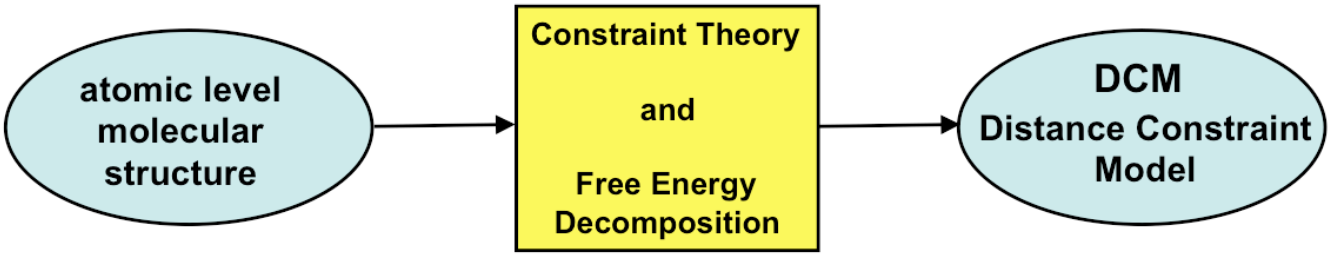
A NEW PERSPECTIVE

D.J. Jacobs,et. al., *Network rigidity at finite temperature: Relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems*. **Physical Reviews E. 68, 061109 1-21 (2003)**

The DCM resolves the problem of non-additivity

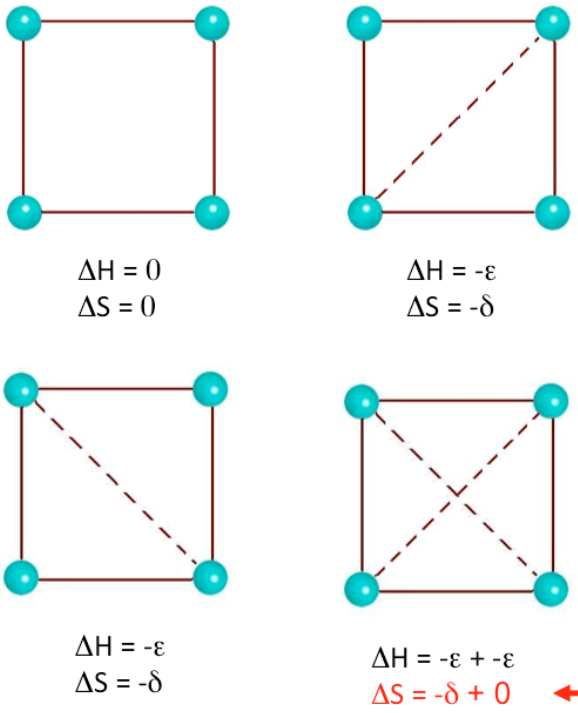
Network rigidity is an enthalpy-entropy compensation mechanism that derives from long-range mechanical interactions.



"*I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it*" ! **--- Lord Kelvin**

Taking the advice of Lord Kelvin, the Father of Thermodynamics, I describe the protein molecule and other molecular systems of interest in terms of a mechanical model. Thermodynamic information is impressed onto mechanical networks to form a complete statistical mechanics theory. Network rigidity becomes a long-range interaction – the critical aspect where I found simplicity in complexity.

Key idea: Combine constraint theory (network rigidity) and free energy decomposition to account for correlations between all degrees of freedom, enabling the required non-additive reconstitution of free energy components within a system.



Jacobs, et al. *Proteins* (2001) 44:150
Jacobs, et al. *Phys. Rev. E* (2003) 68:061109
Jacobs & Dallakyan (2005) *Biophysical J.* 88:903

$$G(F) = H(F) - TS(F)$$

$$H(F) = \sum_c h_c p_c(F)$$

$$S(F) = \sum_c s_c q_c(F) p_c(F)$$

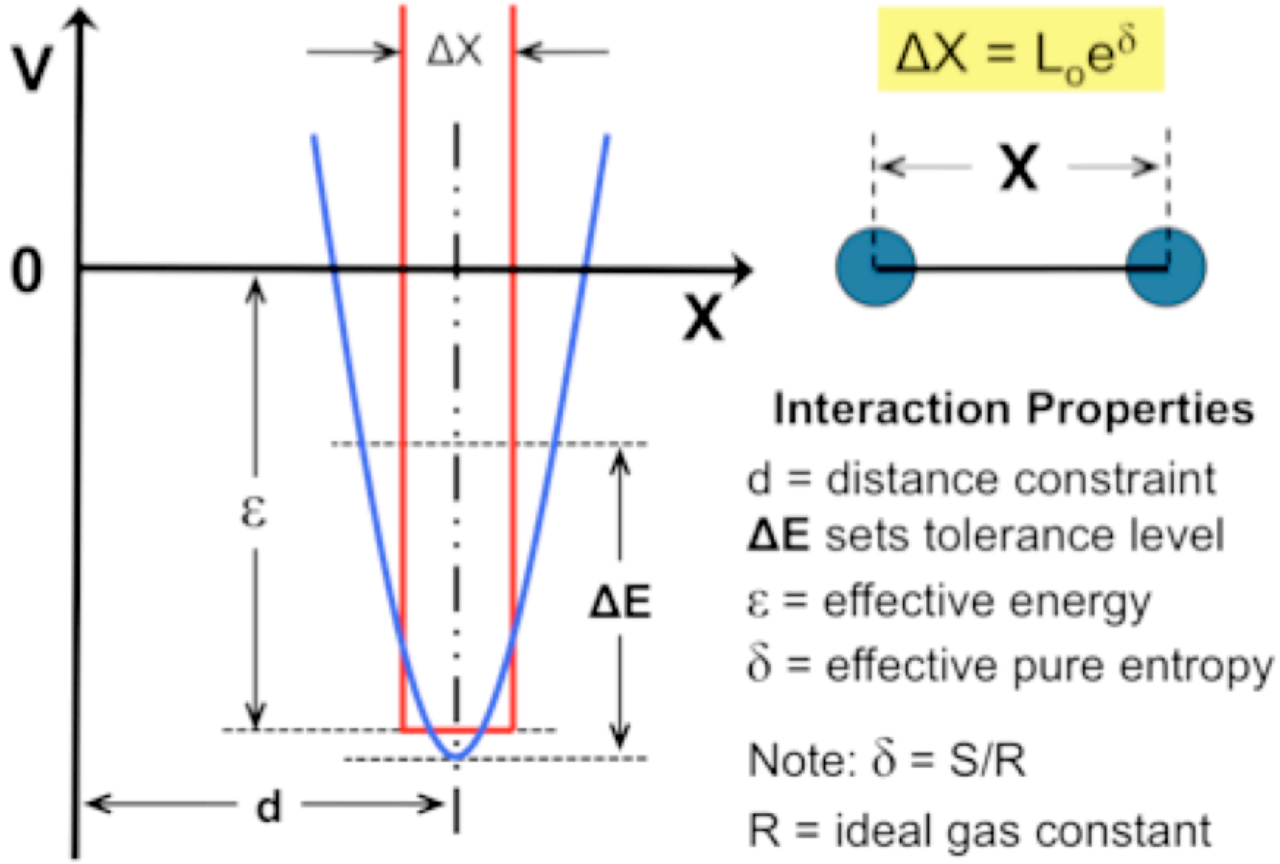
Regarding NETWORK RIGIDITY as a mechanical interaction accounts for NON-ADDITIVITY IN ENTROPY

