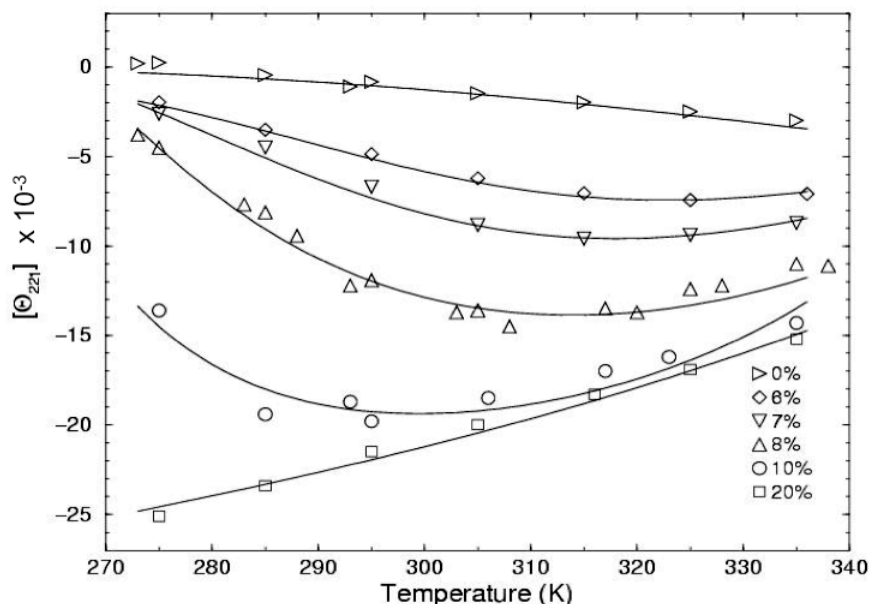D.J. Jacobs and G.G. Wood, Biopolymers. **75**, 1-31 (2004).


Experiment: N. H. Andersen, et. al. J. Am. Chem. Soc. 118, 10309 (1996)
Monomeric peptide in different concentrations (by volume) of HFIP

The figure shows that the DCM when applied to the helix-coil transition, and solved using the exact transfer matrix method, is able to fit to experimental data well. The system has some characteristics as cold denaturation even in a helix-system, but the transition is too shallow to be a strong signal of cold denaturation. Nevertheless, the DCM is able to fit to the experimental data with less than 2 free parameters per curve, compared to the standard Lifson-Roig theories that require a minimum of 3 parameters per curve.

A powerful aspect of the DCM is that it is a full-fledged microscopic statistical mechanical theory, but it is still an empirical model because it has been casted in terms of free parameters to describe solvent effects as interfacial boundary terms. Here, the solvent is water plus various concentrations of **hexafluoroisopropanol**. The solvation terms alter the relative importance of various energy-entropy compensation mechanisms, which are clearly captured by the model very well.

Key idea: A successful model need not have to get every atomic-level interaction perfectly correct to make accurate predictions. The only requirement is that the form and the fundamental characteristics of effective interactions must be correct. By fitting to experimental data, complex phenomena can be expressed in terms of a few adjustable parameters that control interaction strengths that are known to exist in principle, but difficult to precisely obtain by means of ab initio calculations. Consequently, the DCM provides a means to rapidly predict and understand large sets of systematic data, which would otherwise be impossible to obtain through brute-force methods, such as though quantum mechanics calculations or molecular dynamic simulations.
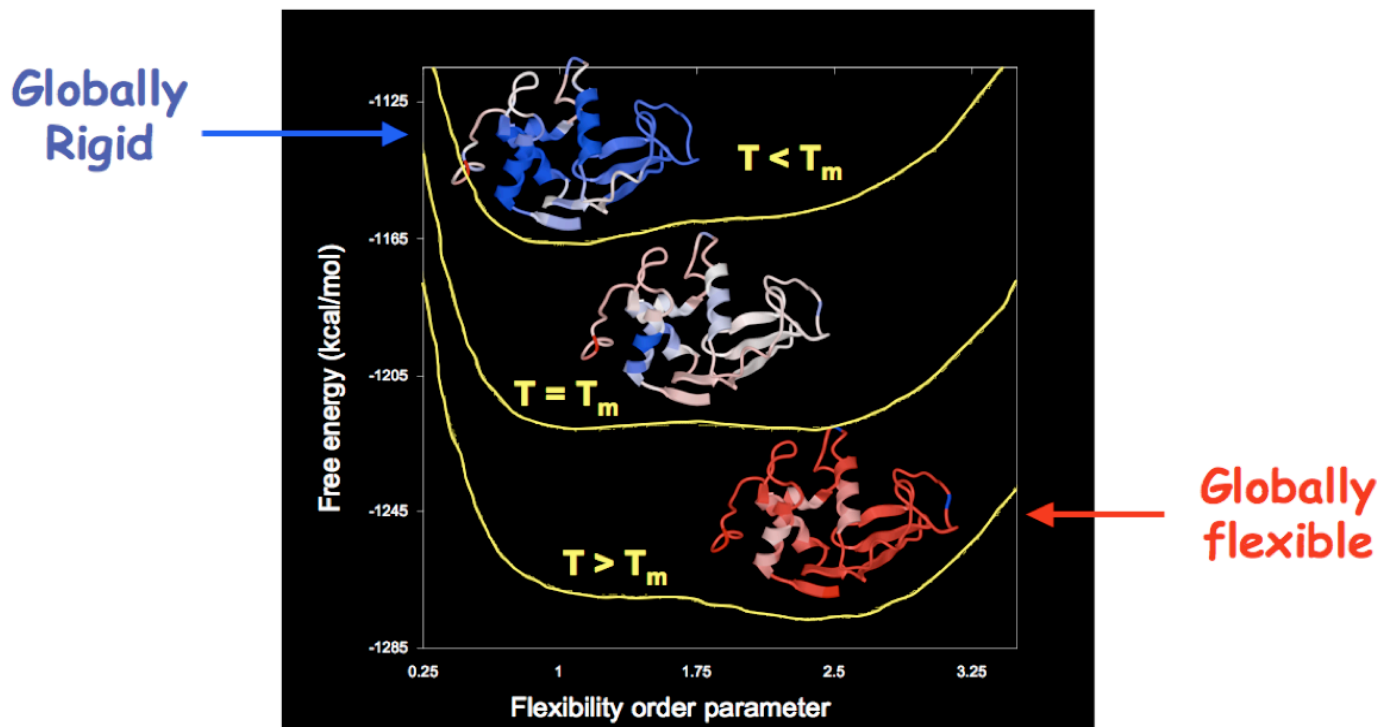
One-dimensional free energy landscapes can be readily calculated using the mDCM that directly link the free energy of a protein to its global flexibility. In the figure above, when the temperature is below the melting temperature of the protein (folding temperature), a protein will typically be globally rigid. In native conditions the protein will not be 100% rigid. Due to fluctuations and the heterogeneous nature of protein structure, pockets of flexible regions will be present in the protein. Flexible and rigid paths within a protein lead to interesting correlated motions. At temperatures above the melting temperature (otherwise called the transition temperature), the protein will be very flexible, but some rigid regions will be present within the protein, at least up to some degree. To traverse between the two states, it is usually the case that there is a free energy barrier to cross, as shown in the figure above. Mechanical and thermodynamic properties of the transition state can be readily calculated using the mDCM.

