UNC CHARLOTTE

A Virtual Pebble Game to Ensemble Average Graph Rigidity

Donald Jacobs, Professor

Department of Physics and Optical Science

Making Models for Protein Flexibility and Stability

Outline

Background: Distance Constraint Model

Modeling protein stability using graph rigidity.

Statistical ensembles of constraint networks.

Results from ensemble averaging graph rigidity.

Approximate Graph Rigidity Models

Maxwell Constraint Counting

Virtual Pebble Game

Virtual Pebble Game Results

Conclusion and Open Questions

Applications and Motivation

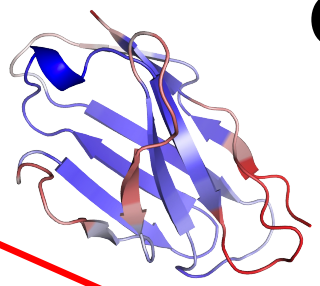
"Essentially, all models are wrong, but some are useful." --- George Box

George Box and Norman Draper (1987). Empirical Model-Building and Response Surfaces, p. 424, Wiley. ISBN 0471810339.

Insight into Thermodynamic Stability

A simple two state model

$$G = H - TS$$

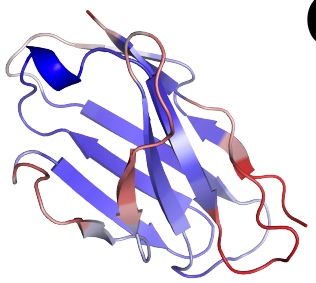


J.A. Schellmann, J. Phys. Chem. 62, 1485-1492 (1958)

Thermodynamic Stability: A Two State Model

Enthalpy-entropy compensation

$$G = H - TS$$

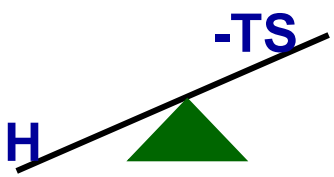


Folded

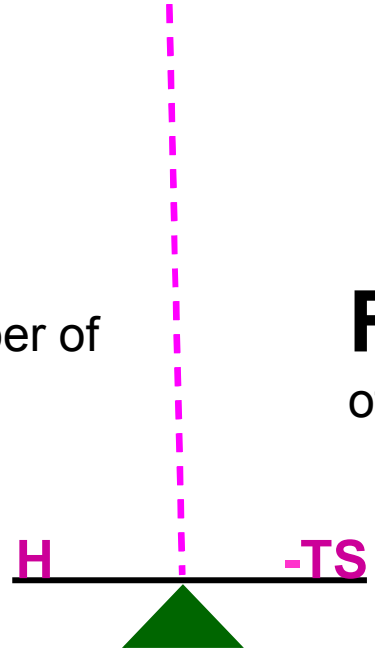
Coil

Rigid, low number of conformational DOF

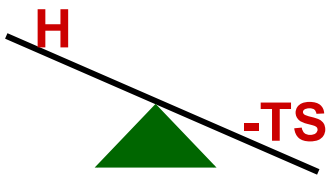
Flexible, high number of conformational DOF



Low T



Transition



High T

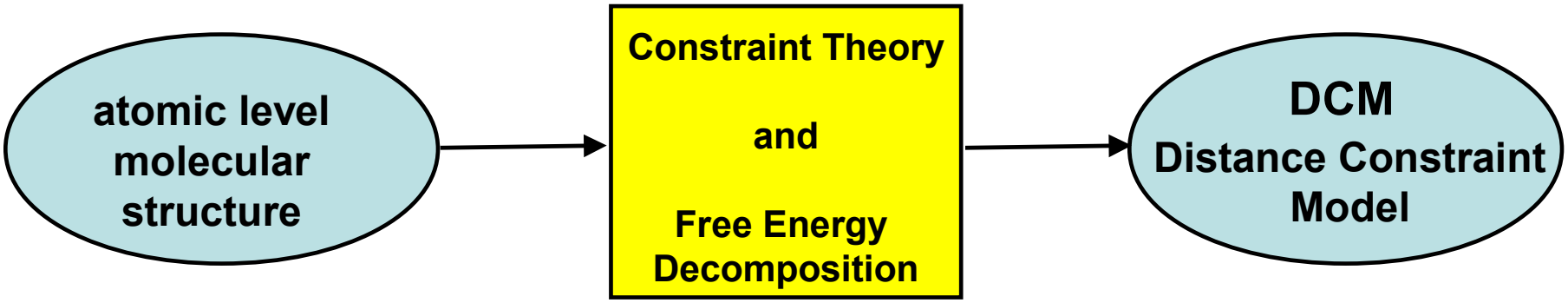
J.A. Schellmann, J. Phys. Chem. 62, 1485-1492 (1958)

Distance Constraint Model (DCM)

Putting thermodynamics into network rigidity

A MECHANICAL PERSPECTIVE

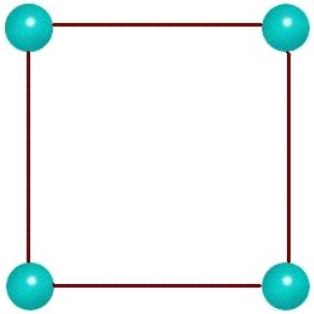
D.J. Jacobs,et. al., *Network rigidity at finite temperature: Relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems*. **Physical Reviews E. 68, 061109 1-21 (2003)**



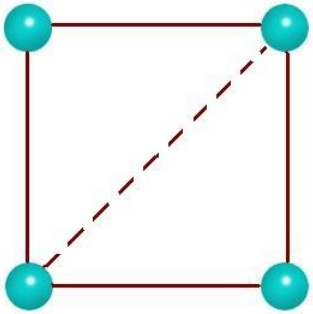
"I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it"! **--- Lord Kelvin**

Tao of the DCM

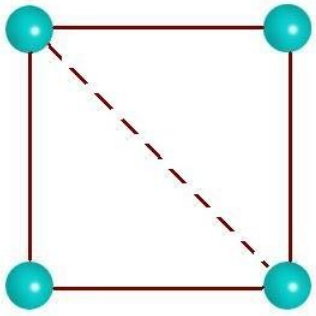
Enthalpy-entropy compensation modeled with mechanical constraints



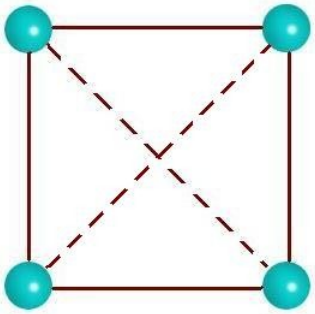
$\Delta H = 0$
$\Delta S = 0$



$\Delta H = -\varepsilon$
$\Delta S = -\delta$



$\Delta H = -\varepsilon$
$\Delta S = -\delta$



$\Delta H = -\varepsilon + -\varepsilon$
$\Delta S = -\delta + 0$

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

Linking Mechanics Directly with Thermodynamics

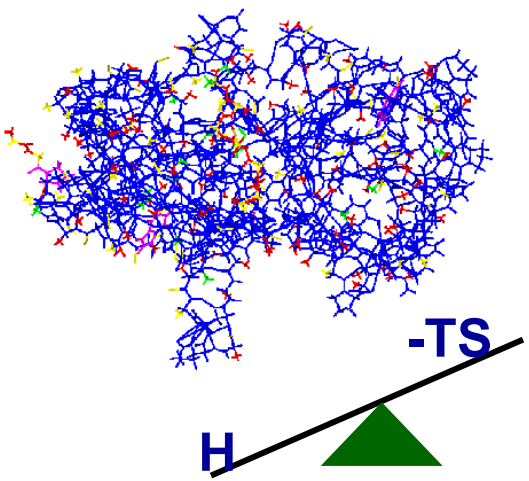
Network rigidity is regarded as an underlying mechanical interaction

Jacobs & Dallakyan (2005) *Biophysical J.* 88:903

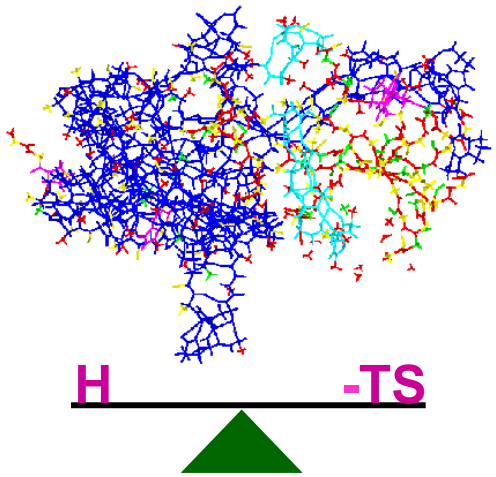
Livesay et al. (2004) *FEBS Letters* 576:468

MECHANICS

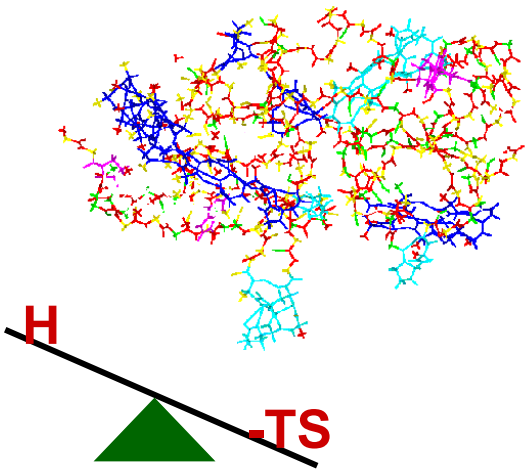
Globally Rigid ⟷ Globally Flexible



STABLE representing the native state (**NS**)



UNSTABLE representing the transition state (**TS**)