Focusing on why **mechanics matters**, consider a two dimensional quadrilateral molecule with four distinct constraint topologies (or frameworks). In the top-left network, the shape is very flexible. In the right-top and bottom-left networks, the diagonal constraint (i.e. a fluctuating hydrogen bond) makes the networks isostatically rigid. In the bottom-right network, two diagonal constraints cause the network to be overconstrained (i.e. having one redundant constraint). Since interactions are modeled as distance constraints, each will have an associated entropy contribution that reflects the amount of (change in length) that is allowed between a given pair of atoms. **See next slide if you are confused about how a distance constraint is associated with a change in length!**

As distance constraints are added to a network, motions are becoming more limited provided the constraints are independent. However, if a region is already rigid, adding a distance constraint will not lower entropy. As such, the lower-right network does not lower entropy due to redundant constraints, which is where non-additivity enters. The entropy is related to the number of independent constraints in a system, not simply total number of constraints. Thus, regions having redundant constraints form cooperative mechanisms during the process of forming/removing specific interactions in a system.

The free energy is broken up into separate enthalpy and entropy terms. According to the formulas, it is clear that the total enthalpy is obtained additively (as is generally assumed) by weighting variable enthalpy contributions by the probability, p_c, that a fluctuating constraint is present in the network. In contrast, entropy terms are non-additive due to the **additional attenuation factor, q_c**, corresponding to the conditional probability that if a constraint is present in a network it must also be independent for it to contribute to the entropy. In the lower-right network, due to symmetry, each diagonal constraint is assigned a q-factor of ½ since there are twice as many constraints than is needed to maintain the region to be isostaticly rigid. Note that each framework, F, represents a different constraint topology, and has its own calculation based on network rigidity.

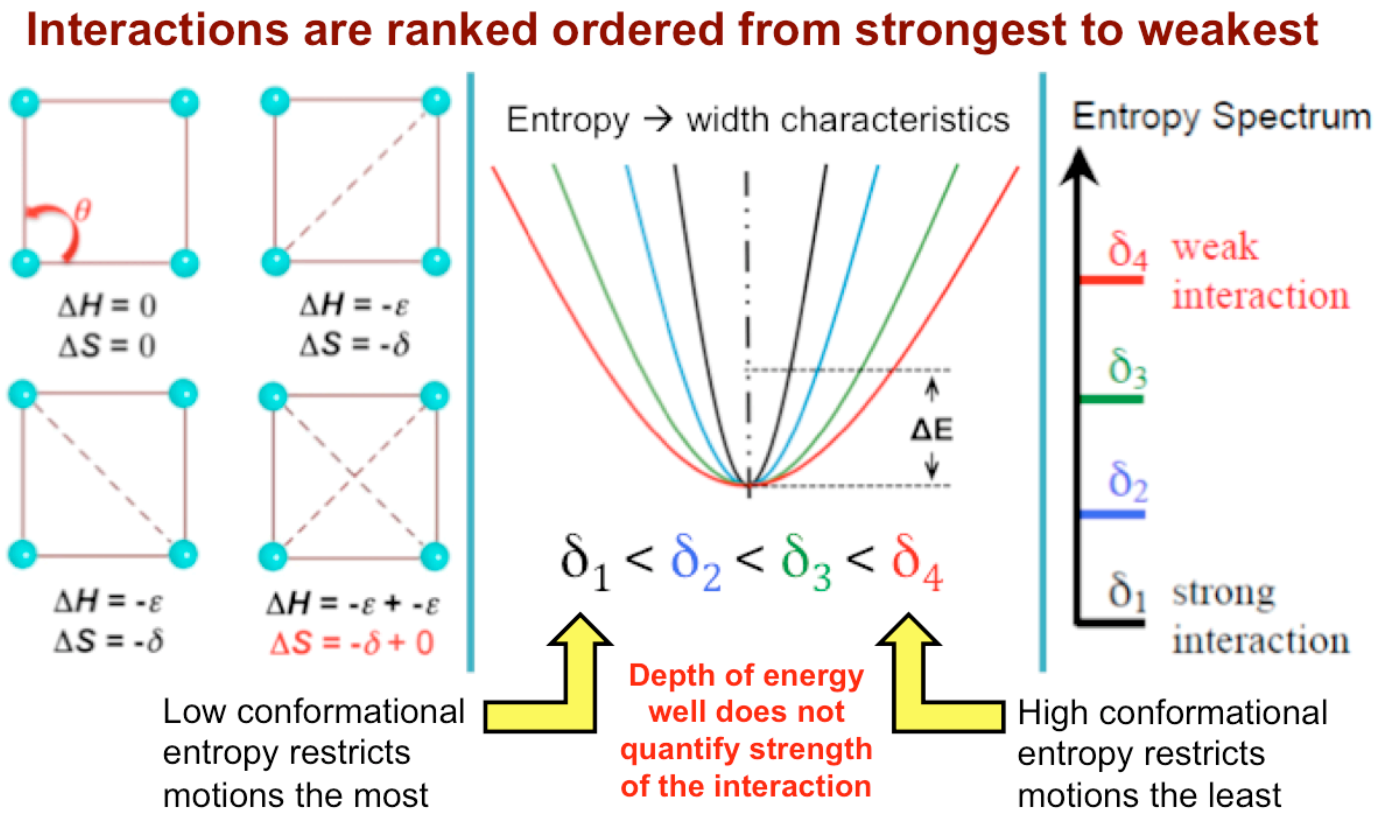
Key idea: Non-additivity is directly linked to network rigidity (mechanical properties), and in particular non-additivity appears because of the presence of redundant constraints.