The backbone flexibility characteristics are displayed using red and blue as mutually flexible and rigid respectively. White shows a marginally mechanical stability where the degrees of freedom available are just about balanced with the number of distance constraints. The interesting aspect of the mDCM is that it predicts which atomic pairs can move relative to one another without ever simulating atomic motions. Note that there is a difference between flexibility and mobility.

Key idea: The detailed flexibility characteristics of a protein can be generated rapidly using a mDCM.
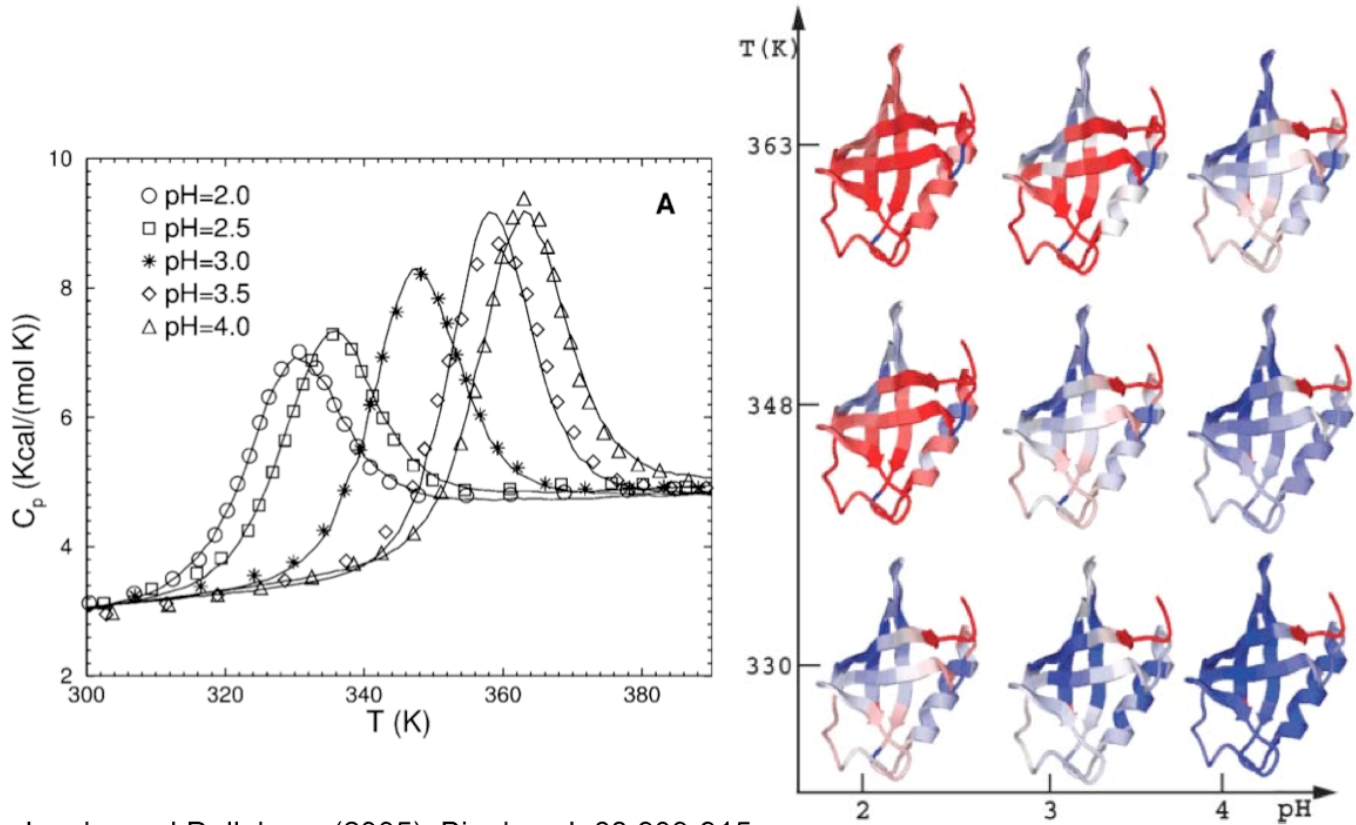
Jacobs and Dallakyan (2005). Biophys J, 88:903-915

Using the mDCM, its empirical parameters can be adjusted to account for pH effects. Interestingly, the backbone flexibility characteristics averaged over the native ensemble at the transition point is very similar for all the distinct situations. This result supports the idea that overall structure is the main determinate of backbone flexibility and backbone mobility, and changing thermodynamic and solvent conditions merely shifts the transition points, but not so much the mechanisms. Having said this, there are some variations, and the mDCM is too simplified to make fine distinctions. Nevertheless, it does make sense that over a narrow range of perturbing conditions, large-scale mechanical properties will not change dramatically.

Key idea: Backbone characteristics of network rigidity remain conserved within the native ensemble defined by thermodynamic/solvent conditions at the transition point defined in part by the melting temperature.