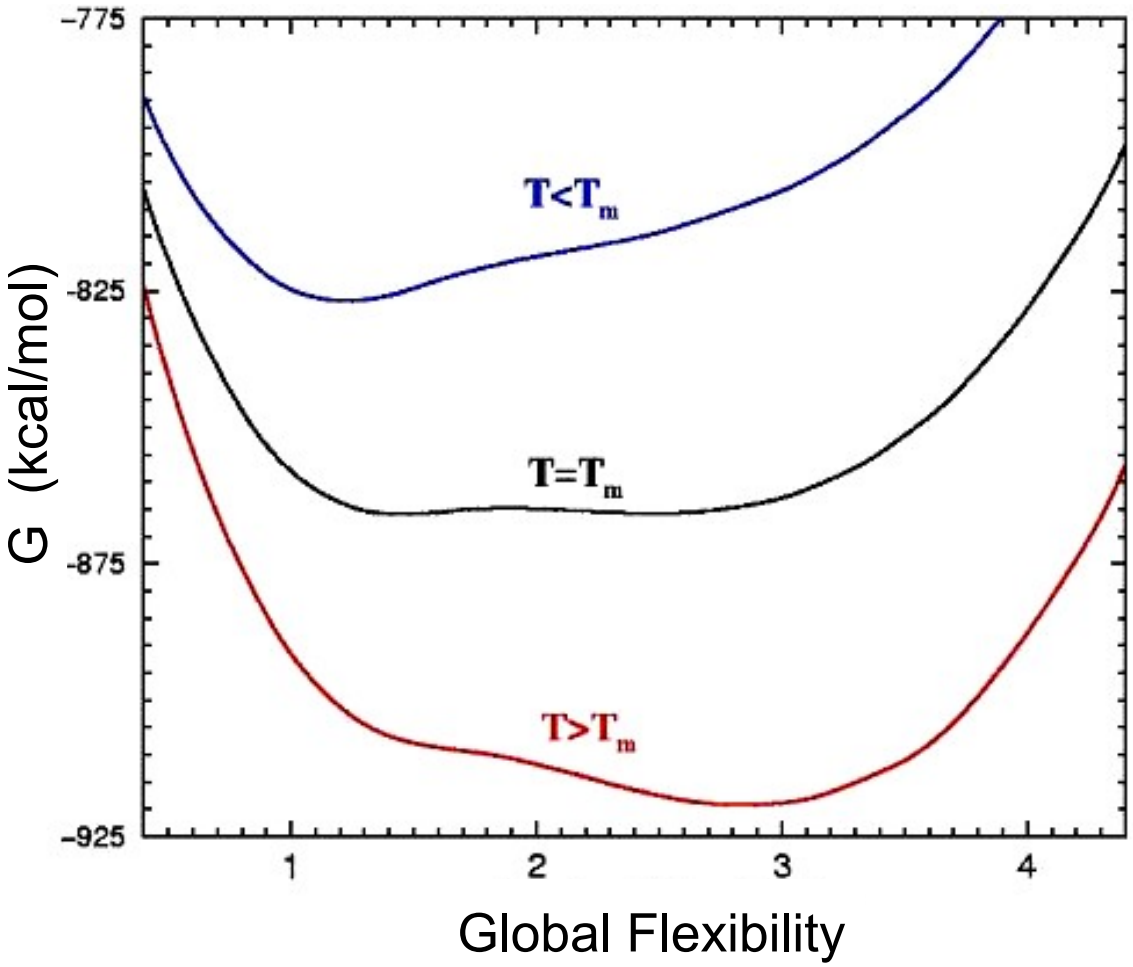
STABLE representing the unfolded state (**US**)

THERMODYNAMICS

1D Free Energy Landscape

Free energy is directly related to the global flexibility of a protein

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

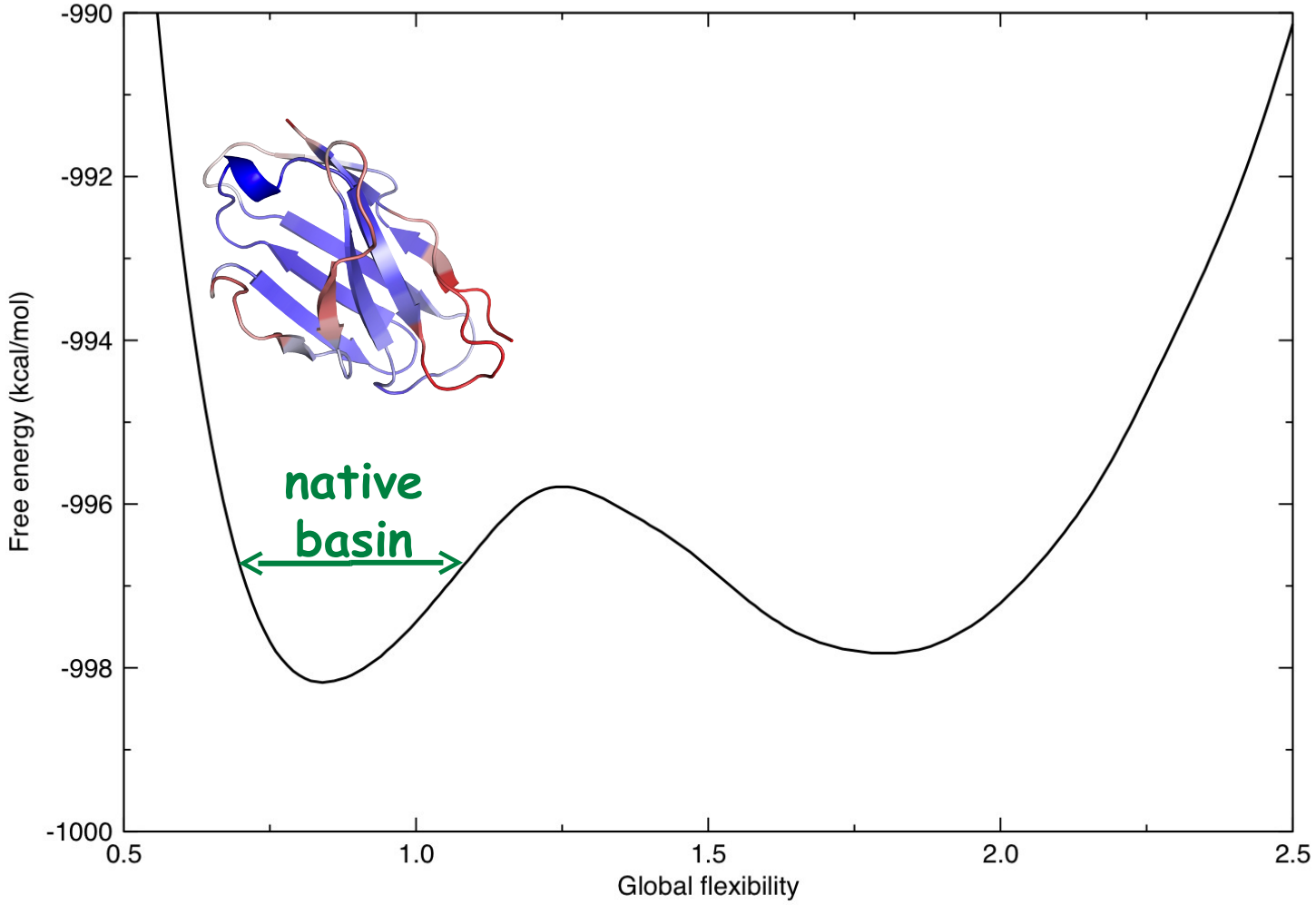


Ensemble Based Methods Probe Fluctuations

Native state fluctuations reflect properties of network rigidity

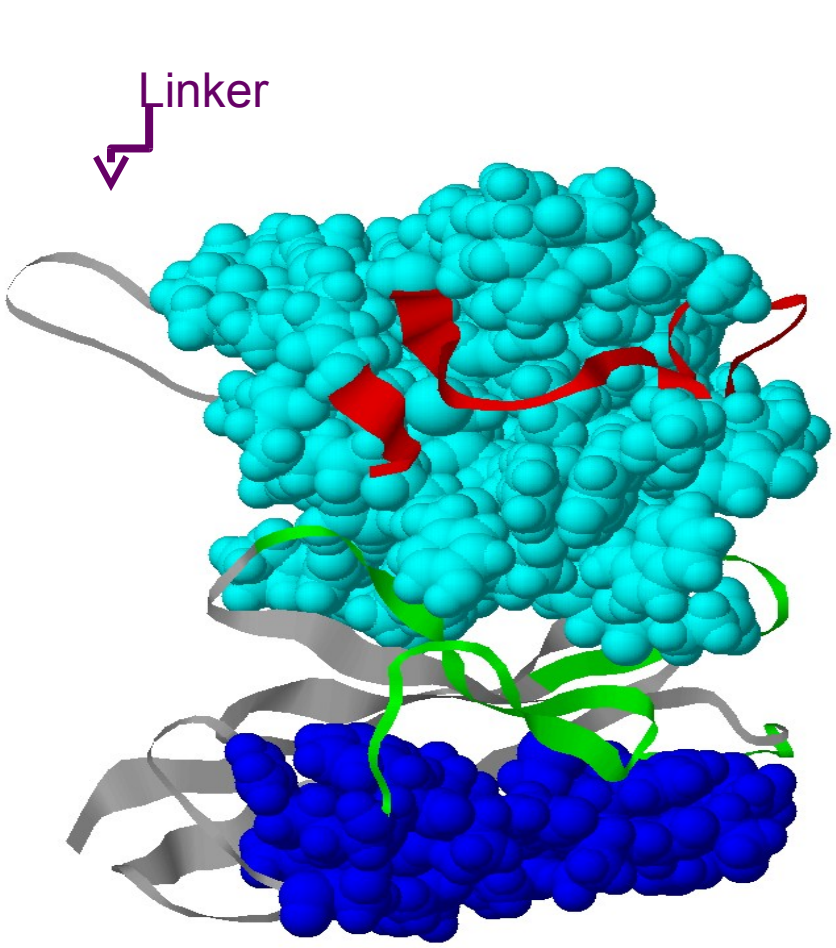
Use known X-ray crystal structure as a geometrical template