This schematic picture shows how a distance constraint has variable length. When network rigidity is calculated, it is based on **generic rigidity**. This means the exact length of the constraints do not matter (within linear response). The rigidity properties of a network are dependent only on **topology**, defined by the relative placement of the distance constraints, but not their precise lengths. As such, to know what is rigid and what is flexible does not require knowing the exact lengths of distance constraints. Indeed, there will be fluctuations in the length between any pair of atoms. We can characterize an interaction by the depth of the potential well and its width, which is determined by a set tolerance in the energy, given by ΔE. The greater curvature corresponds to smaller entropy. That is, the flatter the potential well, the greater entropy is associated with the distance constraint.

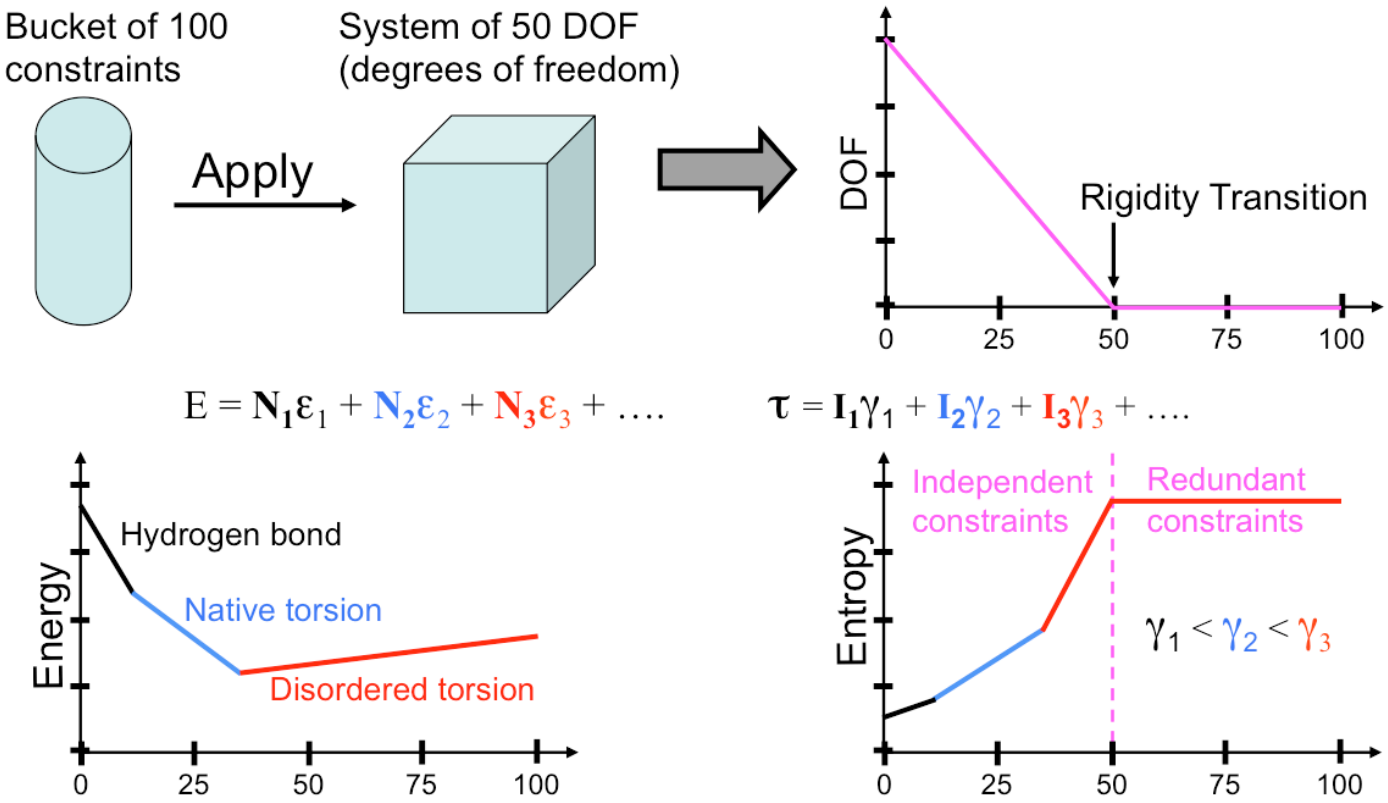
By selecting some arbitrary global scale factor for distance, L_o, it is possible to define pure entropies, δ, associated with the tolerance in length assigned to each distance constraint. The smaller tolerance corresponds to stronger interactions having low entropy. Greater tolerance has a flatten curvature, corresponding to weak interactions and large entropy. In chemical bonds, high curvature corresponds normally to low depth in energy and vice versa, although this relationship is not technically required.

Key idea: Apply network rigidity in the sense of generic rigidity, and by accounting for tolerances on distance constraints (i.e. the wiggling room allowed by a generic distance constraint in the network), an estimate for the total conformational entropy of the system can be made.



