.132550](https://doi.org/10.1146/annurev.biophys.36.040306.132550) PMID: [17201676](https://pubmed.ncbi.nlm.nih.gov/17201676/)
7. Zhou HX, Gilson MK. Theory of free energy and entropy in noncovalent binding. *Chem Rev*. 2009; 109:4092–4107. doi: [10.1021/cr800551w](https://doi.org/10.1021/cr800551w) PMID: [19588959](https://pubmed.ncbi.nlm.nih.gov/19588959/)
8. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J*. 1997; 72:1047–1069. doi: [10.1016/S0006-3495\(97\)78756-3](https://doi.org/10.1016/S0006-3495(97)78756-3) PMID: [9138555](https://pubmed.ncbi.nlm.nih.gov/9138555/)
9. Wereszczynski J, McCammon JA. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q Rev Biophys*. 2012; 45:1–25. doi: [10.1017/S0033583511000096](https://doi.org/10.1017/S0033583511000096) PMID: [22082669](https://pubmed.ncbi.nlm.nih.gov/22082669/)
10. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins: Struct, Func, Gen*. 2004; 55:383–394. doi: [10.1002/prot.20033](https://doi.org/10.1002/prot.20033)
11. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*. 2006; 106:1589–1615. doi: [10.1021/cr040426m](https://doi.org/10.1021/cr040426m) PMID: [16683746](https://pubmed.ncbi.nlm.nih.gov/16683746/)
12. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*. 1999; 314:141–151. doi: [10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9)
13. Singh H, Misra N, Hnizdo V, Fedorowicz A, Demchuk E. Nearest neighbours estimates of entropy. *J Math Manag Sci*. 2003; 23:301–321.
14. Hnizdo V, Darian E, Fedorowicz A, Demchuk E, Li S, Singh H. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J Comput Chem*. 2007; 28:655–668. doi: [10.1002/jcc.20589](https://doi.org/10.1002/jcc.20589) PMID: [17195154](https://pubmed.ncbi.nlm.nih.gov/17195154/)

15. Hnizdo V, Tan J, Killian BJ, Gilson MK. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J Comput Chem.* 2008; 29:1605–1614. doi: [10.1002/jcc.20919](https://doi.org/10.1002/jcc.20919) PMID: [18293293](https://pubmed.ncbi.nlm.nih.gov/18293293/)
16. Huggins DJ. Comparing distance metrics for rotation using the k-nearest neighbors algorithm for entropy estimation. *J Comput Chem.* 2014; 35:377–385. doi: [10.1002/jcc.23504](https://doi.org/10.1002/jcc.23504) PMID: [24311273](https://pubmed.ncbi.nlm.nih.gov/24311273/)
17. Mukherjee A. Entropy Balance in the Intercalation Process of an Anti-Cancer Drug Daunomycin. *J Phys Chem Lett.* 2011; 2:3021–3026. doi: [10.1021/jz2013566](https://doi.org/10.1021/jz2013566)
18. Numata J, Wan M, Knapp EW. Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. *Genome Inform.* 2007; 18:192–205. doi: [10.1142/9781860949920_0019](https://doi.org/10.1142/9781860949920_0019) PMID: [18546487](https://pubmed.ncbi.nlm.nih.gov/18546487/)
19. Wang L, Abel R, Friesner RA, Berne BJ. Thermodynamic properties of liquid water: an application of a nonparametric approach to computing the entropy of a neat fluid. *J Chem Theory Comput.* 2009; 5:1462–1473. doi: [10.1021/ct900078k](https://doi.org/10.1021/ct900078k) PMID: [19851475](https://pubmed.ncbi.nlm.nih.gov/19851475/)