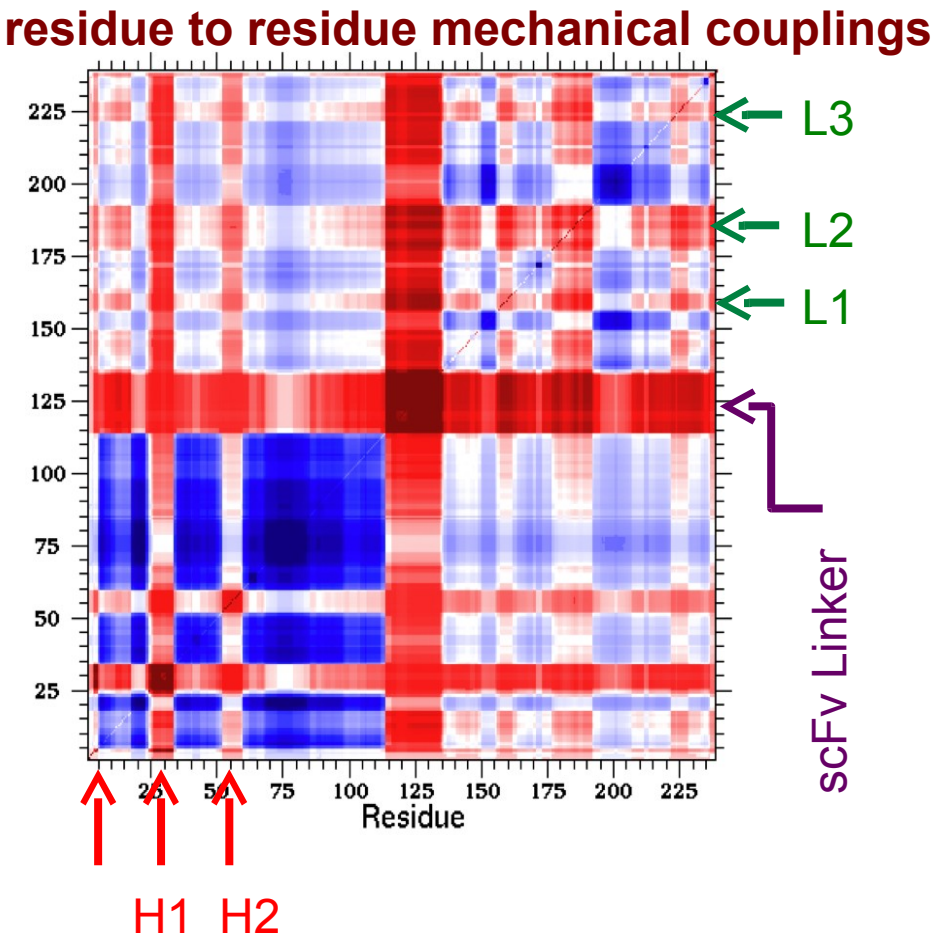
Perturb structure by breaking native state H-bonds (random dilution)



A P104D mutant scFv anti-body fragment

Correlations are found in native state fluctuations





Li T, Tracka MB, Uddin S, Casas-Finet J, Jacobs DJ and Livesay DR (2014) Redistribution of Flexibility in Stabilizing Antibody Fragment Mutants Follows Le Châtelier's Principle. PLoS ONE 9(3): e92870

Sub-ensembles of constraint networks

Rigidity properties change depending on number of H-bonds

Total number of constraints = covalent bond constraints + H-bond constraints

$NC = NCB + NHB$

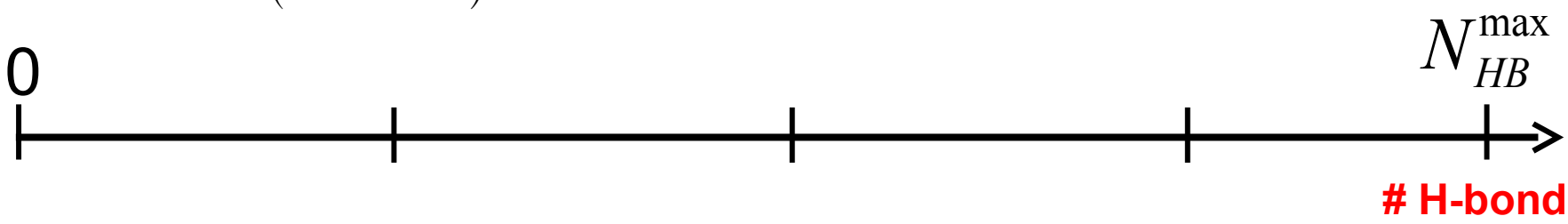
Quenched (always ON)

Fluctuating (ON or OFF)

A *typical* value for maximum number of H-bonds $N_{HB}^{\max} = 200$

$$2^{N_{HB}^{\max}} = \sum_{N_{HB}=0}^{N_{HB}^{\max}} \binom{N_{HB}^{\max}}{N_{HB}}$$

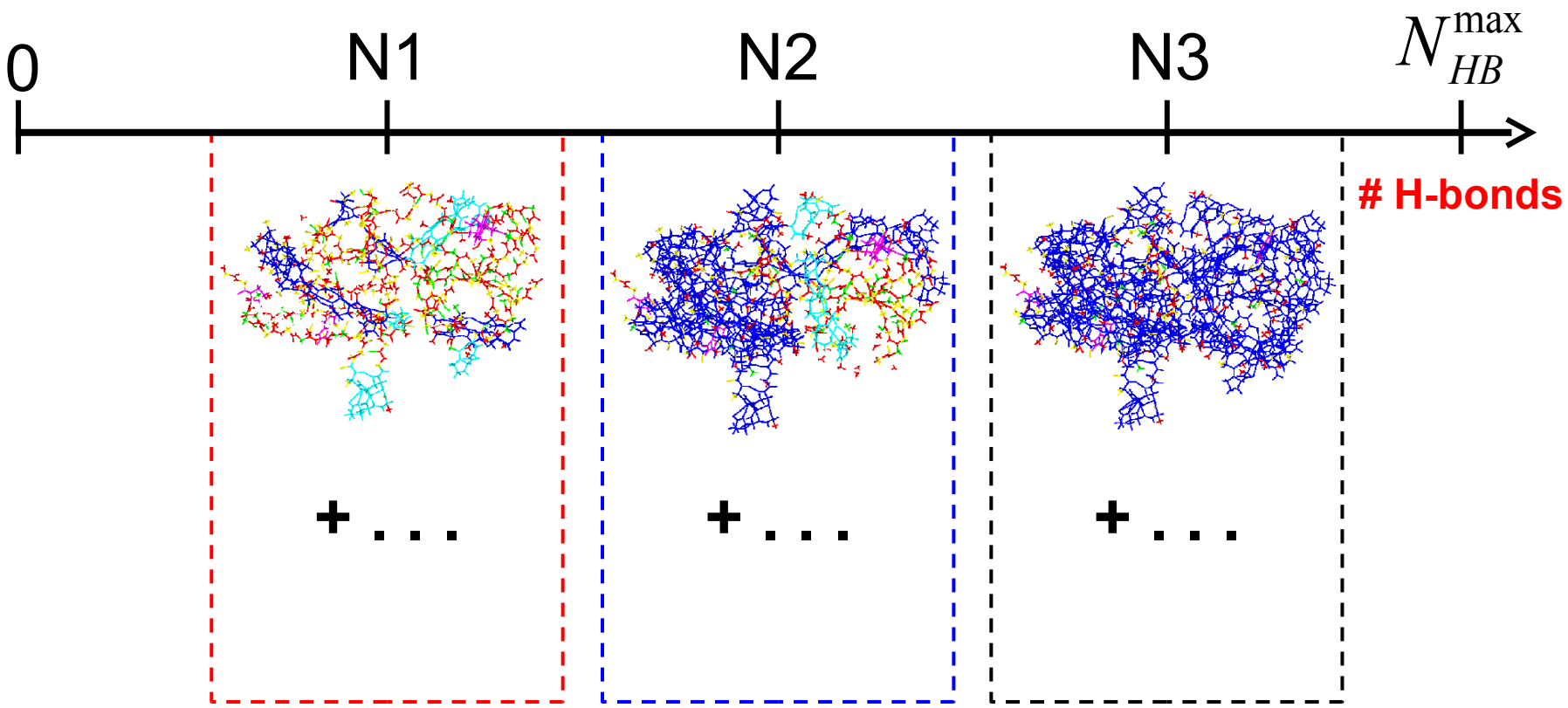
Binomial coefficients give the number of distinct constraint networks with NHB H-bonds present.



Sub-ensembles of constraint networks

Rigidity properties change depending on number of H-bonds

How to estimate average graph rigidity properties in each sub-ensemble?