If you understand everything said on the previous slide, it should be clear to you that a system having inhomogeneous constraint types will create different entropy contributions due to differences in curvature in the potential energy function associated with the interaction that the distance constraint is modeling. All constraints (corresponding to some type of interaction) are assigned entropy values based on curvature of the potential energy function. Redundant constraints do not contribute to conformational entropy, but independent constraints do. Also important to note is that the assignment of which distance constraint is independent and which is redundant is arbitrary to a large degree. It depends on the order of assignment to which constraint is independent and which is redundant (see FIRST which describes the pebble game). As such, **a unique estimate for conformational entropy is not possible because it depends on which constraints are assigned to be independent. Moreover, not all independent constraints are orthogonal**. As such, the non-additive contributions will not yield a unique or correct answer by simply adding up independent constraints. Instead, each answer provides an upper bound estimate to the true conformational entropy. The distance constraint model gives the lowest possible upper bound estimate for conformational entropy by preferentially placing distance constraints in a network with lower entropy before those with higher entropy to obtain a robust and surprisingly good estimate of total conformational entropy – without moving any atom! A **preferential rule** is implemented by presorting all the constraints from lowest to highest entropies in a queue for identifying if a constraint is independent during the pebble game. **The queue defines an entropy spectrum** as depicted on the right panel.

Key idea: The lowest upper bound estimate of conformational entropy is obtained by systematically placing distance constraints with the lowest pure entropy in the network before those with higher pure entropies.



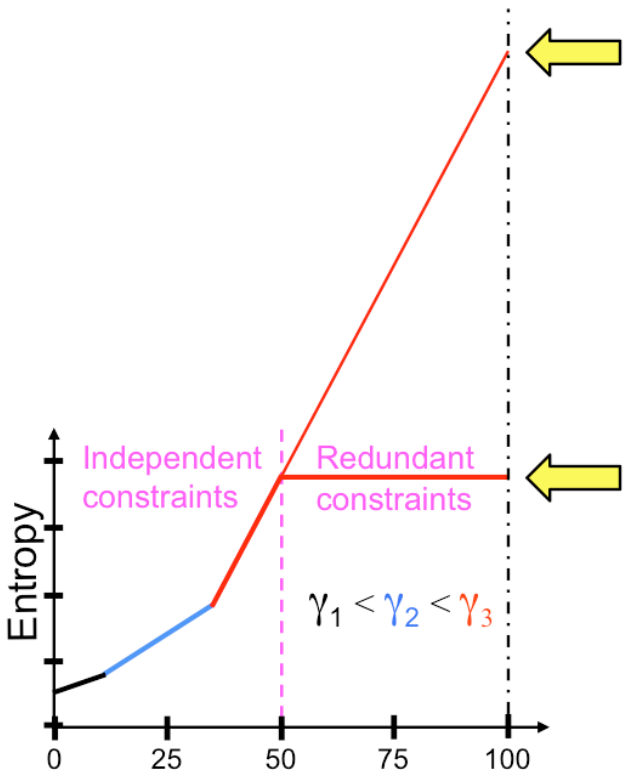
D. Jacobs and M. Fairchild, *Progress in Biopolymer Research,* pp 45-76
Editor: Pablo C. Sanchez, ISBN: 1-60021-984-5 45-76 (2007).

An instructive simple example: Suppose for the moment that Maxwell-counting is exact. Maxwell, (the famous theoretical physicist that worked out the equations for electromagnetic radiation) worked on the mechanics of networks and provided a quick an dirty approximate estimate for finding the number of distance constraints that will be redundant in a homogenous network of constraints.

The steps of the DCM are to place constraints in the network according to the preferential rule that places constraints in order of lowest entropy to highest entropy. In this example, as is commonly the case, assume the constraints with lowest energy also have the lowest entropy. Also, assume there are three types of constraints, each characterized by an energy and entropy contribution (ε,γ). Based on Maxwell counting, as the constraints are placed in the system they are independent until there are just enough constraints to make the entire network globally rigid. At that point there are no internal degrees of freedom remaining in the system, and this point defines a **rigidity threshold** (see top-right figure). The system undergoes a **rigidity transition** from being globally flexible to globally rigid across the rigidity threshold, denoted by the vertical dashed line in the bottom-right figure.

Total energy is just the sum of the energy contributions from all the distance constraints (that model the interactions) as shown in the bottom-left figure. Note that the same net energy would be obtained independent of the order the constraints are placed, because as you know, the result you get in adding up a list of numbers do not depend on the order you add them.

Total entropy is assumed to equal the lowest upper bound estimate, which is obtained by using the preferential rule and placing the lowest entropy constraints before the higher entropy constraints. Once the rigidity transition point is reached, all constraints placed thereafter are redundant. Before that, all constraints are independent. This is why the total for the entropy does not accumulate after the rigidity transition. In other words, q=1 and 0 below and above the rigidity threshold respectively.