Global flexibility = $\dfrac{\text{average number of independent degrees of freedom}}{\text{number of residues}}$

Native basin of thioredoxin (TRX) has greatest rigid ⇔ flexible fluctuations
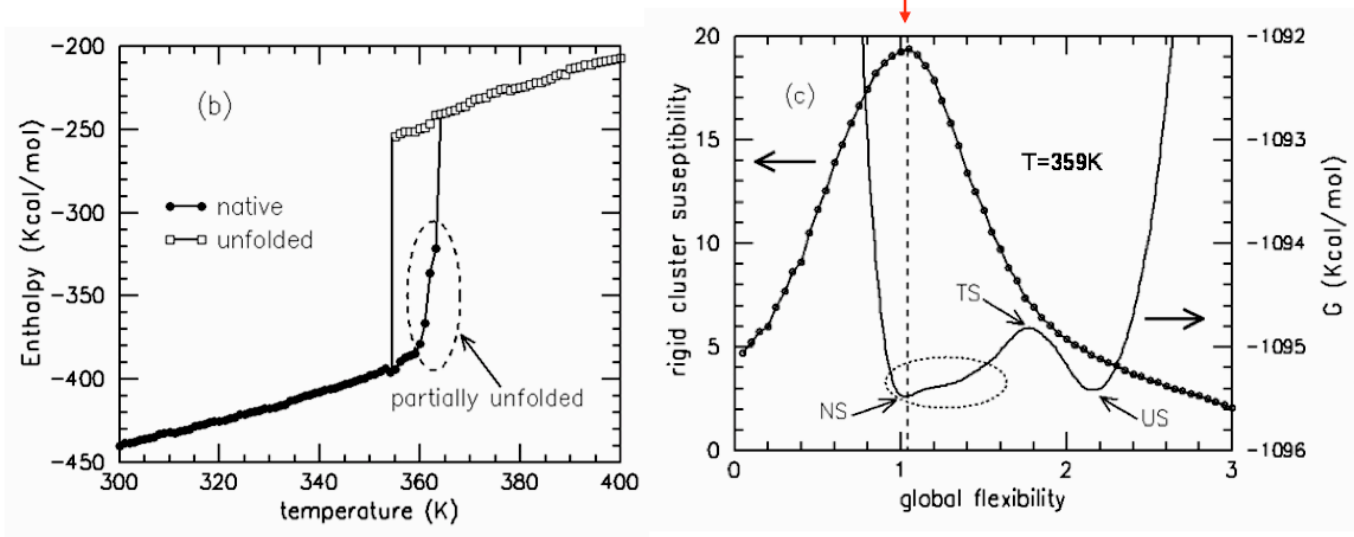
D.J. Jacobs, et. al., Journal of Molecular Biology. **358**, 882-904 (2006)

An interesting discovery with thioredoxin **is that it undergoes partial unfolding** (left figure) **and its largest fluctuations in rigid cluster size,** also called **Rigid Cluster Susceptibility (RCS)**, occur in the native state basin (right figure). Different proteins have the peak in RCS at different points along the global flexibility order parameter, where it has been located beyond, at, or below the transition state. By looking at the relationship between mechanical and thermodynamic properties provides insight in how the protein functions. For example, hydrophobic regions within aqueous proteins tend to be more dynamically flexible, with larger RCS.

Key idea: Reasonably accurate predictions can be made at high throughput speed using the mDCM or other variants of DCM.
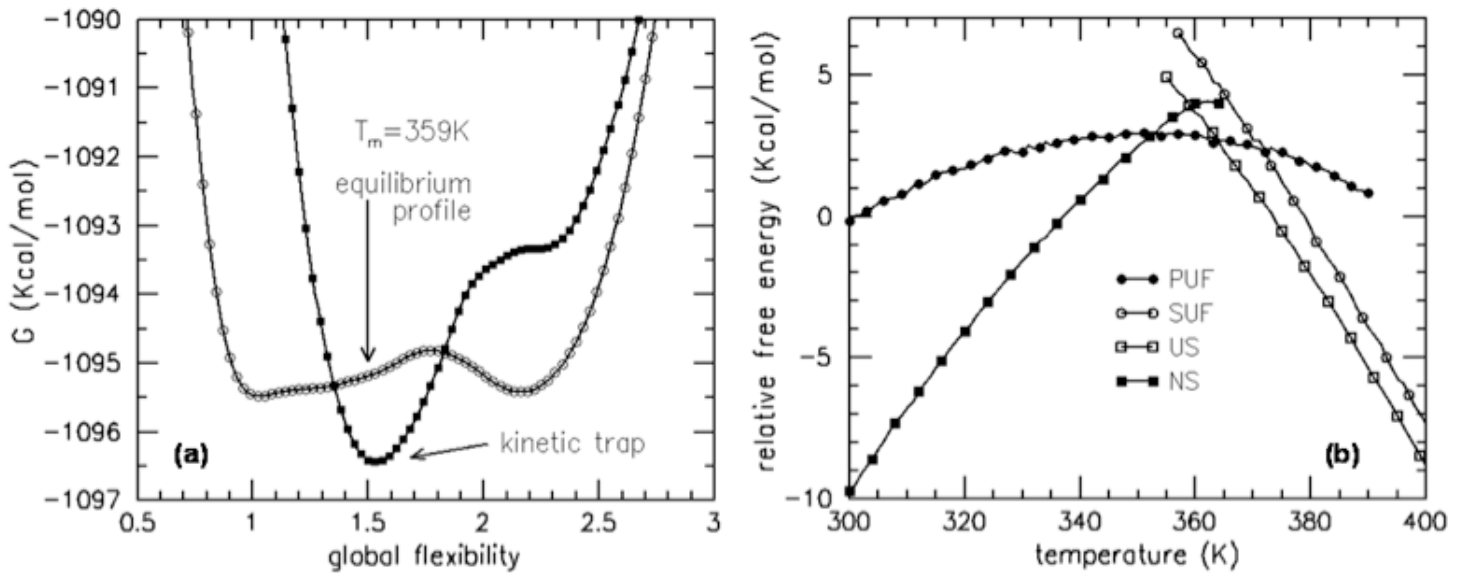
D.J. Jacobs, et. al., Journal of Molecular Biology. 358, 882-904 (2006)

Due to partial unfolding events, multiple macrostates of thioredoxin can be identified. A native state is defined when **partially unfolded regions** undergo thermodynamic fluctuations without encountering a long-lived partial unfolding event. This means kinetics comes into the picture from an experimental viewpoint. At low temperature the native state will be most populated, and it would be difficult to detect the partial unfolding event. On the other hand, it is possible that the mechanically unstable region remains unfolded for relatively long times. In this case, the thermodynamics and dynamics of the protein will be different because the ensemble has a conditional restriction (a section of the protein is already unfolded and has passed through its own free energy barrier). As such, the free energy landscape of the protein in the partially unfolded state (shown in the left figure) can be more or less stable depending on the thermodynamic/solvent conditions. The partially unfolded native state is actually more stable than the normal native state for a small temperature range (right figure). Similar considerations apply to the unfolded state, except the partially unfolded part remains unfolded. These predicted changes in thermodynamic quantities are consistent with kinetic experiments.

