# Of fishes and birthdays: Efficient estimation of polymer configurational entropies

Ilya Nemenman,[1] Michael E Wall,[2] and Charlie E Strauss[3]

[1] *Departments of Physics and Biology, Emory University, Atlanta, GA 30322**

[2] *Computer, Computational, and Statistical Sciences*†

[3] *Biosciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545*‡

(Dated: February 9, 2015)

We present an algorithm to estimate the configurational entropy $S$ of a polymer. The algorithm uses the statistics of coincidences among random samples of configurations and is related to the catch-tag-release method for estimation of population sizes, and to the classic "birthday paradox". Bias in the entropy estimation is decreased by grouping configurations in nearly equiprobable partitions based on their energies, and estimating entropies separately within each partition. Whereas most entropy estimation algorithms require $N \sim 2^S$ samples to achieve small bias, our approach typically needs only $N \sim \sqrt{2^S}$. Thus the algorithm can be applied to estimate protein free energies with increased accuracy and decreased computational cost.

PACS numbers: 87.15.A-, 89.70.Cf

Computational estimation of protein free-energy differences (e.g., between ligand-bound and ligand-free states) is an unsolved problem with broad applications in molecular biology and medicinal chemistry. Present approaches to the problem may be divided into two classes: *difference methods*, in which the difference in free energy between two states is directly estimated, and *end-point methods*, in which absolute free energies are calculated for the states being compared. Difference methods suffer from slow convergence when there is little overlap between the states, though it may be possible to overcome this limitation [1]. In contrast, end-point methods are independent of the overlap between the states being compared. They are also trivially more efficient when pairwise free-energy differences among a large number of states are required. The main challenge for end-point methods is overcoming difficulties in estimating configurational entropy, which contributes substantially to protein free energy [2, 3]. Recent progress has followed several threads: calibration against reference potentials for which the free energy can be exactly calculated [4], selective sampling about local minima in the energy landscape [5], hierarchical estimation of chain-elongation transition probabilities using Monte Carlo simulations [6], and transforming the degrees of freedom to occupy minimally coupled subspaces [7]. While the calibration method uses only randomly sampled configurations and their corresponding energies, the others often require the ability to calculate the energy of an *arbitrary* configuration.

Entropy estimation has been recognized as a crucial problem in other disciplines, such as computational neuroscience and cell biology [8, 9]. Properties of entropy estimators have been studied extensively [10–12]. Bias, rather than variance, is the dominant problem. For common estimators, the bias $\langle \delta S \rangle$ is negative and scales as

$$\langle \delta S \rangle \equiv \langle S_{\rm est}(\mathbf{n}) - S_{\rm true}(\mathbf{p}) \rangle \propto -2^{S_{\rm true}}/N. \qquad (1)$$

Here $S_{\rm true}$ is the true entropy of the unknown probability distribution $\mathbf{p}$, $\dim \mathbf{p} = K$. $S_{\rm est}$ is the entropy estimated from the measured frequencies $\mathbf{n}$, $\sum_{i=1}^K n_i = N$, and the averaging $\langle \dots \rangle$ is taken over the random samples. Both $S_{\rm true}$ and $S_{\rm est}$ are measured in bits. In particular, the Maximum Likelihood (ML) estimator

$$S_{\rm ML}(\mathbf{n}) = -\sum_{i=1}^K \frac{n_i}{N} \log_2 \frac{n_i}{N}, \qquad (2)$$

which uses the observed frequencies instead of the unknown probabilities, has bias that scales as in Eq. (1). This sets the limit on data requirements for traditional configurational entropy estimation methods.