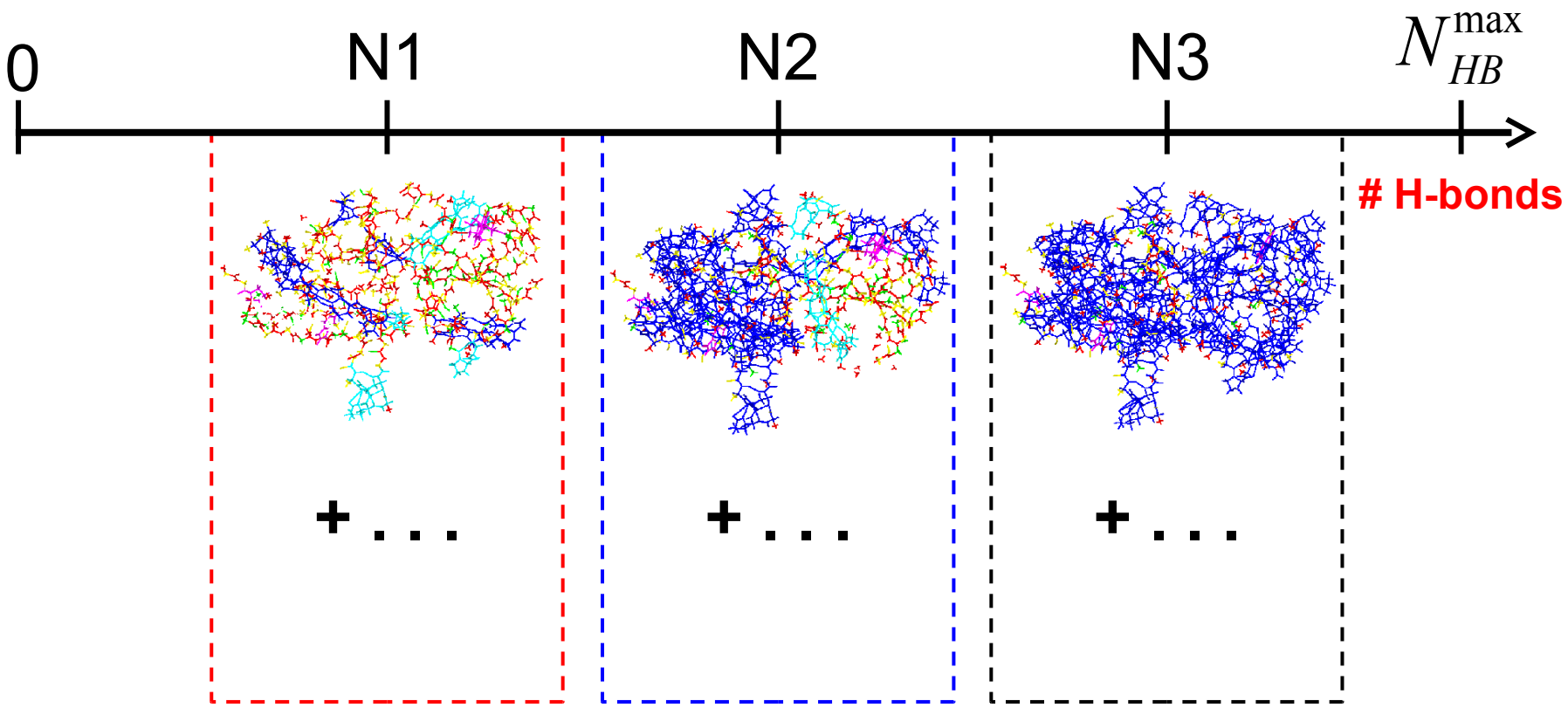


Sub-ensembles of constraint networks

Rigidity properties change depending on number of H-bonds

How to estimate average graph rigidity properties in each sub-ensemble?

Method 1: Monte Carlo sampling (typically run 200 pebble games)

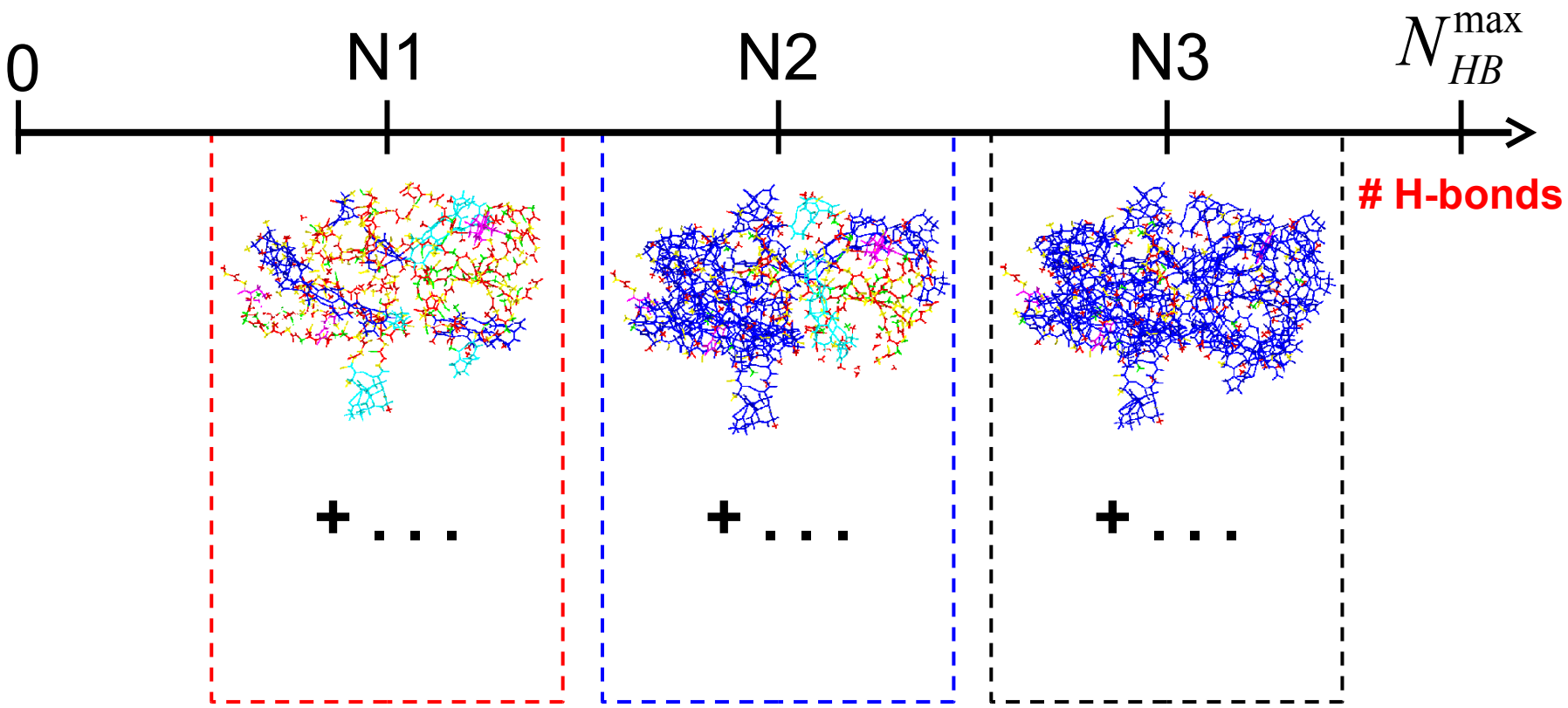


Sub-ensembles of constraint networks

Rigidity properties change depending on number of H-bonds

How to estimate average graph rigidity properties in each sub-ensemble?

Method 2: Maxwell Constraint Counting (estimates number of DOF only)