Key idea: The mDCM results provide a wealth of information that other methods cannot address.
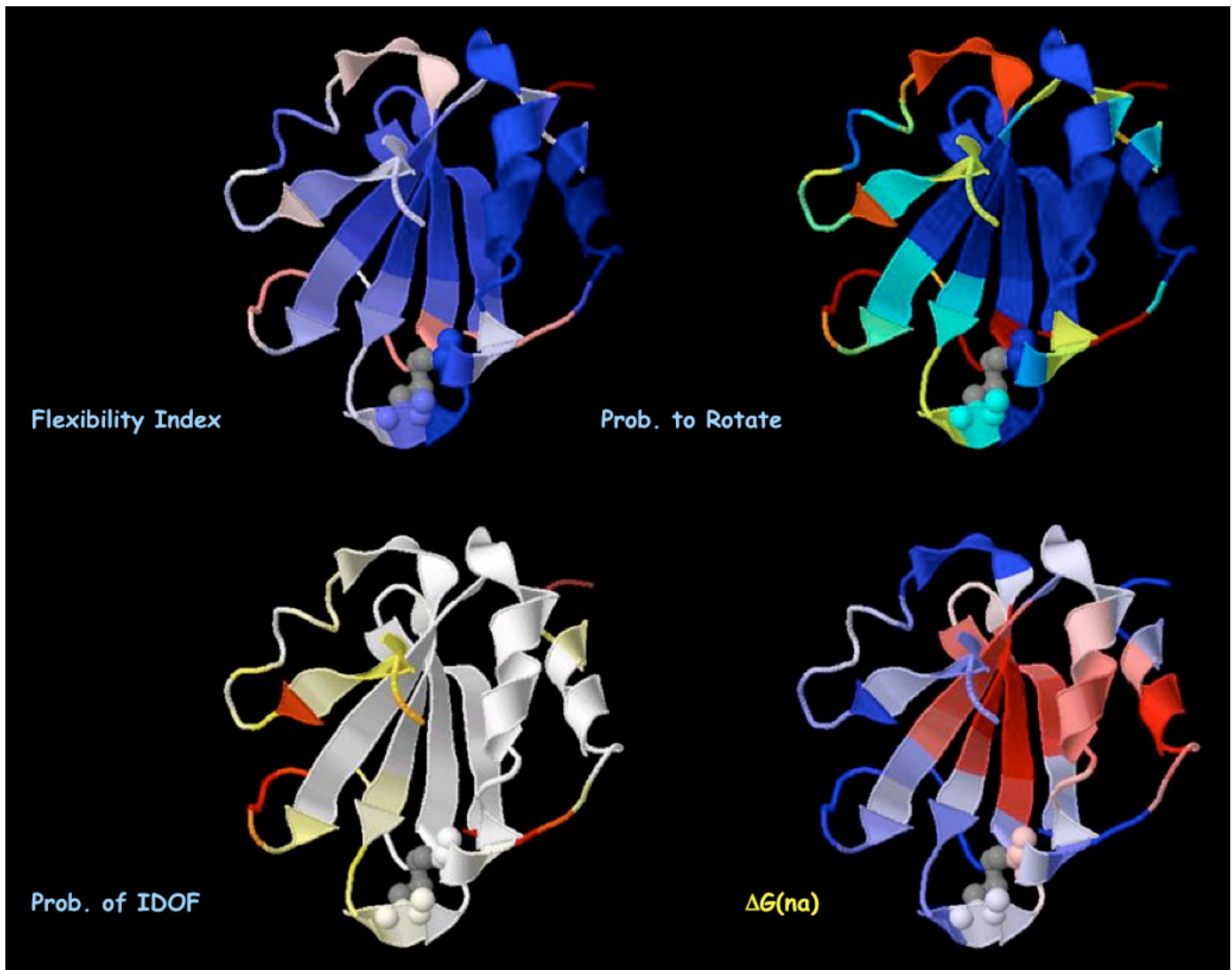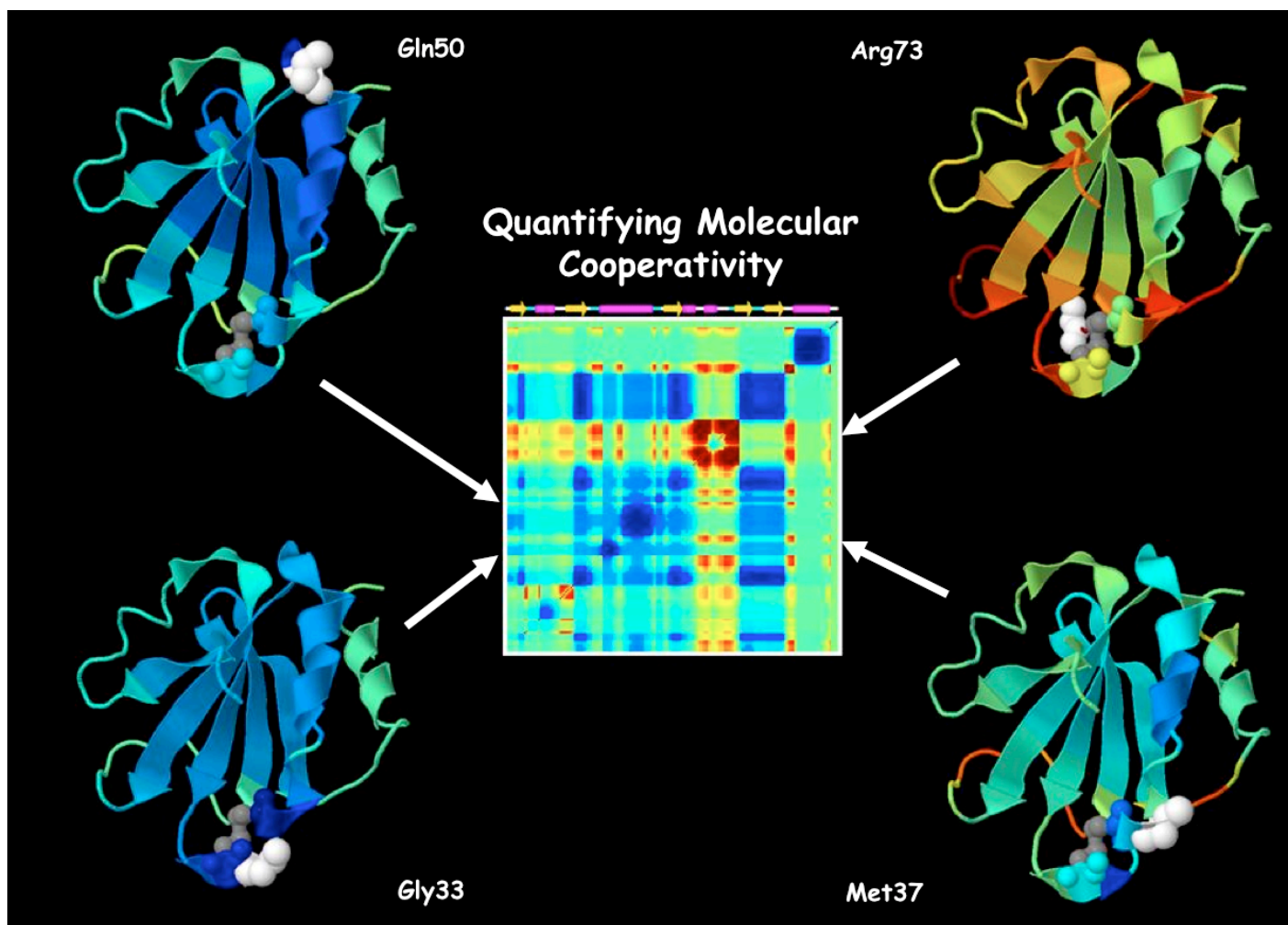
D.J. Jacobs, et. al., Journal of Molecular Biology. 358, 882-904 (2006)