When the asymptotic bias follows Eq. (1), it can be subtracted from the estimate, making the latter nearly unbiased for $N \gg 2^{S_{\rm true}}$ [11, 13]. For $N < 2^{S_{\rm true}}$, universally unbiased estimation is impossible [10, 11]. Then a priori assumptions about the underlying probability distribution are needed to regularize the inference. One such Bayesian prior, $\mathcal{P}(\mathbf{p})$, is known as the NSB (Nemenman-Shafee-Bialek) method [14]. It has been useful in neuroscience, but to our knowledge has not yet been applied to macromolecular entropy estimation. The approach starts with noting that seemingly reasonable prior assumptions $\mathcal{P}(\mathbf{p})$ may result in unexpected assumptions $\mathcal{P}(S)$. For example, consider a family of Dirichlet priors over $\mathbf{p}$, indexed by a parameter $\beta$,

$$\mathcal{P}(\mathbf{p}|\beta) = \frac{1}{Z}\,\delta\left(1 - \sum_{i=1}^K p_i\right) \prod_{i=1}^K p_i^{\beta-1}. \qquad (3)$$

Here the first term normalizes $\mathcal{P}$, and the $\delta$-function constrains the normalization of the distribution $\mathbf{p}$ itself. The product of $p_i^{\beta-1}$'s introduces biases towards peaked ($\beta \to 0$) or uniform ($\beta \to \infty$) distributions $\mathbf{p}$. Maximum likelihood inference of $\mathbf{p}$ with this prior is equivalent to adding $\beta$ *pseudocounts* to every possible outcome $i$. Importantly, for large $K$, these pseudocounts bias the

resulting Bayesian entropy estimator strongly [14]. The entropy becomes "known" before any samples are measured! One sees this by calculating the a priori entropy expectation $S_0(\beta)$ and its rms error $\delta S_0(\beta)$ at a fixed $\beta$, and the latter turns out to be very small [14]. One then uses a new prior over $\mathbf{p}$ and $\beta$,

$$\mathcal{P}_{\rm NSB}(\mathbf{p},\beta) = \frac{1}{Z}\,\delta\left(1-\sum_{i=1}^{K} p_i\right)\prod_{i=1}^{K} p_i^{\beta-1}\,\frac{dS_0(\beta)}{d\beta}. \quad (4)$$

This is different from Eq. (3) by the Jacobian $dS_0(\beta)/d\beta$, which ensures that, in the limit of a narrow and monotonic a priori expectation of $S_0(\beta)$, one gets $\mathcal{P}_{\rm NSB}(S_0) = \int d\beta d\mathbf{p}\,\mathcal{P}_{\rm NSB}(\mathbf{p},\beta)\delta(S_0(\beta)-S(\mathbf{p})) \approx$ const. Ref. [14] has argued that this procedure creates a more uniform $\mathcal{P}(S)$ and reduces the estimation bias.

The NSB estimator is related to the familiar *birthday problem*: in a year with $K$ days, one only needs $N \sim \sqrt{K}$, but not $N \sim K$, individuals in a room to make it likely that there will be at least one shared birthday. The same idea is behind the *catch-tag-release* estimation of wildlife population sizes: in a pond with $K$ fishes, one will catch a fish that has been previously caught, tagged, and released after $N \sim \sqrt{K}$ fishes caught. It follows that one can estimate $K$ by counting how many previously tagged fishes were caught. If every fish has the same probability to be caught, then $S = \log_2 K$, and both $S$ and $K$ can be estimated with $N \sim \sqrt{K}$, compared to the usual methods that require $N \sim 2^{S_{\rm true}}$, cf. Eq. (1). Such estimation of entropies based on coincidence counting is known as the Ma estimator [15]. No general estimator can function reliably with fewer samples: if one never sees a repeat fish, then one only knows the minimum population size, but nothing about the maximum.