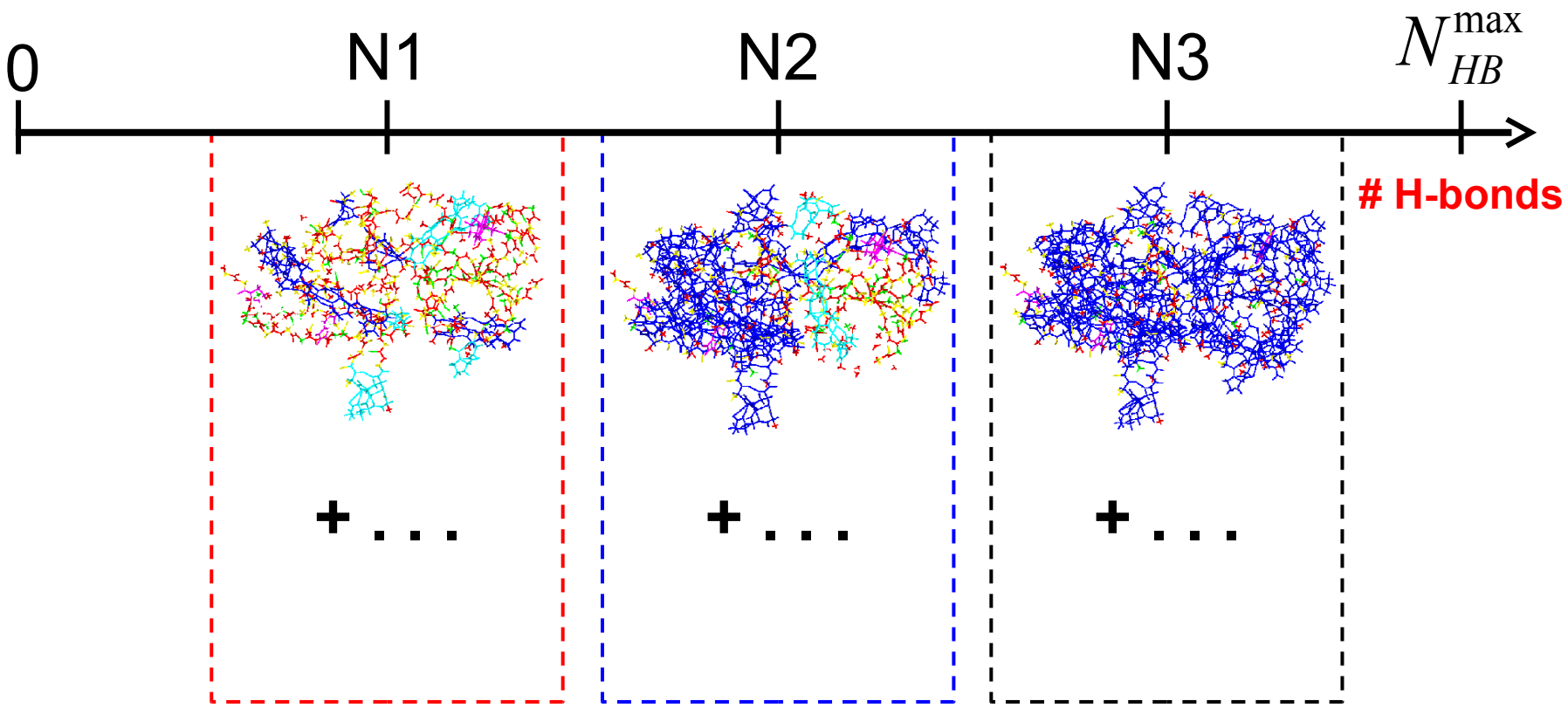


Sub-ensembles of constraint networks

Rigidity properties change depending on number of H-bonds

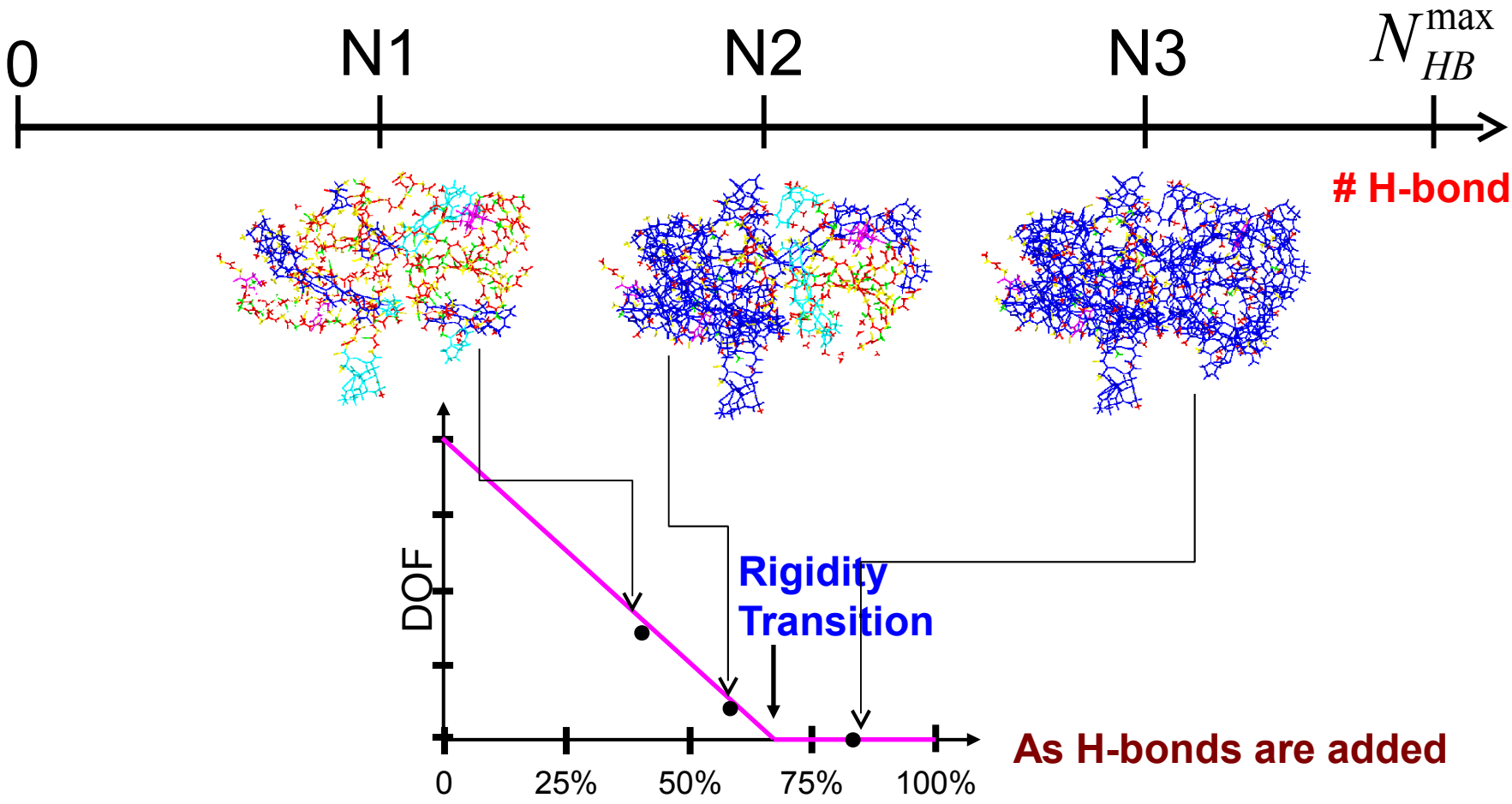
How to estimate average graph rigidity properties in each sub-ensemble?

Method 3: Virtual Pebble Game (estimates all graph rigidity properties)



Maxwell Constraint Counting (MCC)

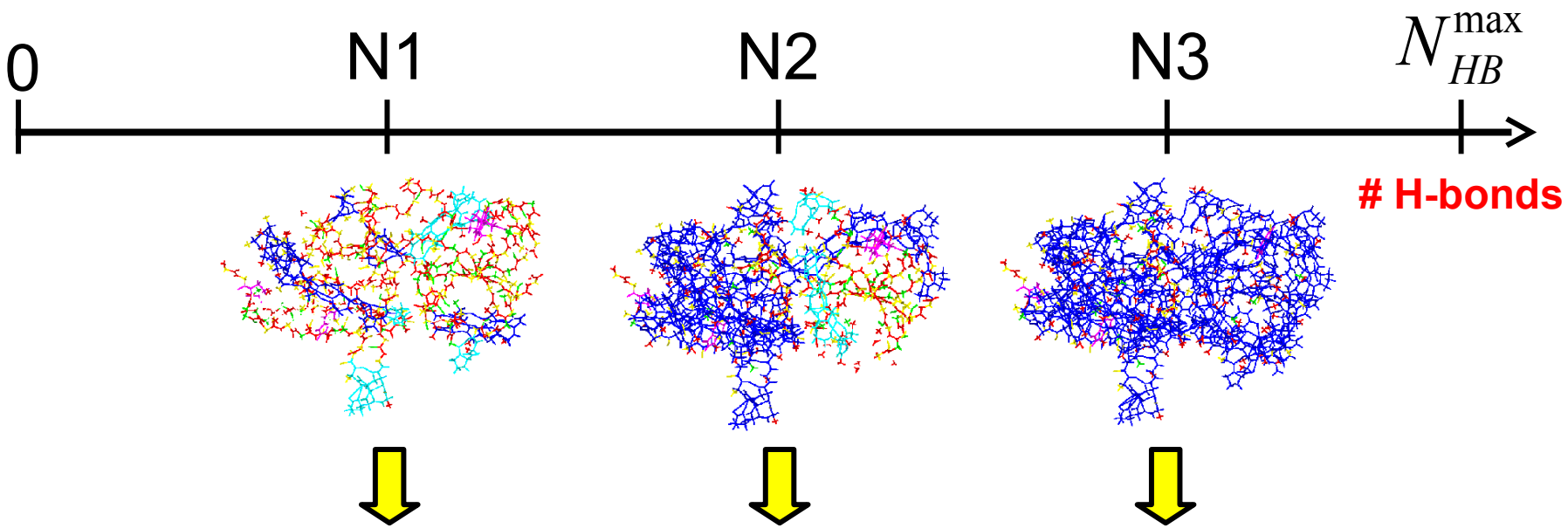
Mean field approximation based on an effective medium approximation



MCC assumes all constraints are independent until the entire network is globally rigid, at which point all additional constraints are redundant.

Maxwell Constraint Counting (MCC)

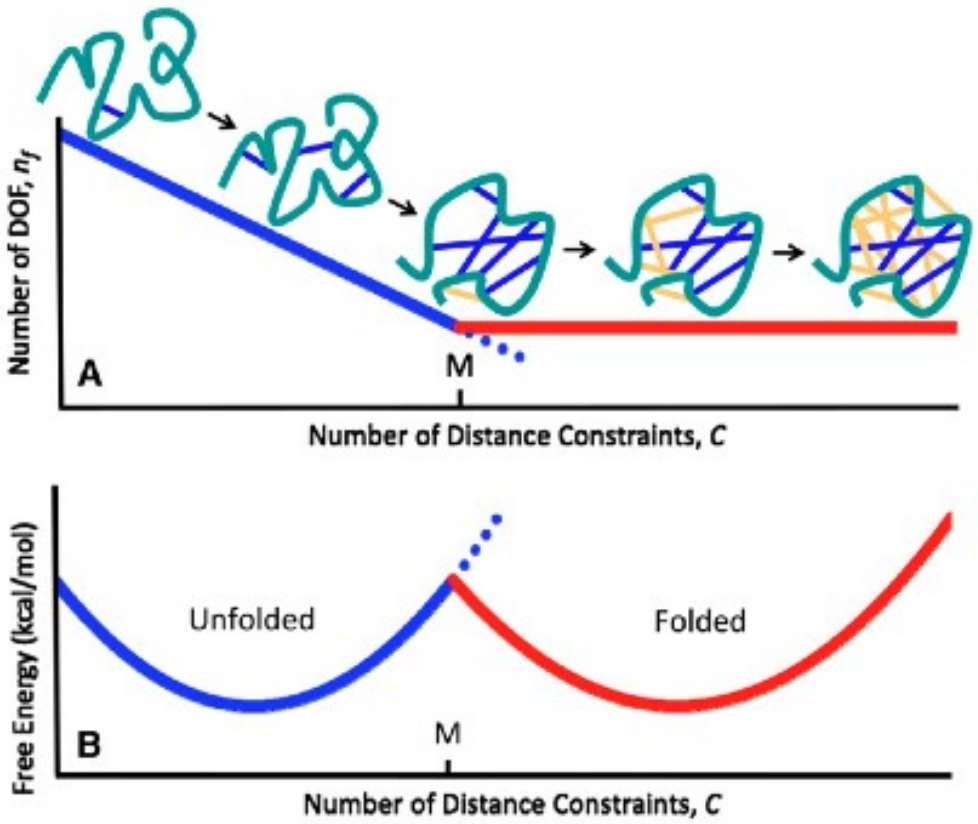
Suppress constraint fluctuations globally



MCC assumes all constraints are independent until the entire network is globally rigid, at which point all additional constraints are redundant. **The mean field approximation defines an effective medium with uniform constraint density.**