Many different backbone measures can be calculated and rendered on a protein structure. The above figure shows four such examples. The results of many different quantities have elucidated the properties of thioredoxin remarkably well. The different metrics provide ample ways to rap your head around the data, and make sense of how the protein wiggles and jiggles to carry out its function.

Key idea: Multiple metrics that characterize backbone properties can be calculated, and the results impressed onto sequence alignments to facilitate comparative studies across protein families.

D.J. Jacobs, et. al., Journal of Molecular Biology. 358, 882-904 (2006)

In addition to backbone metrics, it is possible to look at **pairwise correlations** between residues. The center square matrix represents **cooperativity correlations** in dihedral angles being rotatable. Two dimensional matrix representations of the data are hard to interpret. Thus, the four protein structures show a strip of color taken from one column of the residue-residue pairwise coupling matrix (shown in the middle). The perceived rigidity or flexibility is highly dependent on the observation point, which is highlighted by the space-filled atoms as a reference point. The reference point makes a difference because the information that is being plotted is related to pairwise correlations. If two residues where flexibly correlated the color would be red (or yellowish for less amplitude) while blue is for rigidly correlated (or cyan for less amplitude). The color green indicates no correlation.

Key idea: Design of proteins can take advantage of being able to identify cooperativity correlations that exist between residues.
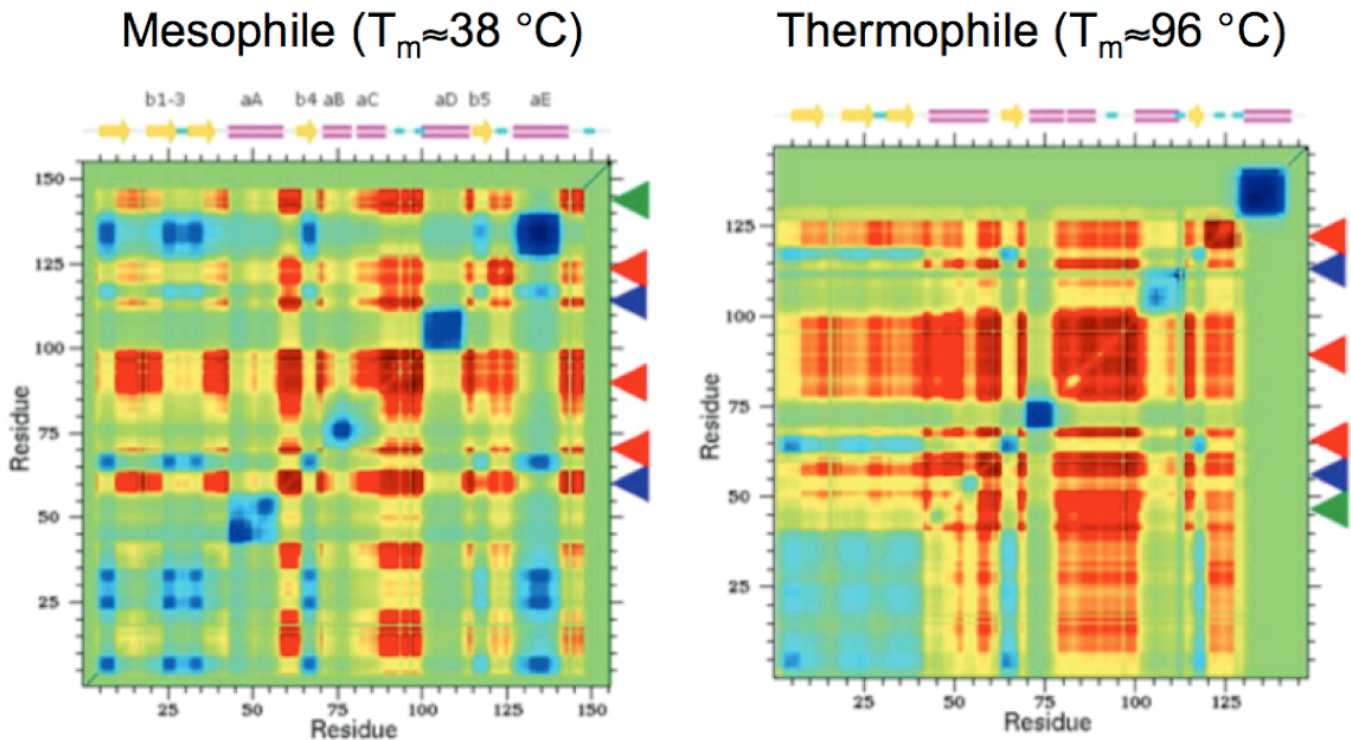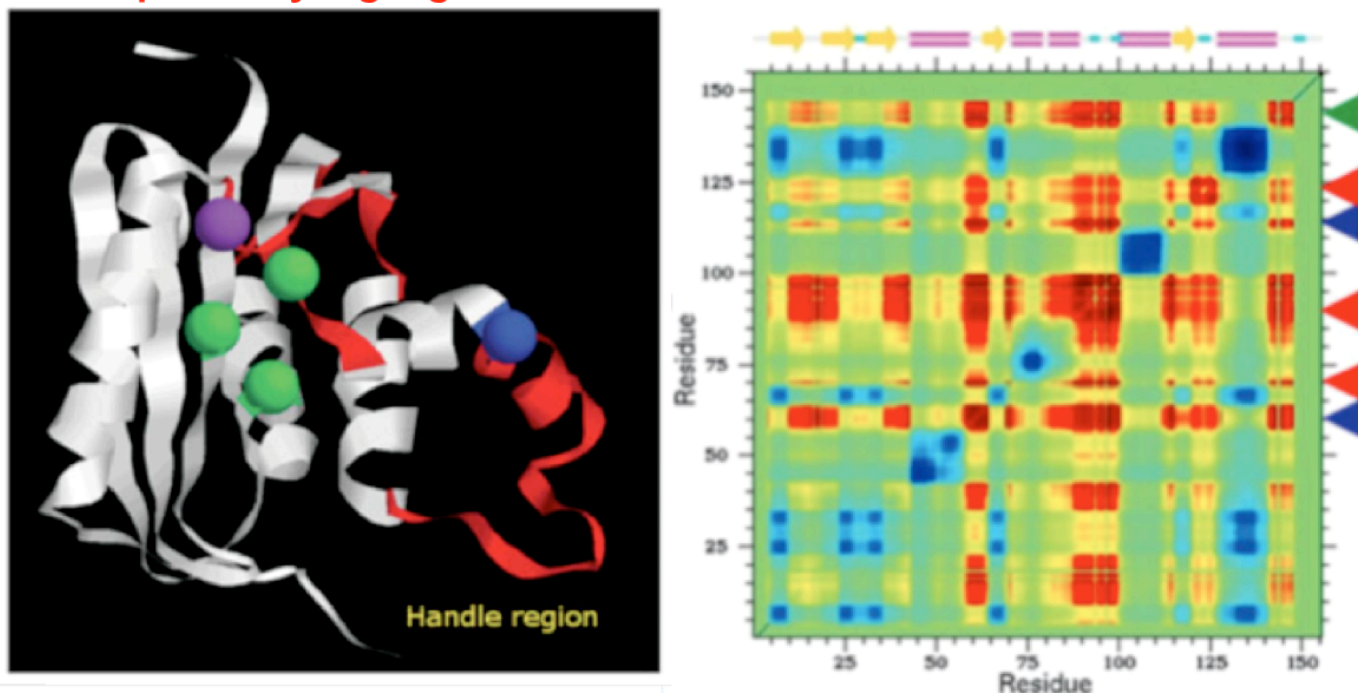
D.R. Livesay and D.J. Jacobs, Proteins, **62**, 130-43 (2006)

The same type of residue-residue coupling analysis was done on a pair of **ortholog RNase proteins**, corresponding to a **mesophilic** (left) and **thermophilic** (right) type. We find considerable differences between the two orthologs in terms of their energy-entropy compensation mechanisms, and overall, the mesophilic protein exhibits more rigidity than the thermophilic protein. However, a comparison can be made between the two at their respective melting temperatures. Averaging over the **native basins** at T_m, we find many similarities with respect to backbone flexibility. Nevertheless, there is also many detailed differences in the residue-residue couplings (as shown above).

Key idea: Looking closely at similarities and differences in a QSFR comparative analysis provides insight into the thermal-mechanical mechanisms that are important (and possibly critical) for function.

Cooperativity highlighted in red

Handle region

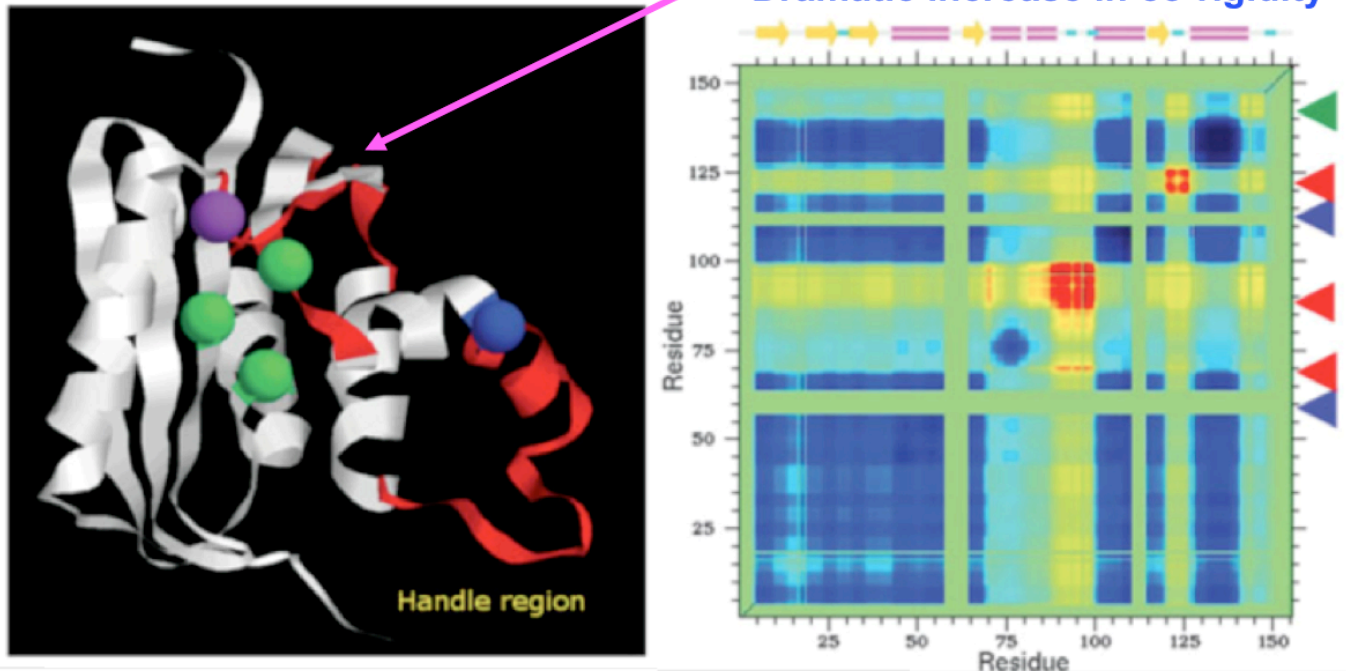D.R. Livesay and D.J. Jacobs, Proteins, **62**, 130-43 (2006)