Unfortunately, since logarithms diverge near zero, the low-probability, poorly sampled tail of $\mathbf{p}$ contributes disproportionally to entropy. Thus coincidence counting cannot transfer easily to non-uniform probability distributions [11]. One needs to use the high-probability events (i.e., coincidences) and extrapolate to the tail. Then the prior $\mathcal{P}(\mathbf{p})$ may be seen as enforcing a certain shape of the tail, and NSB assumes that the tail is not too heavy [14, 16]. When the tail structure has been guessed correctly, the estimate, $S_{\rm NSB}(\mathbf{n})$, converges to the true entropy in the Ma regime, $N \sim 2^{1/2S_{\rm true}}$ [17]. For massive tails, typically there is bias, $\langle \delta S_{\rm NSB}\rangle < 0$. However, $|\langle \delta S_{\rm NSB}\rangle| < |\langle \delta S_{\rm ML}\rangle|$. To verify if a sample size dependent bias is present, one estimates $S_{\rm NSB}(\alpha N)$, where $0 < \alpha \leq 1$ is the fraction of data used. If the estimates at different $\alpha$ agree within the posterior error bars, the bias can be neglected compared to the variance [18].

We wondered whether NSB might improve estimation of configurational entropies of polymer chains. For this, we generated self avoiding random walks of different lengths on a 3D lattice [19–21]. We focused largely on chiral walks on a cubic lattice, so that the $n$'th bond in the chain is allowed to take, at most, three of the five possible orientations, depending on the orientations of the $n-1$'st and the $n-2$'nd bonds. Such chains weave chiral paths through the lattice, approximating the c-alpha secondary structure of real proteins [22]. We considered lattices with bounding cubes up to 4x4x4 (64 total sites) and homopolymeric chains of length $L \leq 50$. With these constraints, all self avoiding configurations and their energies can be enumerated on commodity computers. We then calculated the partition function by direct summation. This makes further sampling of random configurations trivial and decouples, for presentation purposes, the entropy estimation problem from the problem of efficient sampling, which is not the focus of this Letter. To ensure sufficient generality of our results, we explored chiral chains of lengths $16 \leq L \leq 50$, as well as short non-chiral chains. The results were similar for these cases. Thus here we present only chiral polymers with $L = 32$.

In the spirit of Ref. [23], we evaluated the energy of lattice conformations using energy functions that include local and long-range contributions: (1) backbone secondary structure (SS) propensity, measured by a preference among the 3 chiral local configurations [24]; (2) an approximation of solvent exposure per residue [25], measured by the number of vacant sites surrounding each occupied lattice site in a fold; and (3) pair contact energy via a Gō-like model for preferred contacts [20]. The Gō model and SS propensities were derived by simulating arbitrary chains and then arbitrarily selecting a chain representative of a good protein fold (high contact order, low solvent exposure, and low radius of gyration) [20, 26–29] as a reference model. We then assigned energies to the other chains in proportion to their distance from the reference. We quantified contact order as the average distance in sequence of two residues contacting in a fold. Contact order is a key statistic predicting folding rate, with high values indicating that the fold's nucleation requires distal regions to come into contact, making it dependent on meshing of long range side-chain forces in addition to local backbone propensities [28]. In the Gō model, chains with the same secondary structure motifs (short range potential) or the same pairwise long range contacts (not necessarily with the same neighbors) as the reference model have the lowest energy. Since the relative importance of the effects represented by the different energy functions is unknown a priori, we explored each energy function independently, arguing that if NSB works for each function, it will work for their combinations.

The choice of the temperature $T$ for the analysis is very important. Indeed, for $T \to 0$, the configurational distribution is dominated by a few highly probable configurations. The entropy is low, and hence it is easy to estimate, cf. Eq. (1). For $T \to \infty$, all configurations are equiprobable, and the Ma estimator is unbiased when $N \sim \sqrt{K}$, where $K$ is now the total number of configurations. Intermediate temperatures are the most inter-