20. Misra N, Singh H, Hnizdo V. Nearest neighbor estimates of entropy for multivariate circular distributions. *Entropy (Basel).* 2010; 12:1125–1144. doi: [10.3390/e12051125](https://doi.org/10.3390/e12051125)
21. Nguyen CN, Young TK, Gilson MK. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys.* 2012; 137:044101. doi: [10.1063/1.4733951](https://doi.org/10.1063/1.4733951) PMID: [22852591](https://pubmed.ncbi.nlm.nih.gov/22852591/)
22. Fenley AT, Killian BJ, Hnizdo V, Fedorowicz A, Sharp DS, Gilson MK. Correlation as a Determinant of Configurational Entropy in Supramolecular and Protein Systems. *J Phys Chem B.* 2014; 118:6447–6455. doi: [10.1021/jp411588b](https://doi.org/10.1021/jp411588b) PMID: [24702693](https://pubmed.ncbi.nlm.nih.gov/24702693/)
23. Nguyen CN, Cruz A, Gilson MK, Kurtzman T. Thermodynamics of water in an enzyme active site: grid-based hydration analysis of coagulation factor Xa. *J Chem Theory Comput.* 2014; 10:2769–2780. doi: [10.1021/ct401110x](https://doi.org/10.1021/ct401110x) PMID: [25018673](https://pubmed.ncbi.nlm.nih.gov/25018673/)
24. Porcelli SA, Modlin RL. The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu Rev Immunol.* 1999; 17:297–329. doi: [10.1146/annurev.immunol.17.1.297](https://doi.org/10.1146/annurev.immunol.17.1.297) PMID: [10358761](https://pubmed.ncbi.nlm.nih.gov/10358761/)
25. Saper MA, Bjorkman PJ, Wiley DC. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J Mol Biol.* 1991; 219:277–319. doi: [10.1016/0022-2836\(91\)90567-P](https://doi.org/10.1016/0022-2836(91)90567-P) PMID: [2038058](https://pubmed.ncbi.nlm.nih.gov/2038058/)
26. Verdone G, Corazza A, Viglino P, Pettrossi F, Giorgetti S, Mangione P, et al. The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci.* 2002; 11:487–499. doi: [10.1110/ps.29002](https://doi.org/10.1110/ps.29002) PMID: [11847272](https://pubmed.ncbi.nlm.nih.gov/11847272/)
27. de Rooij J, Zwartkuis FJ, Verheijen MH, Cool RH, Nijman SM, Wittinghofer A, et al. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature.* 1998; 396:474–477. doi: [10.1038/24884](https://doi.org/10.1038/24884) PMID: [9853756](https://pubmed.ncbi.nlm.nih.gov/9853756/)
28. VanSchouwen B, Selvaratnam R, Fogolari F, Melacini G. Role of dynamics in the autoinhibition and activation of the exchange protein directly activated by cyclic AMP (EPAC). *J Biol Chem.* 2011; 286:42655–42669. doi: [10.1074/jbc.M111.277723](https://doi.org/10.1074/jbc.M111.277723) PMID: [21873431](https://pubmed.ncbi.nlm.nih.gov/21873431/)
29. Guyer CA, Morgan DG, Staros JV. Binding specificity of the periplasmic oligopeptide-binding protein from *Escherichia coli*. *J Bacteriol.* 1986; 168:775–779. PMID: [3536860](https://pubmed.ncbi.nlm.nih.gov/3536860/)
30. Wang T, Wade RC. Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. *J Med Chem.* 2002; 45:4828–4837. doi: [10.1021/jm020900i](https://doi.org/10.1021/jm020900i) PMID: [12383008](https://pubmed.ncbi.nlm.nih.gov/12383008/)