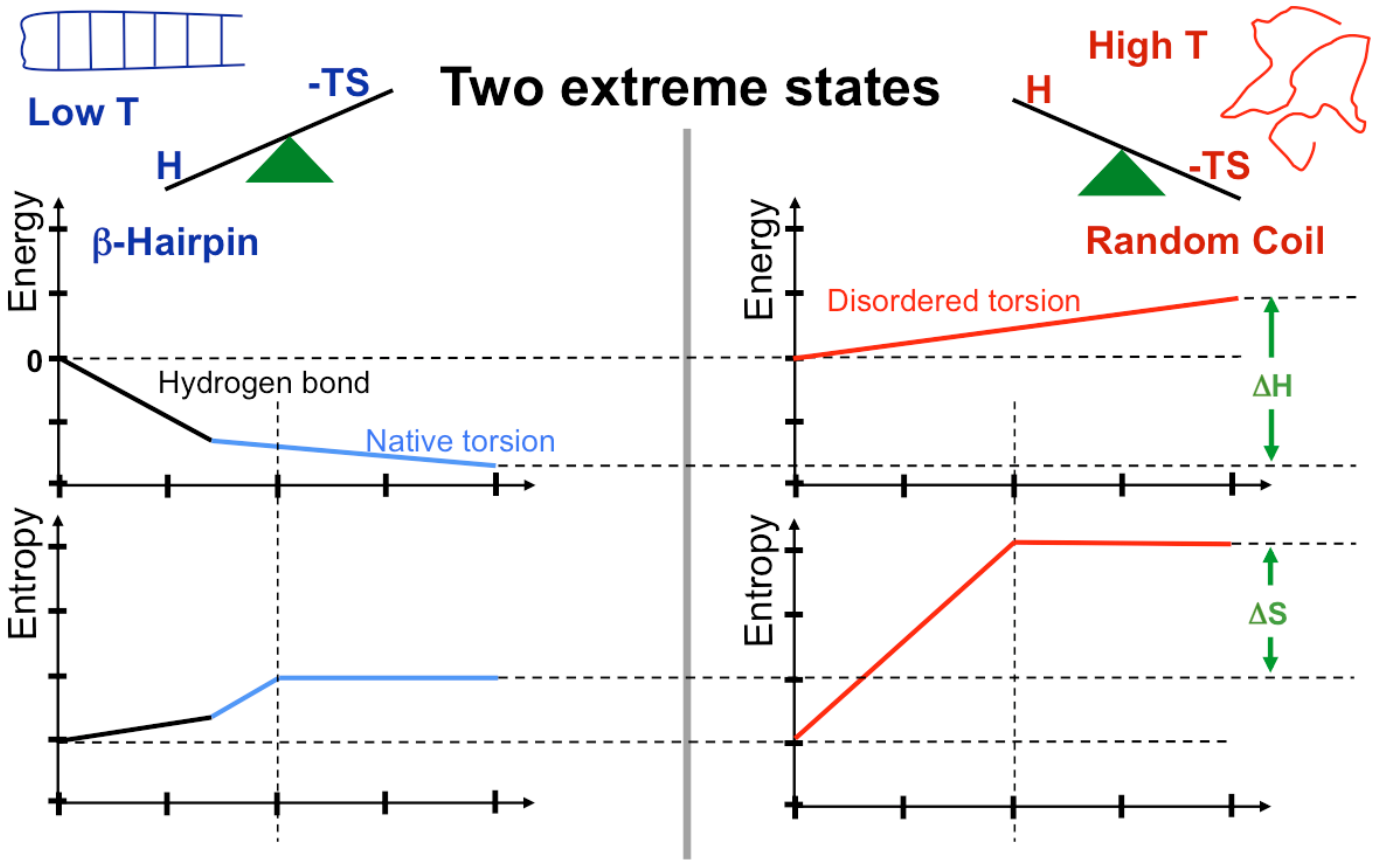
Result from an additive model:
If **free energy reconstitution** is performed by simply adding the entropy contributions from all constraints, one gets a really bad overestimate of the true conformational entropy. This is why all linear models fail in the general case, and why free energy decomposition schemes are not used.

Result from a non-additive model:
If **free energy reconstitution** is performed by invoking the preferential rule of placing constraints using network rigidity, a lowest upper bound of entropy is obtained, which appears to provide a very good estimate!

D. Jacobs and M. Fairchild, *Progress in Biopolymer Research,* pp 45-76
Editor: Pablo C. Sanchez, ISBN: 1-60021-984-5 45-76 (2007).

An instructive simple example (continued): If rigidity theory were not invoked (in this case using Maxwell-counting as a fast and dirty approximation, instead of exact pebble game calculations), we would resort back to an additive model. Ising-like models that do not include long-range mechanical interactions using network theory will make errors in overestimating the entropy as the figure above demonstrates. By simply adding up all the entropy components from all independent constraints, one cannot account for correlations between degrees of freedom, thereby tremendously over estimating the conformational entropy. A common way to correct for these over estimations is to rescale the **local entropy parameters**, but the problem is that the actual mechanical interactions are missing. Consequently, Ising-like models that do not accounting for the fundamental mechanical interactions will miserably fail to be an accurate predictive model. There is some wiggling room depending on what the variables are, so several Ising-like models, such as the Zimm-Bragg and Lifson-Roig models have limited utility. Nevertheless, even in those models, they do not have transferable parameters in large part because non-additivity is important to take into account as I have shown in a number of publications on the helix-coil transition. In fact, application of rigidity theory does not depend on the specific details of the molecular system. The next example that I show considers a **beta-hairpin turn**, transitioning into a coil state, and back. In the next example, I revisit Shellman's two-state model for energy-entropy compensation, but I cast the problem in terms of the DCM to illustrate how it works!

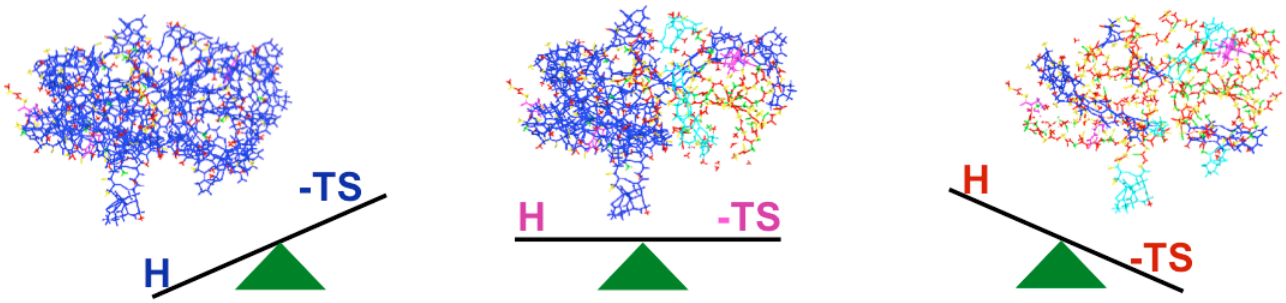
Key idea: Including **mechanical correlations** between degrees of freedom using a DCM provides a **pragmatic approximation** to restore the utility of employing **free energy decomposition** because non-additivity in conformational entropy is taken care of in the process of **free energy reconstitution**.



