

Understanding Protein Stability And Flexibility Using Network Rigidity

D. J. Jacobs

Physics and Astronomy, California State University, Northridge 18111 Nordhoff St.
Northridge CA 91330-8268 U.S.A. donald.jacobs@csun.edu

ABSTRACT

Protein structure is represented as a mechanical framework defined by constraint topology. Various constraint types are used to model covalent bonds, hydrogen bonds, salt bridges and torsion angles. The set of constraint types defines a decomposition of free energy where each constraint is quantified by a local molecular free energy function, having enthalpy and entropy contributions. A coarse grain statistical mechanical Distance Constraint Model (DCM) is presented that explicitly accounts for network rigidity among constraints. For each constraint topology, the well-known problem of non-additivity of component entropies from a free-energy decomposition is explicitly accounted for using network rigidity. Network rigidity is a long-range underlying mechanical interaction that provides a mechanism for enthalpy-entropy compensation. For fixed constraint topology, an efficient graph algorithm is employed to identify flexible and rigid regions, and to determine independent constraints. Total enthalpy is obtained by summing over all enthalpy contributions. Total conformational entropy is obtained by summing entropy contributions from preferentially selected independent constraints to provide a lowest upper bound estimate. Ensemble averaging over accessible constraint topologies allows thermodynamic stability and molecular cooperativity (flexibility) to be predicted. The DCM provides a novel modeling scheme that probes stability-flexibility relationships important for protein engineering. A hybrid computational method that combines Monte Carlo sampling and Landau theory is employed. Phenomenological DCM parameters are determined by fitting to measured heat capacity data. Flexibility characteristics, stability curves and free energy landscapes are calculated. The DCM runs 10 orders of magnitude faster than Molecular Dynamics to obtain comparable sampling statistics.

1. Introduction

Since the late 1950s with the introduction of phenomenological Ising-like models to explain the alpha-helix to coil transition, there has been a desire for *reduced protein models* to capture the essential physics of protein folding. Also important, are many related issues pertaining to protein stability and molecular cooperativity that need to be accurately modeled so that function efficacy can be engineered by computer aided design in fast turn-around times. To this end, it would be desirable to have a simple model that will balance accurate estimations of energy and entropy among molecular constituents, while representing solvation effects in terms of enthalpy and entropy component contributions. A common approach that can satisfy the *simple* and *fast* criteria has been to invoke free energy decomposition schemes. Unfortunately, models based on the additivity assumption of free energy components become inaccurate when component

parts within a protein or associated with protein-solvent interactions strongly couple. These long-standing problems are directly addressed using a *Distance Constraint Model* (DCM) [1] that provides a computationally tractable approximation scheme. The DCM is based on free energy decomposition, but it invokes network rigidity as an underlying mechanical interaction to accurately account for non-additivity in component entropies.

2. Method

Microscopic interactions are modeled as constraints that may be quenched or fluctuate. Generally, a constraint, t , consists of m_t (more than one) elementary distance constraints. The DCM obtains thermodynamic properties by working with an ensemble of distinct constraint topologies, each defining a mechanical framework, \mathcal{F} . The DCM is solved within Landau theory by defining a macrostate of a protein in terms of (N_{hb}, N_{nt}) giving the number of (crosslinking H-bonds, native-like torsion-angles). Moreover, these two order parameters define a two-dimensional grid in constraint space. For each node on the grid, a Landau free energy is written in terms of phenomenological parameters as:

$$G(N_{hb}, N_{nt}) = \sum_t E_t \overline{\eta}_t - u N_{hb} + v N_{nt} - RT \sum_t [\gamma_t \overline{\sigma}_t - [p_t \ln(p_t) + (1-p_t) \ln(1-p_t)]] \quad (1)$$

where constraint type, t , has an energy, E_t , and maximal *local* conformational entropy, γ_t . Energy of H-bonding to solvent is accounted for by u , ($u < 0$ is favorable), and v being negative favors native-like torsion constraints. An Ising-like variable η_t is (0,1) when a fluctuating constraint (is not, is) present. Variable σ_t gives the number of independent elementary distance constraints, ranging between 0 to m_t requiring non-trivial calculation. Distance constraints within framework \mathcal{F} are sorted from smallest γ_t to largest $\forall t$. A test for independence employs a generic rigidity graph-algorithm [2], except here, the order of constraint placement is defined by the sorted list. This preferential ordering provides a lowest upper bound estimate for conformational entropy, where $S_c(\mathcal{F}) \leq R \sum_t \gamma_t \sigma_t$. Mixing entropy, $S_m(\mathcal{F})$, is accounted for in Eqn. (1) by the last two terms where p_t has the form of a Fermi-Dirac occupation probability because each constraint acts as a two level system.