The information in the residue-residue coupling matrix (shown to the right) can be coarse-grained to show the **communication pathways** that are present in the protein structure (shown to the left). The red regions highlight a **flexibly correlated region** that couples to the **handle region**. The handle region is known to be flexible and highly mobile and that it must swing back and forth to fold over onto the active site, which is highlighted by space-filled atoms. The mDCM predicts the handle region to be flexible, but moreover, that it is flexibly correlated with loop regions that extend to the "backside" of the protein. While the mobility of this extended cooperativity is not high, it nevertheless can potentially affect the flexibility (and mobility) in the handle region. Interestingly, identifying these correlations in the degrees of freedom of the system (otherwise called communication pathways) provides insight into how long-range (distal) information can exchange through the protein, albeit in the form of energy transfer in the form of motion, or some other **population shift mechanism**. A population shift occurs when one looks at the relative statistical weights of different accessible microstates. No direct mechanical linkage is necessary to get a shift, as simple toy examples can be devised to demonstrate this. In general, **mechanical and thermodynamics effects are occurring simultaneously and are intimately coupled**. Consideration of one without the other is like trying to eat ice cream without ice and just drink rich milk, or without the cream, and just eat some ice cubes. The linking of mechanics and thermodynamics is critical to discover simplicity in otherwise perceived complexity.

Key idea: Mapping out where flexible and rigid regions are located within a protein, and quantifying molecular correlations and how these correlations are linked to the thermodynamic/solvent conditions provides a powerful interrogation tool to identify communication pathways in proteins.

Add constraints in flexibly correlated region far from active site

Cooperativity highlighted in red

Dramatic increase in co-rigidity

Handle region

D.R. Livesay and D.J. Jacobs, Proteins, **62**, 130-43 (2006)

By perturbing the opposite side of the protein relative to the active site, the cooperativity correlations can be dramatically affected in the mesophilic protein (compared with the previous slide). This is an example of an allosteric effect, where a small perturbation can induce a large effect at far distances away from the perturbation site. Interestingly, even a stronger perturbation than that used in the mesophilic protein does not induce an allosteric response in the thermophilic protein. This is because the thermophilic RNase is intrinsically more flexible, and is governed by different energy-entropy mechanisms. This type of interrogation using "what if" scenarios provides a means to perform in silco high throughput screening to design proteins with specific target responses.

Key idea: The DCM approach is ideal for studying allostery in proteins, because it considers both the thermodynamic and mechanical properties in a harmonious way at the all-atom level that is also fast enough to do in a high throughput setting.
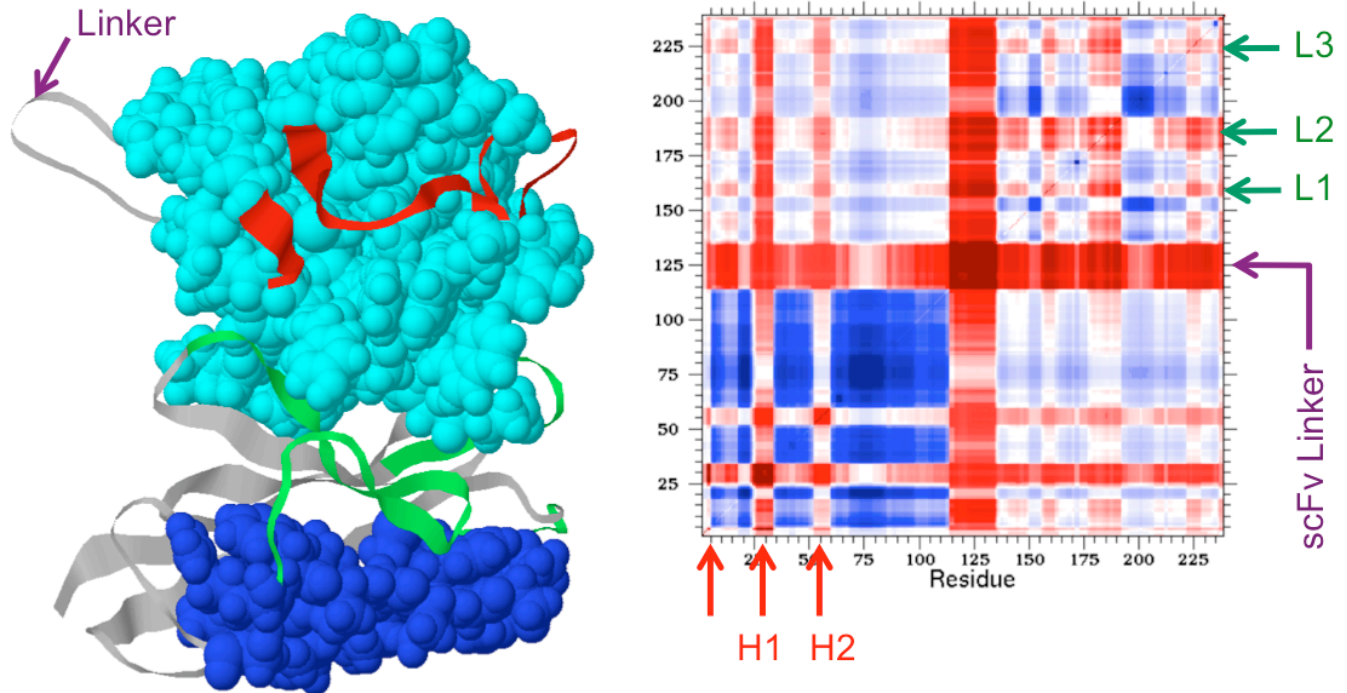
The mDCM approach is applied to antibody fragments, such as in scFv with an engineered linker. Here is an example where the molecular cooperativity plot shown to the right (using a different quantity to track the cooperativity than the previous slides) is coarse grained onto the structure shown to the left. The cyan and blue regions on the top and bottom domains are rigidly correlated. The red and green regions, as well as the linker (uncolored) are flexibly correlated. The interesting phenomena that one sees from these rigidity calculations are that flexibly correlated motions propagate over long distances through rigidly correlated regions. Thinking about this allosteric mechanism carefully should prove to be perplexing, at least for a while. How can you wiggle the red flexible region in the top domain to make the flexible green region wiggle, even though this motion is being propagated through a rigid region? This type of effect is commonly seen in proteins, and is the main reason why allostery is so difficult to characterize. But how can this be possible using the mDCM since it is based on network rigidity calculations, which would not allow such a possibility? --- Or would it? Since these are the results from mDCM, obviously it is possible, and thus, we know it must be possible experimentally too, since we are using statistical physics and mechanics to describe what is happening. As is true with many things in physics, complex phenomena can often be associated with a simple explanation.