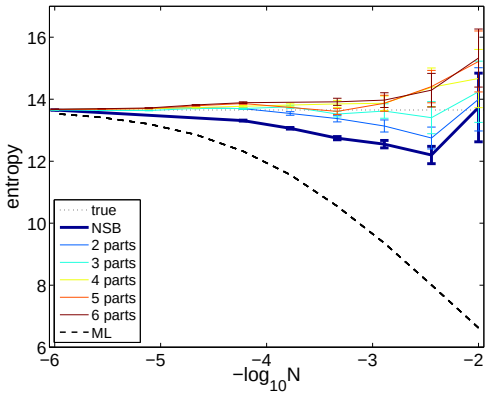


FIG. 1. Configurational entropy estimation for the first energy function and $T = 1$ a. u. NSB (thick line) and the grouping estimator with $M = 2 \ldots 6$ partitions are shown in comparison to $S_{\rm true}$ and the ML estimator. We sampled up to $N = 10^6 \approx 2^{20}$ from the Boltzmann distribution. Error bars correspond to one posterior standard deviation. For NSB and the grouping estimator, we bounded the number of configurations in each partition and in total as $\approx 2.74^L$. The bias of ML can be inferred and subtracted out when $\log_2 N \gtrsim S_{\rm true} \approx 14$, so that $N \gtrsim 10^4$, which corresponds to the bend in the ML curve [13]. However, both ML and NSB are biased for smaller $N$. In contrast, the grouping estimators is unbiased for $M = 3$ in the Ma regime of $\log_2 N \gtrsim S_{\rm true}/2 \approx 7$, or $N \gtrsim 10^2$.

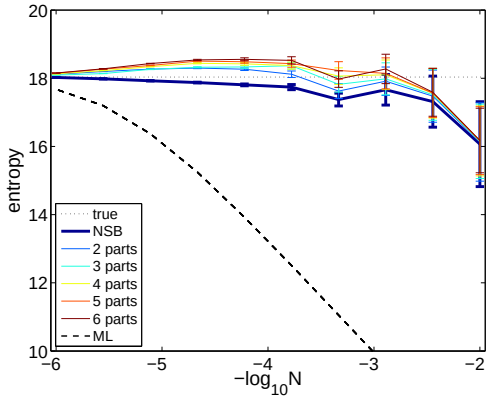


FIG. 2. Configurational entropy estimation for the first energy function and $T = 3$ a. u. Same conventions are used as in Fig. 1. Note that the bias is much less of a problem for this higher entropy case for NSB. As before, the grouping estimator can be made unbiased in the Ma regime, $\log_2 N \gtrsim S_{\rm true}/2 \approx 9$, or $N \gtrsim 10^3$.

esting. Here entropy is too high for simple methods to work, while the Ma estimator cannot be used because configurations are not equiprobable. For our chiral polymers, we estimate numerically $K \approx 2.74^L$, where 2.74 replaces 3 due to self-avoidance and finite volume. Thus, for $T \to \infty$, $S_{\rm true} \sim \log_2 2.74^L$, which is about 46 bits for $L = 32$. We are most interested in entropies substantially smaller than this, but much larger than 1 bit.

Typical results of applying NSB to samples from the protein configurational distribution for the first energy function are illustrated in Fig. 1 at an intermediate temperature $T = 1$ a. u., when $S_{\rm true} \approx 13.65$ bits. By the time $\log_2 N \sim S_{\rm true}/2 \approx 7$, many coincidences have occurred. The estimator is reporting small posterior variances, but it is biased, though always less than ML. NSB remains biased even when $N \sim 2^{S_{\rm true}}$. The bias finally disappears only when even the naive ML estimator is nearly unbiased, $N \gg 2^{S_{\rm true}}$, that is, many samples per typical configuration. Similar failures are observed for different sequence lengths and the other two energy functions. The bias likely stems from the assumptions of NSB being incompatible with the data.

As the temperature increases above $T = 3$ a. u. and the entropy grows beyond $S_{\rm true} \approx 18$ bit, the bias of $S_{\rm NSB}$ becomes small at $\log_2 N > S_{\rm true}/2 \approx 9$, see Fig. 2. While the bias is nonzero, it is comparable to the standard deviation, making the estimator useable. Thus long tails disappear from the distribution of configurations, and the estimator works at $< 40\%$ of the maximum possible entropy (46 bits) for this polymer! Since no general entropy estimator can work until $\log_2 N \gtrsim S_{\rm true}/2$, in this regime, NSB performs nearly optimally.