Example of a beta hairpin turn to coil transition: Using the **Maxwell-counting approximation** in this example, the calculations required to estimate energy and entropy can be done graphically as shown on the top and bottom graphs respectively. The two graphs on the left side correspond to an intact beta-hairpin turn with all the hydrogen bond crosslinks formed, while the two graphs on the right side correspond to an unfolded coil state. **The rigidity calculation is applied to each mechanical network separately**. In this case, only two states are considered, and by making a direct comparison of energy and entropy of both mechanical networks (representing the folded and unfolded states) the change in enthalpy and change in entropy of the system can be figured out in a similar way Shellman estimated changes for the alpha-helix to coil transition. Constraints placed in the system beyond the rigidity transition (denoted by the vertical dashed lines) are redundant and do not contribute to the entropy. The propagation of rigidity through a molecular system causes there to be non-additivity in the fundamental nature of conformational entropy. **The two-state model is a crude approximation**.

More generally for a given molecular system, such as a protein, an estimate for the energy (using an additive summation) and conformational entropy (using the non-additive preferential rule to network rigidity calculations of some type) must be performed on **every** mechanical framework. Furthermore, to fully account for microscopic fluctuations, all accessible constraint topologies must be considered by building a complete ensemble of mechanical frameworks. Just two such mechanical frameworks are considered above, representing two extremes.

Key idea: Apply network rigidity calculations combined with the free energy decomposition scheme to each accessible mechanical framework, and calculate the partition function of the system.



Network of constraints defines microstate of the system

Let $\gamma_t \equiv S_t/R$ Provides a dimensionless pure entropy

Suppose: $\gamma_1 < \gamma_2 < \gamma_3 < \ldots$ Ordering of entropies is important

$$E = N_1\varepsilon_1 + N_2\varepsilon_2 + N_3\varepsilon_3 + \ldots. \qquad \tau = I_1\gamma_1 + I_2\gamma_2 + I_3\gamma_3 + \ldots.$$

$$Q = \sum_F e^{\tau(F)} e^{-\beta E(F)}$$

$e^{-\beta E(F)}$ → Total energy at fixed constraint topology

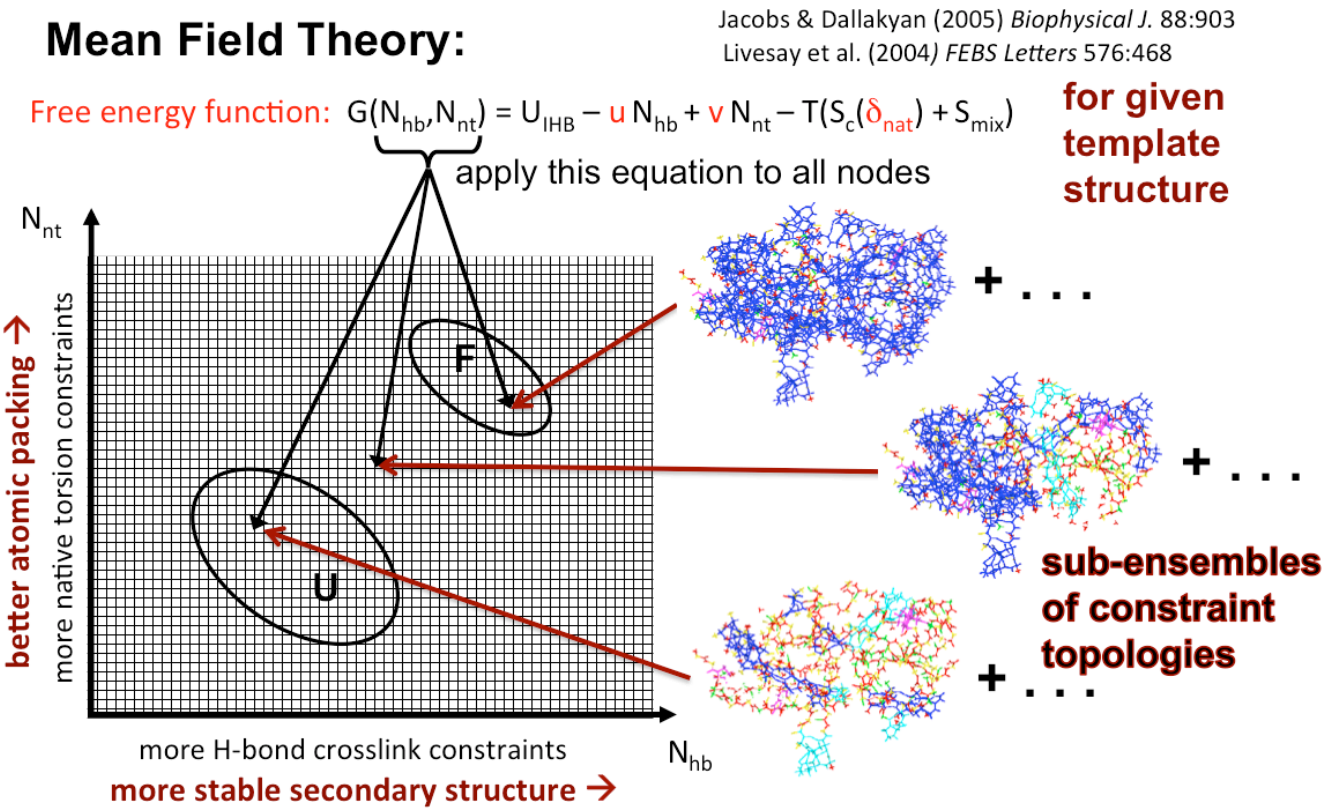
$e^{\tau(F)}$ → Total pure entropy. Conformational degeneracy = e^{τ}

$\sum_F$ → Sum over all topologically distinct mechanical frameworks

Jacobs, et. al., Phys. Rev. E 68, 061109 (2003).