Two State Thermodynamics is Captured by MCC

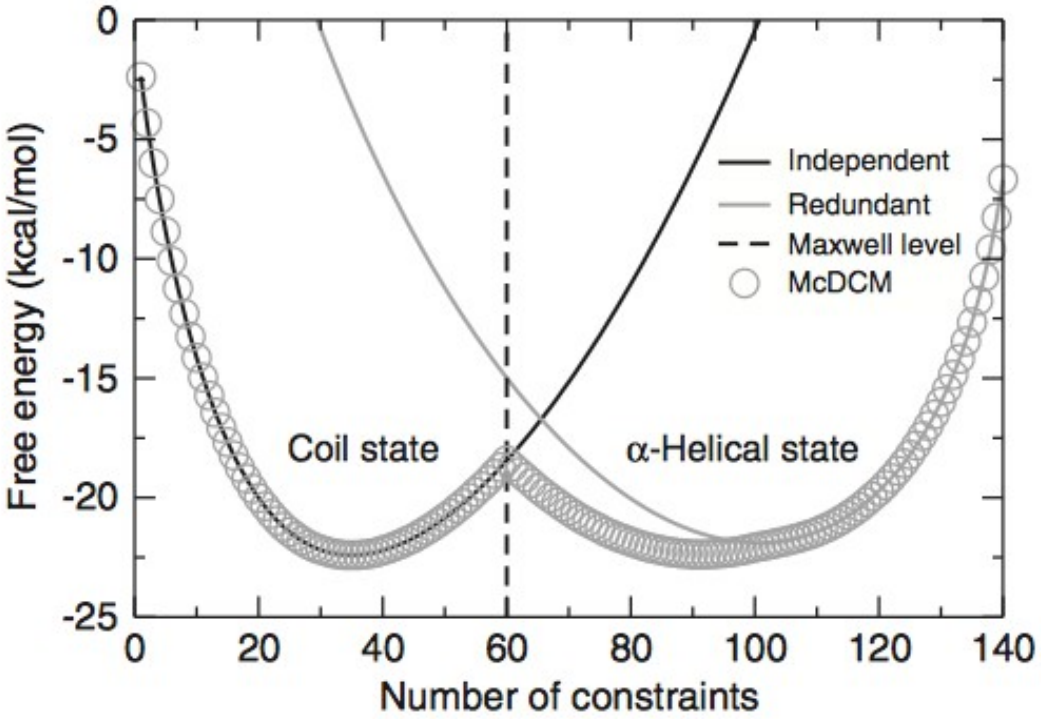
Two extreme basins form



Vorov, Livesay and Jacobs, *Biophysical J.* **100**:1129-38 (2011)

Two State Thermodynamics is Captured by MCC

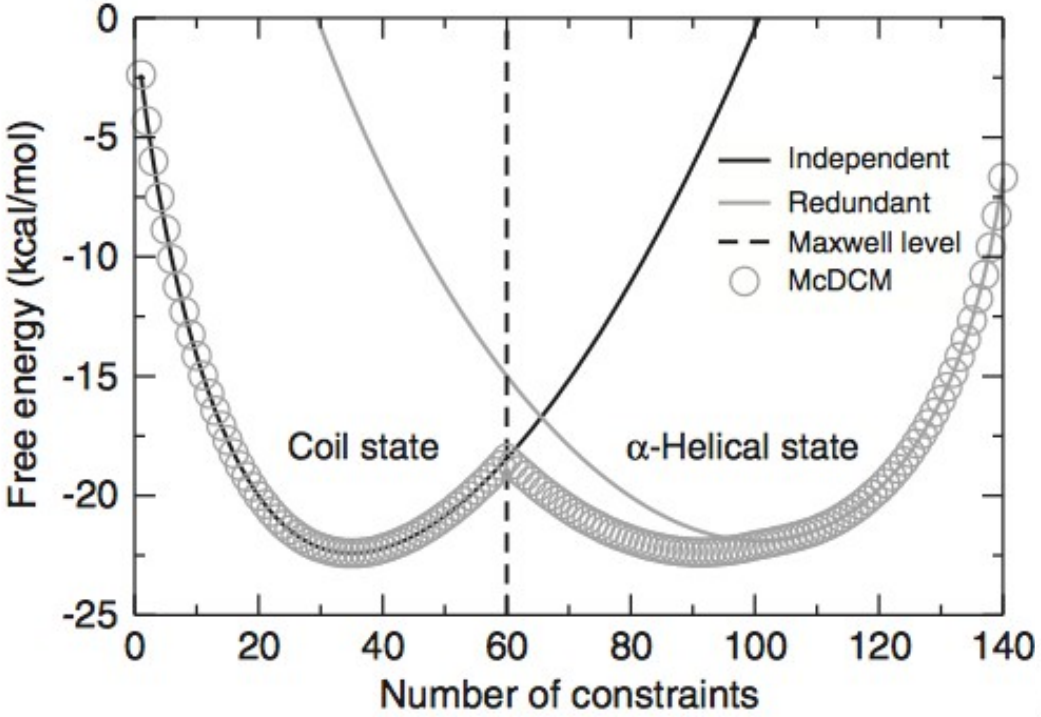
Two extreme basins form

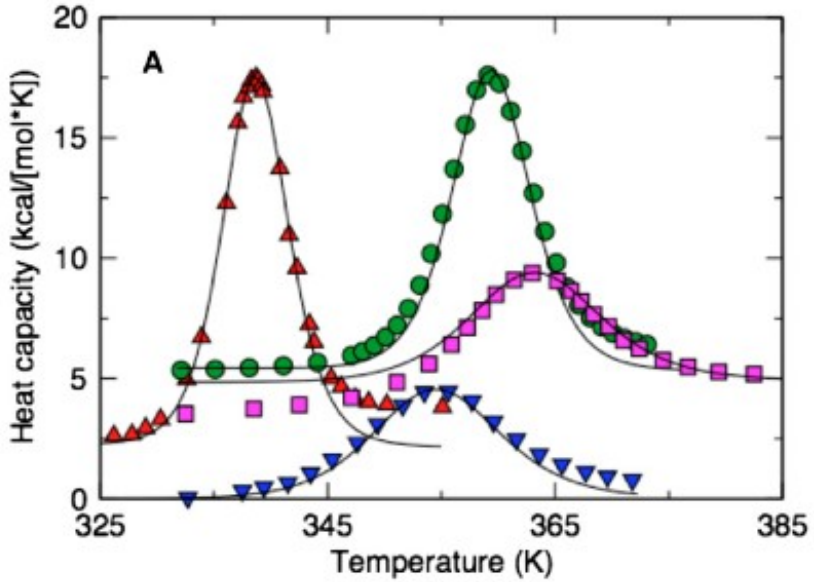


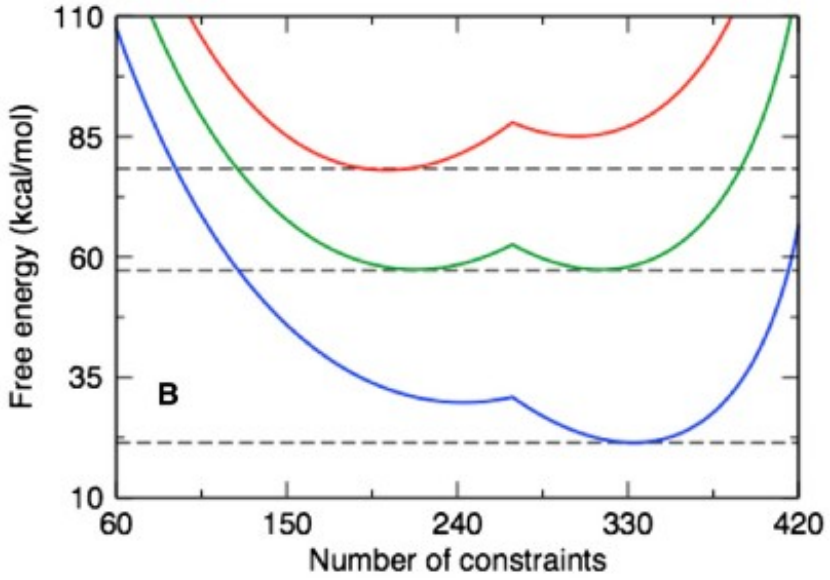
Vorov, Livesay and Jacobs, *Biophysical J.* **97**:3000-09 (2009)

Two State Thermodynamics is Captured by MCC

Two extreme basins form







Vorov, Livesay and Jacobs, *Biophysical J.* **100**:1129-38 (2011)

Vorov, Livesay and Jacobs, *Biophysical J.* **97**:3000-09 (2009)

The Virtual Pebble Game (VPG)

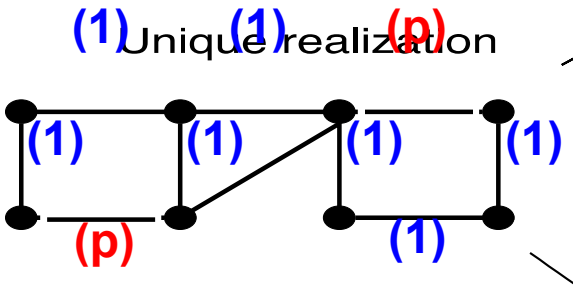
An effective medium approximation applied to fluctuating constraints

Simple example:

(a)

Quenched covalent bond

Fluctuating H-bond



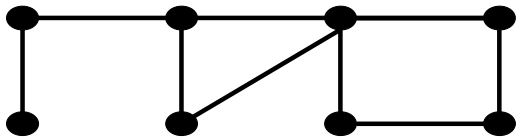
Note: 22 distinct constraint networks

(b)

1 H-bond

Prob = p(1-p)

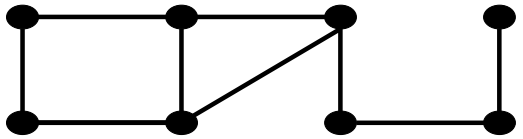
Realization 1



1 H-bond

Realization 2

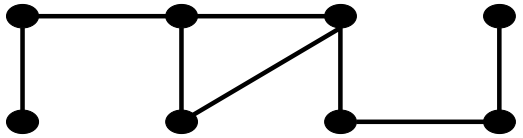
Prob = p(1-p)