We can capitalize on the accurate performance of NSB for near-uniform distributions at high $T$. In the fish counting problem, basses, carps, and catfishes may have different probabilities of being caught, while the probabilities may be closer to uniform within the species. Thus counting each species *separately* will improve population estimates, but at the cost of needing a larger $N$ to ensure that coincidences (catching a tagged fish) happen for each species. Similarly, suppose the space of possible protein configurations is split into partitions $\nu_\mu$, $\mu = 1, \ldots, M$. Then by the grouping axiom for entropy, [30]

$$S(\mathbf{p}) = \sum_{\mu=1}^{M} \pi_\mu S(\nu_\mu) + S(\boldsymbol{\pi}), \tag{5}$$

where $S(\nu_\mu)$ stands for the entropy of the partition $\nu_\mu$, $\pi_\mu = \sum_{i \in \nu_\mu} p_i$ is the probability of a particular partition, and $S(\boldsymbol{\pi})$ is the entropy of the partition choice. While the overall $\mathbf{p}$ may be incompatible with NSB, the estimator may perform better on each of the partitions separately, resulting in the new *grouping* estimator [31]:

$$S_{\rm gr}(\mathbf{n}, M) = \sum_{\mu=1}^{M} \phi_\mu S_{\rm NSB}(\nu_\mu) + S_{\rm NSB}(\boldsymbol{\phi}), \tag{6}$$

$$\begin{aligned} \delta^2 S_{\rm gr}(\mathbf{n}, M) &= \sum_{\mu=1}^{M} \left[\delta^2 \phi_\mu S_{\rm NSB}(\nu_\mu) + \phi_\mu \delta^2 S_{\rm NSB}(\nu_\mu)\right] \\ &\quad + \delta^2 S_{\rm NSB}(\boldsymbol{\phi}) \\ &\approx \sum_{\mu=1}^{M} \phi_\mu \delta^2 S_{\rm NSB}(\nu_\mu) + \delta^2 S_{\rm NSB}(\boldsymbol{\phi}). \end{aligned} \tag{7}$$

Here $\phi_\mu = \sum_{i\in\nu_\mu} n_i$ are the empirical frequencies of each partition, and $\delta^2$ is the posterior variance. For $S_{\rm gr}(M)$ to be unbiased, the partitions should be chosen such that either (i) distributions of configurations within each partition are more uniform (allowing Ma's arguments to work), or (ii) structure of tails within each partition is compatible with NSB. If a non-NSB estimator is used in the r. h. s. of Eq. (6), partitions should be chosen instead to make that estimator unbiased.

Since each polymer configuration has an energy value that is known, and configurations with similar energies are nearly equiprobable, a natural partitioning exists in this context. We expect reduction in bias if one assigns a configuration with energy $E$ to the partition $\mu$, for which $E_{\rm min} + (E_{\rm max} - E_{\rm min})(\mu-1)/M \le E < E_{\rm min} + (E_{\rm max} - E_{\rm min})\mu/M$, where $E_{\rm min}$ and $E_{\rm max}$ are the minimum and the maximum energy in a given sample.

However, such partitioning comes at a cost. First, each of the terms in Eq. (5) has statistical errors. The errors add in quadratures, so that the estimator variance, $\langle\delta^2 S_{\rm gr}\rangle$, typically grows with $M$. Second, when $M \to K$, $S(\pi)$ approaches $S_{\rm true}$ and becomes equally hard to estimate. Third, for $M > 1$, one needs coincidences in *each* partition. This requires more data, and would lead to $\langle\delta S_{\rm gr}\rangle > 0$ if some of the partitions have no coincidences. Finally, one doesn't know the maximum possible number of configurations $K_\nu$ in each partition, and has to take $K_\nu = K$. Larger $K$ results in a larger $S_{\rm NSB}$, though the dependence is weak [17]. Thus $\langle\delta S_{\rm gr}\rangle > 0$ for $M \gg 1$. Combined, these concerns indicate that success of the grouping estimator in polymer problems is uncertain.