The above figure summarizes the statistical mechanics of the DCM. Different mechanical frameworks are used to define **microstates**, which are not configurations. A configuration is a specific location of all atoms in the network. A mechanical framework defines generic distance constraints between pairs of atoms. Since each generic distance constraint is impressed with a length tolerance to give it some wiggle room, a mechanical framework represents a **mini-ensemble of accessible geometries** of the atoms consistent with a fixed constraint topology (i.e. **rigid cluster decomposition**). Because the tolerances in the distance constraints are set by an energy tolerance, each mechanical framework has an energy associated with it that is effectively a constant for all of its accessible geometries. A partition function is built by summing over Boltzmann factors, where the total energy is additive over all the constraints present in the system. However, flexible frameworks wiggle more than rigid ones, so that the amount of wiggling contributes to the statistical weight in the form of a **degeneracy factor**. The degeneracy factor for each mechanical framework is estimated through the conformational entropy calculation. When these two calculations are combined per framework, the partition function is expressed by the equation given above. For a protein, there will be an astronomical number of accessible frameworks. For example, if there are 400 possible hydrogen bonds (formed/broken) and 1,200 torsion interactions (native/disordered), then the number of distinct frameworks (or microstates) will be $2^{400} \times 2^{1200}$ that must be considered to calculate the partition function.

Key idea: The DCM is a well-defined statistical mechanical model that defines a particular partition function that must be calculated, and the standard tools available to calculate partition functions can be applied in the usual way.



The DCM in its general formulation has very little limitations. Although it is based on a coarse grained approach where an energy tolerance needs to be set, in the limit that this energy tolerance is taken to zero, the approximations essentially cease to exist. However, this would be useless because one would end up with the configuration integrals that cannot be integrated. A good view of the DCM is that it coarse grains the configuration integrals into discrete entities so that they can be summed over instead of integrated over. However, as the previous slide illustrated, the number of terms to be summed is astronomical. Compared to a two-dimensional Ising model on a square lattice of size 40×40, there is just as many microstates (i.e. 2^{1600}). This system is small compared to systems usually considered in condensed matter physics. However, **unlike problems in condensed matter physics, the protein system is highly non-homogeneous with strong finite size effects where interfacial boundary interactions with solvent need to be accounted for**.

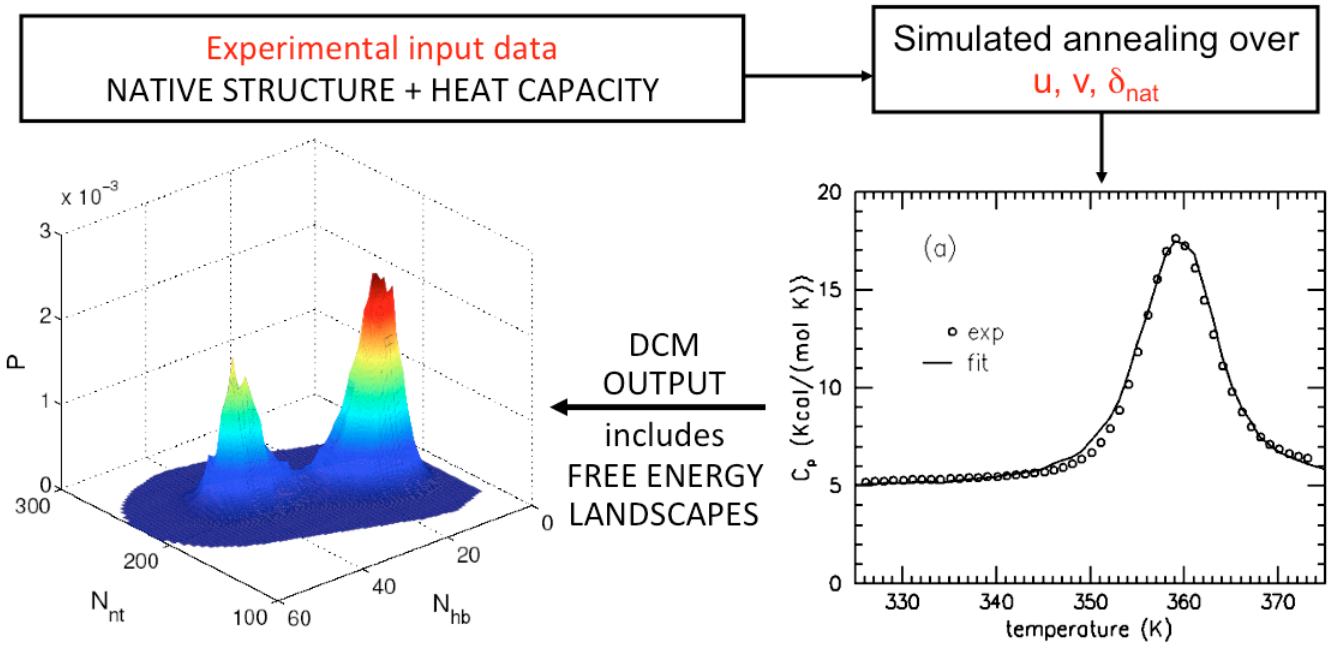
It is possible to calculate a partition function by transforming the problem into solving a **free energy functional** (FEF). In homogeneous systems, solving a FEF is relatively easy because translational symmetry can be used to simplify finding the solution. In a protein, solving the FEF is difficult because of the heterogeneous microenvironments. Approximations are made by identifying relevant order parameters to describe key features of the system. Rather than solving the FEF analytically, it is calculated numerically. Finding the minimum is not the goal (as it often is in condensed matter physics). In a protein, the free energy for different values of the order parameters that define the FEF are also of interest. The complete solution defines the **free energy landscape** (FEL).

minimal DCM (mDCM): A few empirical parameters are applied to all residues irrespective of type and location, which is why the mDCM is a mean field approximation (MFA). Also, the FEF is solved under a MFA. However, much information is retained despite making two MFAs because the mDCM is applied to a known protein structure, which encodes the hydrogen bond network. Exact rigidity calculations are performed using Monte Carlo sampling at each macrostate (defined by the grid points shown in the figure above). Empirical parameters account for much of the oversimplifications.