0 H-bond

Realization 3

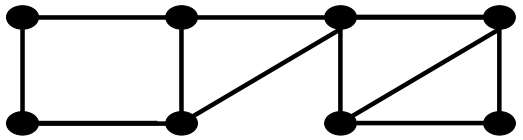
Prob = (1-p)2



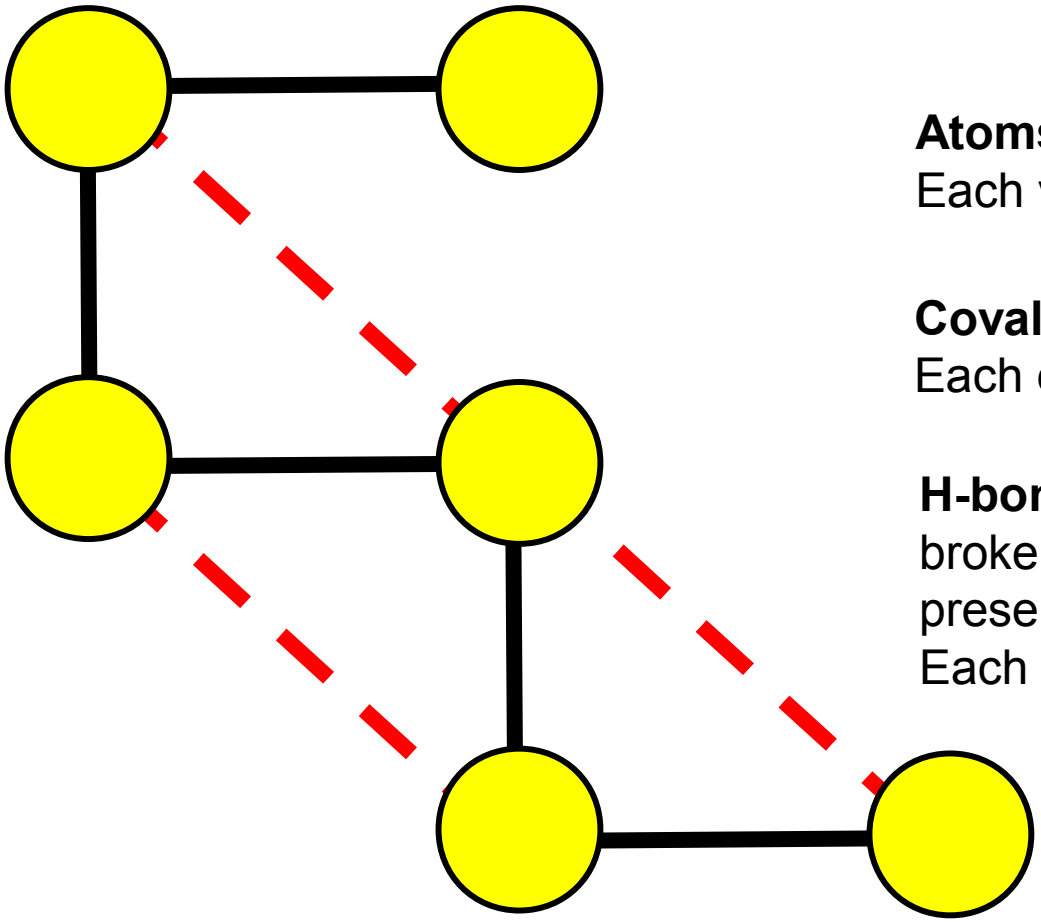
Realization 4

2 H-bond

Prob = p2



Pebble Game (PG)
Body-bar example



Pebble Game Rules
(Body-bar graphs)

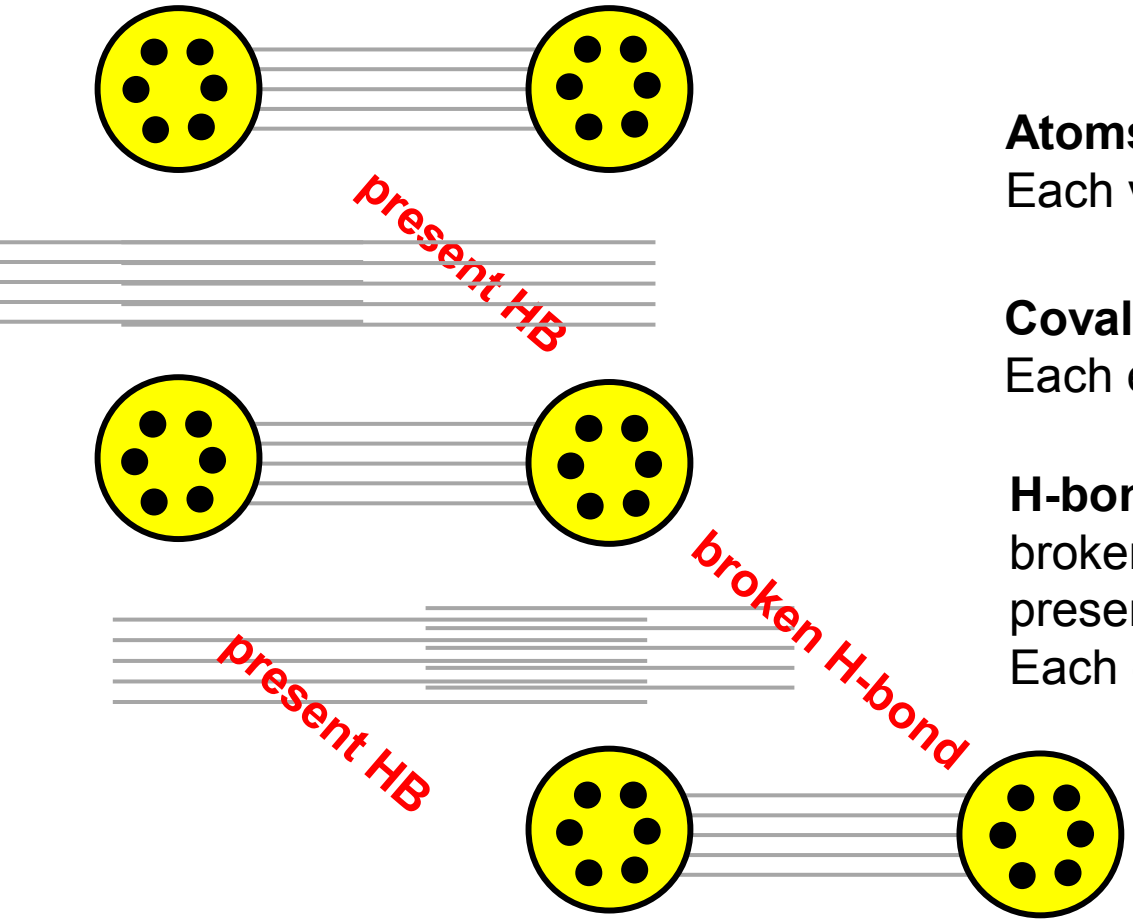
Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

Pebble Game (PG)
Body-bar example

No constraints placed yet.



Pebble Game Rules
(Body-bar graphs)

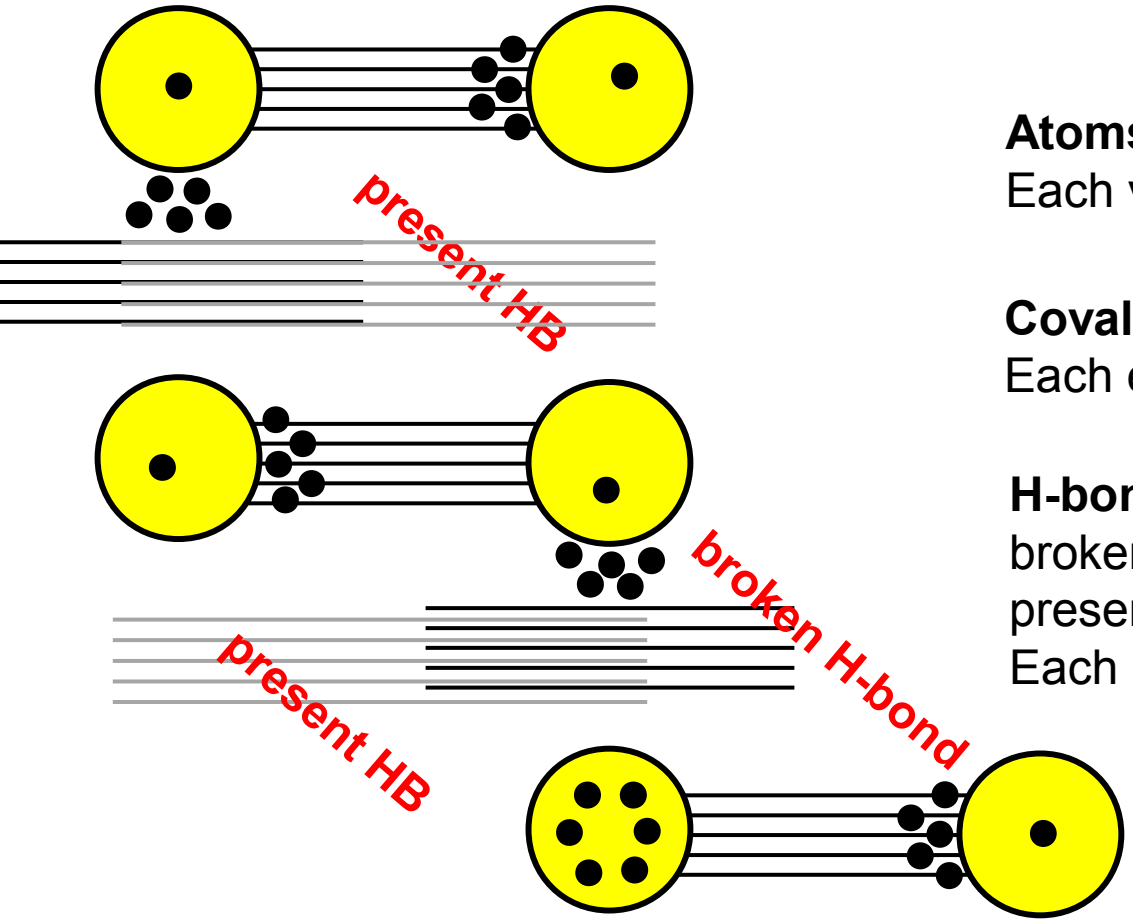
Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

Pebble Game (PG)
Body-bar example

All covalent bonds placed.