We tested the performance of the grouping estimator for different $L$, $T$, and energy functions. The results were consistent with the expectations and similar for all cases. As seen in Fig. 1, increasing the number of partitions first decreases the bias. $\langle\delta S_{\rm gr}\rangle$ is insignificant for $M \sim 2\ldots4$. For $M$ so small, each partition is sampled well, and $S_{\rm gr}$ works in the Ma regime. However, as $M$ grows, the bias changes sign and increases again. Similar results hold for higher temperatures, Fig. 2. Here the bias is small for all $M$, and it is dramatically smaller than the ML bias.

These results suggest a straight-forward algorithm for estimation of polymer configurational entropies. For a given sample of configurations and their energies, one computes $S_{\rm gr}(\alpha N, M)$ using Eq. (7) and $\delta^2 S_{\rm gr}(\alpha N, M)$ for $M = 1$ using Eq. (7), while varying the fraction of the data used for the estimation, $0 < \alpha \le 1$. One looks for the sample size dependent bias by verifying if $S_{\rm gr}(\alpha N, 1)$ drifts by more than the standard deviation as $\alpha$ increases. If the bias is positive, the algorithm cannot be applied (this has never happened in our tests). If the bias is insignificant, then $S_{\rm gr}(N,1) \pm \left(\delta^2 S_{\rm gr}(N,1)\right)^{1/2}$ is the entropy estimate. If the bias is negative, then one increments $M \to M+1$, and repeats the estimation for various $\alpha$. One increments $M$ until it reaches $M^*$, such that $\delta S_{\rm gr}(N, M^*) > 0$. The best estimate, the bias, and the variance are then the means of the corresponding quantities for $S_{\rm gr}(N, M^*-1)$ and $S_{\rm gr}(N, M^*)$. Crucially, unless $M^* \gg 1$, coincidences are present in all partitions. Thus the proposed estimator will work in the Ma regime, $\log_2 N \sim S_{\rm true}/2$, providing a *square root* data requirement reduction compared to simpler approaches. Since coincidences are required for any estimator to work, it is unlikely that other general purpose estimators will substantially outperform the NSB-based grouping algorithm.

For off-lattice polymers, the entropy can be computed by enumerating local minima in the energy landscape and additionally estimating the entropy within each such basin of attraction [32]. For the latter, there are good methods for entropy estimation based on kernel smoothing or nearest neighbor techniques [33, 34]; we expect that these also will be improved by grouping. Alternatively, the entropy in a local basin may be estimated analytically using the normal modes approximation [32]. We expect the present version of the NSB algorithm with grouping to be especially useful for the former, that is for calculating contributions to entropy from many similar local minima, which are observed for rugged energy landscapes that are characteristic of real proteins [35].

In summary, in this Letter, we have verified that the grouping generalization of the NSB algorithm can be used to produce reliable estimates of configurational entropies of polymer chains in the severely undersampled regime $N \sim \sqrt{2^{S_{\rm true}}}$, using only random samples of configurations and their corresponding energies. The estimator is available from `http://nsb-entropy.sourceforge.net` as a C++ and Matlab/Octave code. The estimation is rapid on modern computers. For lattice polymers of length $\sim 30$, it requires only $\sim 1000$ configuration samples. Thus the square-root scaling suggests that the method will be able to work with sequences of previously unaccessible lengths.

This work was partially supported by the US Department of Energy under the contract No. DE-AC52-06NA25396. IN was partially supported by the James S. McDonnell Foundation Award No. 220020321.