The answer (if you haven't already figured it out from my clues) is **fluctuations**. All that is shown in these color plots is the average result over an ensemble of mechanical frameworks. There are flexible mechanisms that connect the two flexibly correlated regions in question, and the loops themselves are almost always flexible. However, on average the regions between the loops are rigidly correlated. It is for this reason that an athermal mechanical model (such as FIRST) or thermodyanic models with out rigidity fail. Both aspects are needed, and both aspects represent pillars of physics.

Key idea: The DCM approach has all the essential physics necessary to predict the complex patterns observed in proteins, although improving model details (relative to the mDCM) are important to obtain very accurate predictions.

Verma et al (2012)
PLoS Comp Biol, 8:1002409.
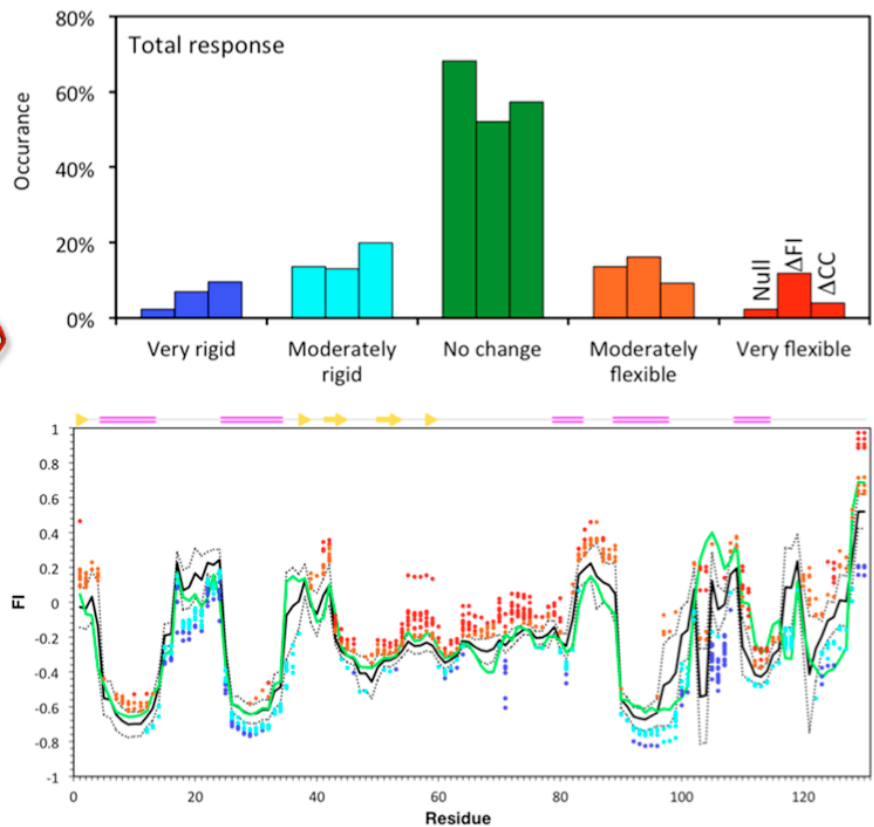
While most of the illustrative examples and descriptions given within this QSFR document refer to earlier results (2004-2006), recent results on anti-bodies (shown on previous slide) and lysozyme single point mutants shown above are no less interesting. Over the last 10 years we have looked at dozens of diverse proteins, and find that despite the simplicity of the mDCM it provides insight into the thermal-mechanical properties of proteins that are consistent with known experimental results. The figure above that shows results for single site mutants of lysozyme, encapsulates our findings over many different protein systems. On top-left is a ribbon cartoon of one particular mutant protein with a coloring that shows the changes in flexibility across the protein due to the mutation. The bottom-right figure looks at backbone flexibility for the wild type and plots in dots the results from the single site mutants but colored coded to show when the changes are statistically significant or not. The top-right figure forms a histogram to summarize the results from a collection of all mutant proteins.

We find that single site mutations in a protein will frequently lead to large and long-range response. Note that most of the time there is no change, or just minimal change that is not physically relevant. So then: What is meant by frequent, large and long-range? We perform the analysis and then we do statistical t-test and other tests to compare actual response with an assumed random Gaussian process. We find there is statistical significance in these large changes in proteins due to a single mutant. Also comparing QSFR in different proteins from the same family, allows the compensation of different mutations to be studied. Indeed the idea of directed evolution can be applied to the mDCM approach, to use mutations in combinations to target desired QSFR characteristics.

Key idea: By targeting specified QSFR patterns, a directed evolution approach can be implemented to design proteins and drugs within an in silico high throughput screening scenario.