The known native 3D protein structure is used as a starting template. An ensemble of frameworks is constructed by perturbing away from this template. This simplification allows us to consider the variables E_t and γ_t fixed. Using empirical formulae for E_t , and assuming γ_t is a linear function of E_t , a 5-parameter model has been developed --- two are modeled as independent of protein and solvent conditions, and 3 remain as effective phenomenological parameters to be determined. With these simplifications, $\{p_t\}$ provides mean-field probabilities that are self-consistently calculated using Lagrange multipliers to fix the desired total number of constraints on the protein conformation per node. Random constraint topologies are then generated using Monte Carlo sampling within each node in accordance with a product probability measure built from $\{p_t\}$ to calculate Eqn. (1). By adaptive grid sampling, a free energy landscape is calculated from which thermodynamic response functions follow. Three free-parameters are determined by fitting heat capacity predictions to experimental data using simulated annealing with LAM-MPI parallel code.

3. Results

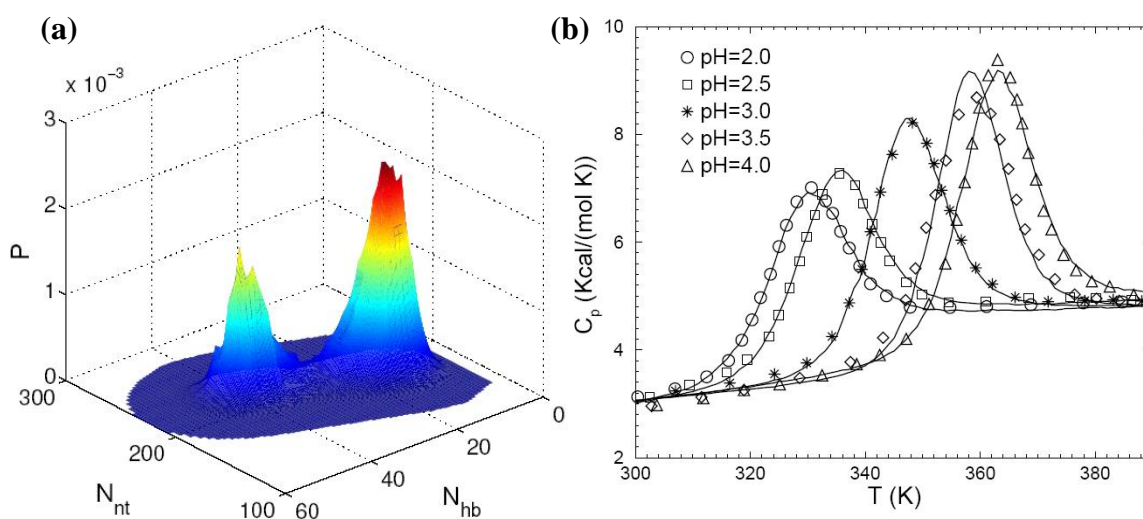


Figure 1: (a) Probability for the protein to have N_{hb} H-bond crosslinks and N_{nt} native-like torsion constraints for ubiquitin (at pH 3.0). (b) Corresponding 3 parameter DCM fits to the heat capacity data taken by P.L. Wintrode et al, *Proteins* **18** 246-253 (1994).

Typical free energy landscapes show two deep basins associated with the native and unfolded states (see Fig. 1a). Considering phenomenological parameters as functions of pH, the essential features of heat capacity are reproduced well, with typical results shown in Fig. 1b. Further, the DCM directly relates protein stability and flexibility in 10^{10} faster CPU time than MD to obtain similar sampling statistics. Current performance scales as $O(N^2)$, taking 2 hrs of CPU time on one 2.4 GHz processor for a $N=238$ residue protein. The goal is to obtain transferable parameters based on local molecular partition functions.

Acknowledgements

Research supported by NIH S06 GM48680-0952 and Research Corporation CC5141.

References

- [1] D.J. Jacobs, S. Dallakyan, G.G. Wood, and A. Heckathorne, *Network rigidity at finite temperature: relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems*, *Phys. Rev. E* **68** 061109-21 (2003); D.J. Jacobs, and G.G. Wood, *Understanding the alpha-helix to coil transition in polypeptides using network rigidity: predicting heat and cold denaturation in mixed solvent conditions*, *Biopolymers* **75** 1-31 (2004).
- [2] D. J. Jacobs, A. J. Rader, L. A. Kuhn, and M. E. Thorpe, *Protein flexibility predictions using graph theory*, *Proteins* **44** 150-165 (2001).