* ilya.nemenman@emory.edu
† mewall@lanl.gov
‡ cems@lanl.gov

[1] F. Ytreberg, R. Swendsen, and D. Zuckerman, J. Chem. Phys. **125**, 184114 (2006).
[2] K. Frederick, M. Marlow, K. Valentine, and A. Wand, Nature **448**, 325 (2007).
[3] M. Shirts, Methods in Molecular Biology **819**, 425 (2012).
[4] F. Ytreberg and D. Zuckerman, J. Chem. Phys. **124**, 104105 (2006).
[5] M. Head, J. Given, and M. Gilson, J. Phys. Chem. A **101**, 1609 (1997).
[6] R. White, J. Funt, and H. Meirovitch, Chem. Phys. Lett.

**410**, 430 (2005).
[7] U. Hensen, O. Lange, and H. Grubmüller, PLoS One **5**, e9179 (2010).
[8] A. Fairhall, E. Shea-Brown, and A. Barreiro, Curr. Opinion Neurobiol. **22**, 4 (2012).
[9] A. Levchenko and I. Nemenman, Curr. Opinion Biotech. **28**, 156 (2014).
[10] A. Antos and I. Kontoyiannis, Random Struct. Alg. **19**, 163 (2001).
[11] L. Paninski, Neural Comput. **15**, 1191 (2003).
[12] Z. Zhang, Neural Comput. **24**, 1368 (2012).
[13] S. Strong, R. Koberle, R. de Ruyter van Steveninck, and W. Bialek, Phys. Rev. Lett. **80**, 197 (1998).
[14] I. Nemenman, F. Shafee, and W. Bialek, in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 14, edited by T. Dietterich, S. Becker, and Z. Gharamani (MIT Press, 2002).
[15] S. Ma, J. Stat. Phys. **26**, 221 (1981).
[16] I. Nemenman, Entropy **13**, 2013 (2011).
[17] I. Nemenman, W. Bialek, and R. de Ruyter van Steveninck, Phys. Rev. E **69**, 056111 (2004).
[18] I. Nemenman, G. Lewen, W. Bialek, and R. de Ruyter van Steveninck, PLoS Comput. Biol. **4**, e1000025 (2008).
[19] M. Levitt and A. Warshel, Nature **253**, 694 (1975).
[20] N. Go and H. Taketomi, Proc. Natl. Acad. Sci. (USA) **75**, 559 (1978).
[21] D. Hinds and M. Levitt, Proc. Natl. Acad. Sci. (USA) **89**, 2536 (1992).
[22] J. Karanicolas and C. Brooks, Protein Sci. **11**, 2351 (2002).
[23] D. Gront, A. Kolinski, and J. Skolnick, J Chem. Phys. **113**, 5065 (2000).
[24] P. Pokarowski, K. Droste, and A. Kolinski, J Chem. Phys. **122**, 214915 (2005).
[25] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).
[26] K. Lau and K. Dill, Macromolecules **22**, 3986 (1989).
[27] K. Dill, Biochemistry **29**, 7133 (1990).
[28] K. Plaxco, K. Simons, and D. Baker, J Mol. Biol. **277**, 985 (1998).
[29] C. Rohl, C. Strauss, K. Misura, and D. Baker, Methods Enzymol. **383**, 66 (2004).
[30] C. Shannon and W. Weaver, *The mathematical theory of communication* (University of Illinois Press, Urbana, IL, 1962).
[31] The grouping property, Eq. (5), was first used to estimate entropy in neuroscience context [18, 36]. However, the optimal choice of the number of partitions was not addressed there.
[32] D. Ming, M. Anghel, and M. Wall, Phys. Rev. E **77**, 021902 (2008).
[33] J. Beirlant, E. Dudewicz, L. Gyorfi, and E. Van der Muelen, Int. J. Math. Stat. Sci. **6**, 17 (1997).
[34] A. Kraskov, H. Stogbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[35] H. Frauenfelder, S. Sligar, and P. Wolynes, Science **254**, 1598 (1991).
[36] C. Tang, D. Chehayeb, K. Srivastava, I. Nemenman, and S. Sober, PLoS Biol. **12**, e1002018 (2014).