Mean Field Theory:

Jacobs & Dallakyan (2005) *Biophysical J.* 88:903
Livesay et al. (2004) *FEBS Letters* 576:468

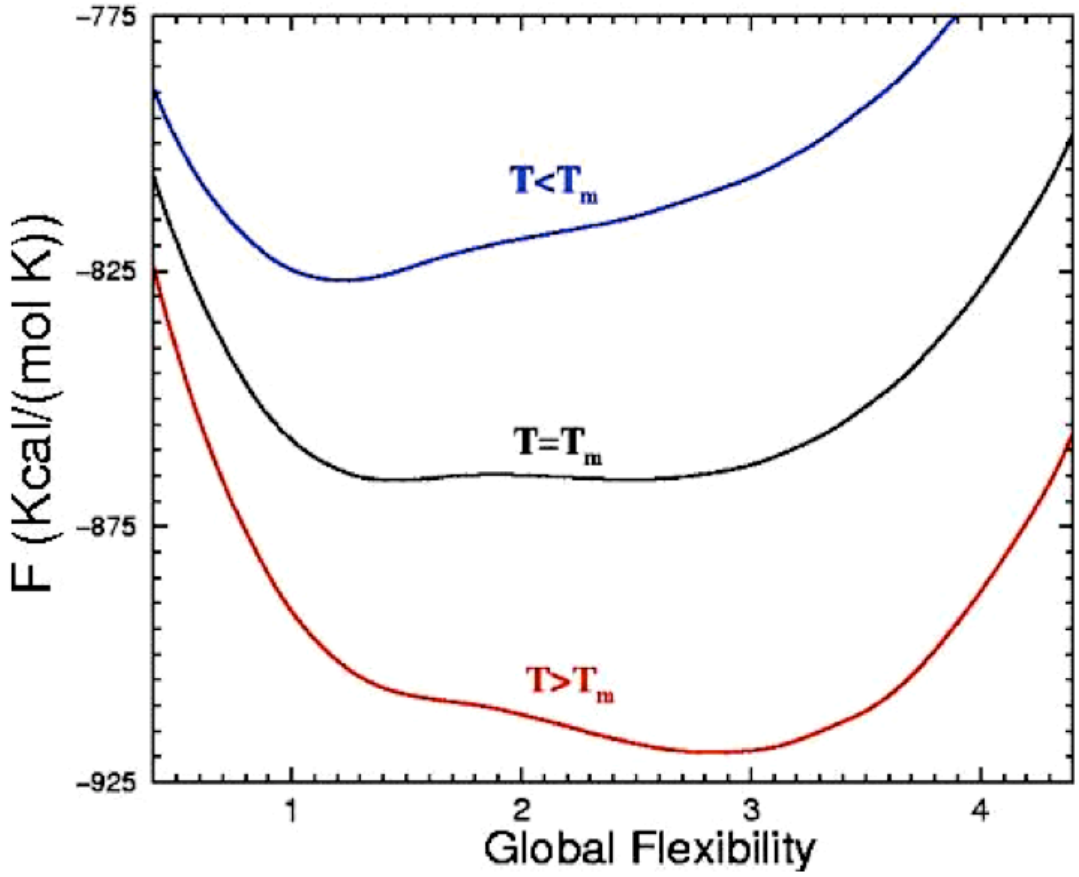
Free energy function: $G(N_{hb},N_{nt}) = U_{IHB} - u\,N_{hb} + v\,N_{nt} - T(S_c(\delta_{nat}) + S_{mix})$



minimal DCM (mDCM): The mDCM is an all atom model. Monte Carlo sampling is employed to generate mechanical frameworks. All mechanical frameworks that are generated at a given grid point will have the same global properties defined by order parameters, which define the number of native torsion interactions and number of native hydrogen bonds. Other than the mean field simplifications, the particularly simple free energy decomposition and statistical errors that appear from the Monte Carlo sampling, it is possible to calculate the partition function accurately and rapidly. Therefore, all thermodynamic properties of interest, including heat capacity, can be calculated numerically. By using the native state topology of the protein, no simulation of the protein motion is necessary. This introduces the approximation that only native contacts are considered. As it can be seen, a lot of approximations were made, and for this reason the qualifier “minimum” has been used to emphasize that only the bare-minimum essence of rigidity interactions is retained carefully in the model. Other types of interactions involving solvation can use improvement. However, this mDCM has proven to be quite useful, generating many results consistent with a large number of proteins and experiments.

Parameter fitting: The problem is that this calculation would be good if the empirical parameters were known. However, the empirical parameters are not known. The procedure is to guess the parameters, and then perform the entire calculation of the free energy landscape, and calculate the partition function, and then the heat capacity. Using simulated annealing, the three parameters are randomly guessed until the predicted heat capacity curve matches the experimental curve well. This empirical approach offers a powerful pragmatic way to tackle the drug discovery process. Over the years, we find that the model parameters are physically reasonable, and are transferable better than that in the Lifson-Roig model for example. In short, the mDCM is an impressive testimony for how a model can capture the essential physics to make useful predictions.

$$\text{Global Flexibility} = \frac{\text{number of independent degrees of freedom}}{\text{number of residues}}$$



The two-dimensional free energy landscape from the previous slides is converted into a one-dimensional free energy landscape by partial integration. Notice that the global flexibility is an intensive quantity that quantifies the number of independent degrees of freedom in a protein. This is not a primary order parameter, but is calculated in the rigidity calculations exactly. By binning the two-dimensional free energy landscapes expressed in terms of two types of constraints, one can quantify how flexible a protein is without reference to which type of constraint is restricting the motions. As such, a direct relation between free energy of a protein and its degree of flexibility is obtained. At low temperature the protein will become globally rigid, and at high temperatures the protein will be globally flexible.

Key idea: Proteins are comprised of rigid and flexible regions to various degrees and distributed in different ways, and the stability of a protein is directly linked to its degree of flexibility.