31. McQuarrie DA. *Statistical Mechanics.* Sausalito, CA, USA: University Science Books; 2000.
32. Roux B, Simonson T. Implicit solvent models. *Biophys Chem.* 1999; 78:1–20. doi: [10.1016/S0301-4622\(98\)00226-9](https://doi.org/10.1016/S0301-4622(98)00226-9) PMID: [17030302](https://pubmed.ncbi.nlm.nih.gov/17030302/)
33. Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. *Structure.* 1994; 2:641–649. doi: [10.1016/S0969-2126\(00\)00065-4](https://doi.org/10.1016/S0969-2126(00)00065-4) PMID: [7922041](https://pubmed.ncbi.nlm.nih.gov/7922041/)
34. Hnizdo V, Fedorowicz A, Singh H, Demchuk E. Statistical thermodynamics of internal rotation in a hindering potential of mean force obtained from computer simulations. *J Comput Chem.* 2003; 24:1172–1183. doi: [10.1002/jcc.10289](https://doi.org/10.1002/jcc.10289) PMID: [12820124](https://pubmed.ncbi.nlm.nih.gov/12820124/)
35. Darian E, Hnizdo V, Fedorowicz A, Singh H, Demchuk E. Estimation of the absolute internal-rotation entropy of molecules with two torsional degrees of freedom from stochastic simulations. *J Comput Chem.* 2005; 26:651–660. doi: [10.1002/jcc.20198](https://doi.org/10.1002/jcc.20198) PMID: [15751106](https://pubmed.ncbi.nlm.nih.gov/15751106/)
36. Karplus M, Kushick JN. Method for estimating the configurational entropy of native macromolecules. *Macromolecules.* 1981; 14:325–332. doi: [10.1021/ma50003a019](https://doi.org/10.1021/ma50003a019)

37. Levy RM, Karplus M, Kushick JN, Perahia D. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an alpha-helix. *Macromolecules*. 1984; 17:1370–1374. doi: [10.1021/ma00137a013](https://doi.org/10.1021/ma00137a013)
38. Karplus M, Ichiye T, Pettitt BM. Configurational entropy of native proteins. *Biophys J*. 1987; 52:1083–1085. doi: [10.1016/S0006-3495\(87\)83303-9](https://doi.org/10.1016/S0006-3495(87)83303-9) PMID: [3427197](https://pubmed.ncbi.nlm.nih.gov/3427197/)
39. Tidor B, Karplus M. The contribution of cross-links to protein stability: a normal mode analysis of the configurational entropy of the native state. *Proteins*. 1993; 15:71–79. doi: [10.1002/prot.340150109](https://doi.org/10.1002/prot.340150109) PMID: [7680808](https://pubmed.ncbi.nlm.nih.gov/7680808/)
40. Schlitter J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem Phys Lett*. 1993; 215:617–621. doi: [10.1016/0009-2614\(93\)89366-P](https://doi.org/10.1016/0009-2614(93)89366-P)
41. Andricioaei I, Karplus M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J Chem Phys*. 2001; 115:6289–6292. doi: [10.1063/1.1401821](https://doi.org/10.1063/1.1401821)
42. Minh DD, Bui JM, Chang CE, Jain T, Swanson JM, McCammon JA. The entropic cost of protein-protein association: a case study on acetylcholinesterase binding to fasciculin-2. *Biophys J*. 2005; 89:L25–27. doi: [10.1529/biophysj.105.069336](https://doi.org/10.1529/biophysj.105.069336) PMID: [16100267](https://pubmed.ncbi.nlm.nih.gov/16100267/)
43. Minh DD, Hamelberg D, McCammon JA. Accelerated entropy estimates with accelerated dynamics. *J Chem Phys*. 2007; 127:154105. doi: [10.1063/1.2794754](https://doi.org/10.1063/1.2794754) PMID: [17949130](https://pubmed.ncbi.nlm.nih.gov/17949130/)
44. Glykos NM. Carma: a molecular dynamics analysis program. *J Comput Chem*. 2006; 27:1765–1768. doi: [10.1002/jcc.20482](https://doi.org/10.1002/jcc.20482) PMID: [16917862](https://pubmed.ncbi.nlm.nih.gov/16917862/)
45. Hnizdo V, Gilson MK. Thermodynamic and Differential Entropy under a Change of Variables. *Entropy (Basel)*. 2010; 12:578–590. doi: [10.3390/e12030578](https://doi.org/10.3390/e12030578)
46. Go N, Scheraga HA. On the use of classical statistical mechanics in the treatment of polymer chain conformation. *Macromolecules*. 1976; 9:535–542. doi: [10.1021/ma60052a001](https://doi.org/10.1021/ma60052a001)
47. MacKerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem*. 2004; 25:1400–1415. doi: [10.1002/jcc.20065](https://doi.org/10.1002/jcc.20065) PMID: [15185334](https://pubmed.ncbi.nlm.nih.gov/15185334/)
48. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010; 78:1950–1958. doi: [10.1002/prot.22711](https://doi.org/10.1002/prot.22711) PMID: [20408171](https://pubmed.ncbi.nlm.nih.gov/20408171/)
49. Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? *Biophys J*. 2011; 100:L47–49. doi: [10.1016/j.bpj.2011.03.051](https://doi.org/10.1016/j.bpj.2011.03.051) PMID: [21539772](https://pubmed.ncbi.nlm.nih.gov/21539772/)
50. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic validation of protein force fields against experimental data. *PLoS ONE*. 2012; 7:e32131. doi: [10.1371/journal.pone.0032131](https://doi.org/10.1371/journal.pone.0032131) PMID: [22384157](https://pubmed.ncbi.nlm.nih.gov/22384157/)
51. Wang G, Dunbrack RL. PISCES: a protein sequence culling server. *Bioinformatics*. 2003; 19:1589–1591. doi: [10.1093/bioinformatics/btg224](https://doi.org/10.1093/bioinformatics/btg224) PMID: [12912846](https://pubmed.ncbi.nlm.nih.gov/12912846/)
52. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol*. 1999; 285:1735–1747. doi: [10.1006/jmbi.1998.2401](https://doi.org/10.1006/jmbi.1998.2401) PMID: [9917408](https://pubmed.ncbi.nlm.nih.gov/9917408/)
53. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*. 1996; 14:51–55. doi: [10.1016/0263-7855\(96\)00009-4](https://doi.org/10.1016/0263-7855(96)00009-4) PMID: [8744573](https://pubmed.ncbi.nlm.nih.gov/8744573/)
54. Fogolari F, Corazza A, Varini N, Rotter M, Gumral D, Codutti L, et al. Molecular dynamics simulation of beta2-microglobulin in denaturing and stabilizing conditions. *Proteins*. 2011; 79:986–1001. doi: [10.1002/prot.22940](https://doi.org/10.1002/prot.22940) PMID: [21287627](https://pubmed.ncbi.nlm.nih.gov/21287627/)
55. Huyinh DQ. Metrics for 3D rotations: comparison and analysis. *J Math Imaging Vis*. 2009; 35:155–164. doi: [10.1007/s10851-009-0161-2](https://doi.org/10.1007/s10851-009-0161-2)
56. Miles RE. On random rotations in R^3 . *Biometrika*. 1965; 52:636–639. doi: [10.1093/biomet/52.3-4.636](https://doi.org/10.1093/biomet/52.3-4.636)
57. van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, et al. Gromacs User Manual version 4.5.4; 2010. Available from: <http://www.gromacs.org>
58. MacKerell ADJ, Bashford D, Bellott M, Dunbrack RLJ, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J Phys Chem B*. 1998; 102:3586–3616. doi: [10.1021/jp973084f](https://doi.org/10.1021/jp973084f) PMID: [24889800](https://pubmed.ncbi.nlm.nih.gov/24889800/)
59. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, et al. NAMD2: greater scalability for parallel molecular dynamics. *J Comp Phys*. 1999; 151:283–312. doi: [10.1006/jcph.1999.6201](https://doi.org/10.1006/jcph.1999.6201)
60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79:926. doi: [10.1063/1.445869](https://doi.org/10.1063/1.445869)

61. Humphrey W, Dalke A, Schulten K. VMD Visual Molecular Dynamics. *J Mol Graph.* 1996; 14:33–38. doi: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: [8744570](https://pubmed.ncbi.nlm.nih.gov/8744570/)
62. Martyna G, Tobias D, Klein M. Constant pressure molecular dynamics algorithms. *J Chem Phys.* 1994; 101:4177–4189. doi: [10.1063/1.467468](https://doi.org/10.1063/1.467468)
63. Feller S, Zhang Y, Pastor R, Brooks B. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys.* 1995; 103:4613–4621. doi: [10.1063/1.470648](https://doi.org/10.1063/1.470648)
64. Berrera M, Molinari H, Fogolari F. Amino acid empirical contact energy definition for fold recognition in the space of contact maps. *BMC Bioinformatics.* 2003; 4:8. doi: [10.1186/1471-2105-4-8](https://doi.org/10.1186/1471-2105-4-8) PMID: [12689348](https://pubmed.ncbi.nlm.nih.gov/12689348/)
65. Anandakrishnan R, Aguilar B, Onufriev AV. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 2012; 40 (Web Server issue):W537–541. doi: [10.1093/nar/gks375](https://doi.org/10.1093/nar/gks375) PMID: [22570416](https://pubmed.ncbi.nlm.nih.gov/22570416/)
66. Kokh DB, Corni S, Winn PJ, Hoefling M, Gottschalk KE, C WR. ProMetCS: An Atomistic Force Field for Modeling Protein-Metal Surface Interactions in a Continuum Aqueous Solvent. *J Chem Theory Comput.* 2010; 6:1753–1768. doi: [10.1021/ct100086j](https://doi.org/10.1021/ct100086j)
67. Kirkpatrick S, D GC, Vecchi MP. Optimization by Simulated Annealing. *Science.* 1983; 220:671–680. doi: [10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671) PMID: [17813860](https://pubmed.ncbi.nlm.nih.gov/17813860/)
68. Iori F, Di Felice R, Molinari E, Corni S. GoIP: an atomistic force-field to describe the interaction of proteins with Au(111) surfaces in water. *J Comp Chem.* 2009; 30:1465–1476. doi: [10.1002/jcc.21165](https://doi.org/10.1002/jcc.21165)
69. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, Flexible, and Free. *J Comp Chem.* 2005; 26:1701–1718. doi: [10.1002/jcc.20291](https://doi.org/10.1002/jcc.20291)
70. Brancolini G, Corazza A, Vuano M, Fogolari F, Mimmi MC, Bellotti V, et al. Probing the Influence of Citrate-Capped Gold Nanoparticles on an Amyloidogenic Protein. *ACS Nano.* 2015; 9:2600–2613. doi: [10.1021/nn506161j](https://doi.org/10.1021/nn506161j) PMID: [25695203](https://pubmed.ncbi.nlm.nih.gov/25695203/)
71. Wüthrich K. NMR of Proteins and Nucleic Acids. A Wiley-interscience publication. New York: Wiley; 1986.
72. Doig AJ, Sternberg MJ. Side-chain conformational entropy in protein folding. *Protein Sci.* 1995; 4:2247–2251. doi: [10.1002/pro.5560041101](https://doi.org/10.1002/pro.5560041101) PMID: [8563620](https://pubmed.ncbi.nlm.nih.gov/8563620/)
73. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J Mol Biol.* 1994; 239:249–275. doi: [10.1006/jmbi.1994.1366](https://doi.org/10.1006/jmbi.1994.1366) PMID: [8196057](https://pubmed.ncbi.nlm.nih.gov/8196057/)
74. Killian BJ, Yundenfreund Kravitz J, Gilson MK. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J Chem Phys.* 2007; 127:024107. doi: [10.1063/1.2746329](https://doi.org/10.1063/1.2746329) PMID: [17640119](https://pubmed.ncbi.nlm.nih.gov/17640119/)
75. McGregor MJ, Islam SA, Sternberg MJ. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol.* 1987; 198:295–310. doi: [10.1016/0022-2836\(87\)90314-7](https://doi.org/10.1016/0022-2836(87)90314-7) PMID: [3430610](https://pubmed.ncbi.nlm.nih.gov/3430610/)
76. Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol.* 1976; 104:59–107. doi: [10.1016/0022-2836\(76\)90004-8](https://doi.org/10.1016/0022-2836(76)90004-8) PMID: [957439](https://pubmed.ncbi.nlm.nih.gov/957439/)
77. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature.* 1977; 267:585–590. doi: [10.1038/267585a0](https://doi.org/10.1038/267585a0) PMID: [301613](https://pubmed.ncbi.nlm.nih.gov/301613/)
78. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc.* 1979; 74:829–836. doi: [10.1080/01621459.1979.10481038](https://doi.org/10.1080/01621459.1979.10481038)
79. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2008. ISBN 3-900051-07-0. Available from: <http://www.R-project.org>
80. Li Z, Raychaudhuri S, Wand AJ. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* 1996; 5:2647–2650. doi: [10.1002/pro.5560051228](https://doi.org/10.1002/pro.5560051228) PMID: [8976574](https://pubmed.ncbi.nlm.nih.gov/8976574/)
81. Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. *Protein Eng.* 1989; 3:1–3. doi: [10.1093/protein/3.1.1](https://doi.org/10.1093/protein/3.1.1) PMID: [2813338](https://pubmed.ncbi.nlm.nih.gov/2813338/)