Pebble Game Rules
(Body-bar graphs)

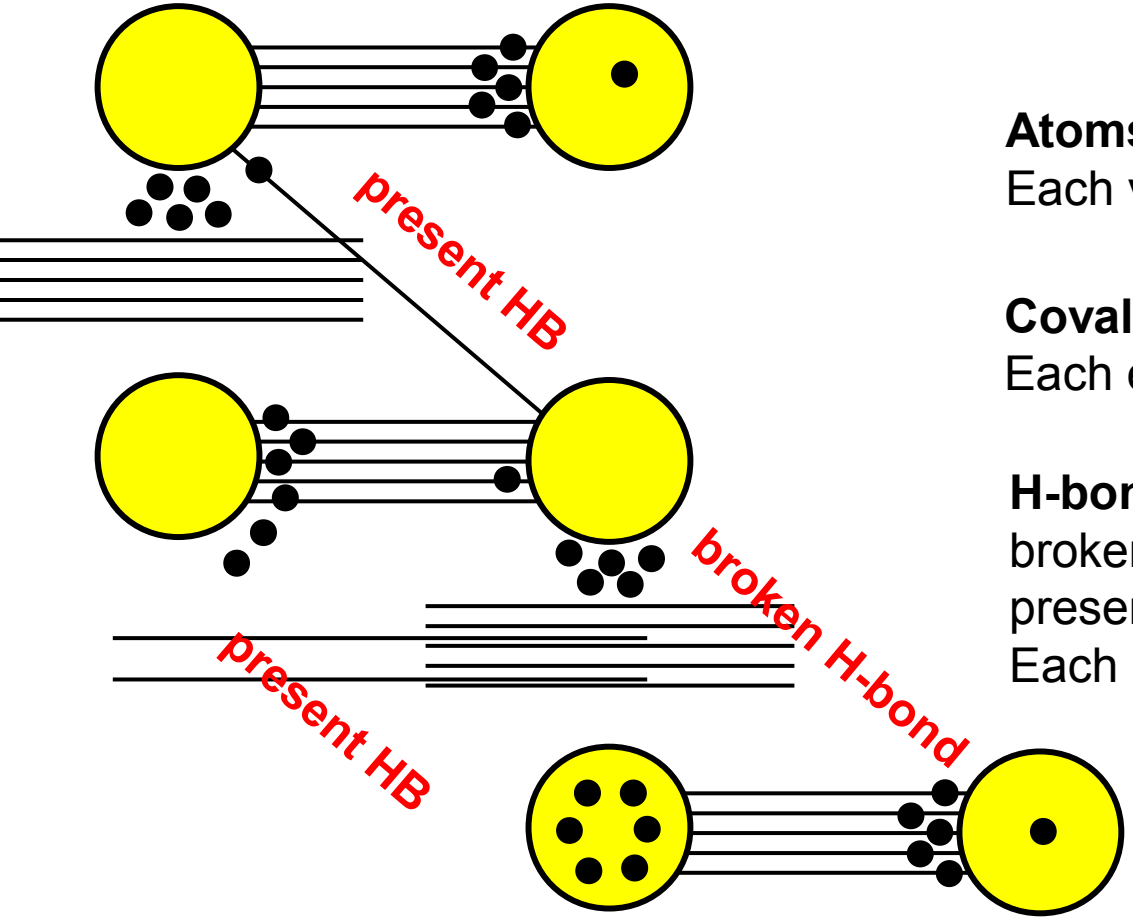
Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

Pebble Game (PG)
Body-bar example

All covalent bonds placed.
Two H-bonds are present and placed.



Pebble Game Rules
(Body-bar graphs)

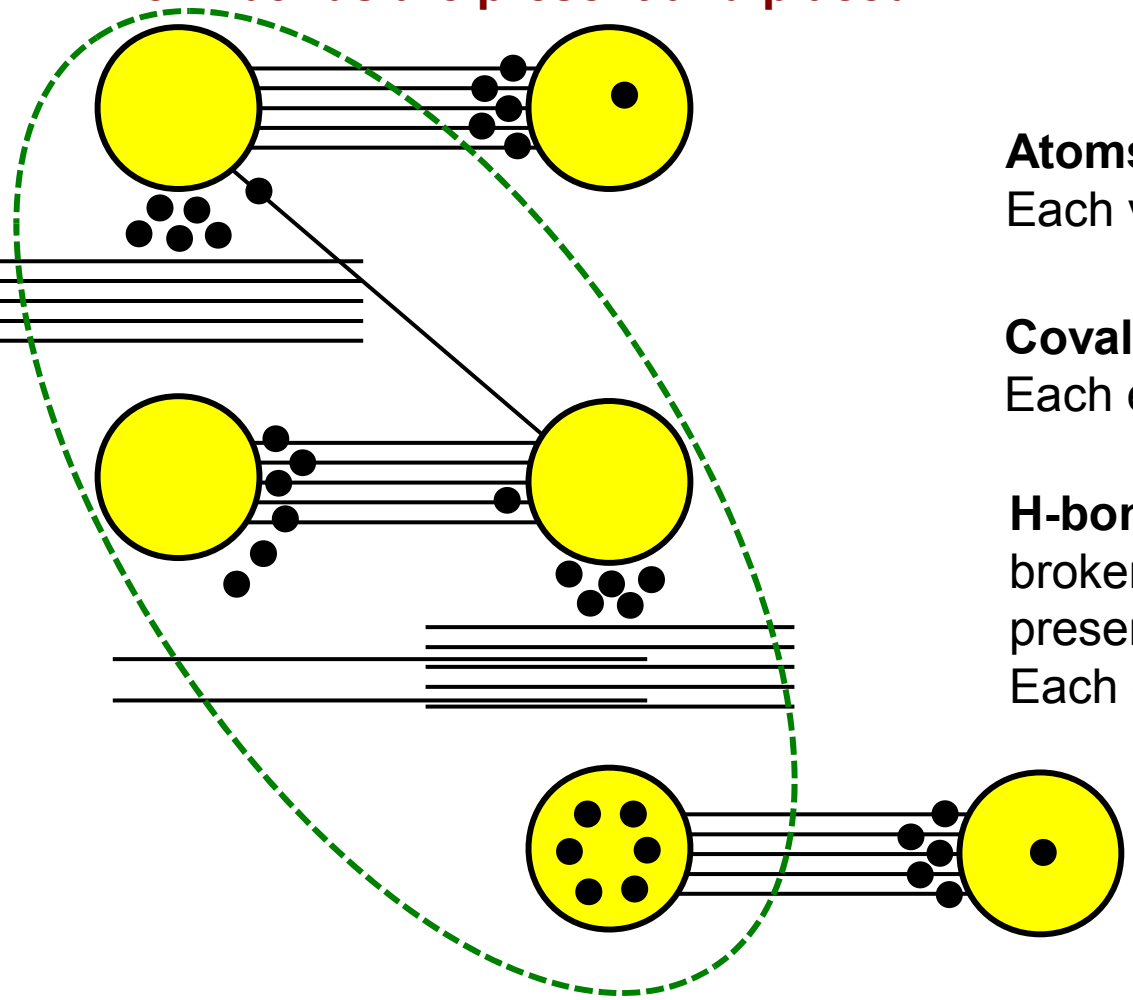
Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

Pebble Game (PG)
Body-bar example

All covalent bonds placed.
Two H-bonds are present and placed.



Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars

The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

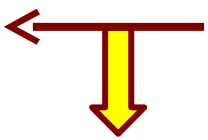
H-bonds
average state □ **5p + 0(1-p)**
Each edge is assigned **5p** bars

Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
broken □ no edge with probability (1-p)
present □ edge with probability p
Each edge is assigned 5 bars



Suppress (ON/OFF) fluctuations within an edge
Retain spatial location of the fluctuating edges.

The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

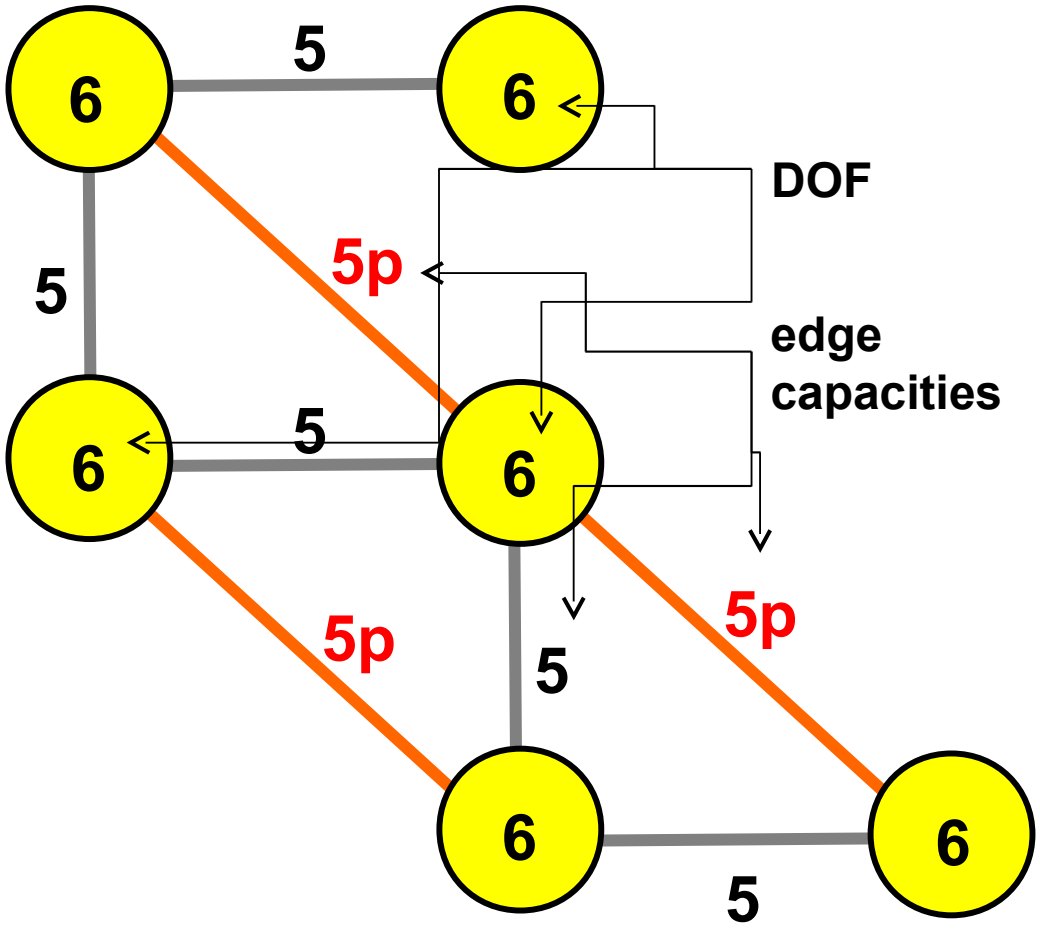
Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
average state □ **5p + 0(1-p)**
Each edge is assigned **5p** bars
Suppress intra-edge fluctuation

No constraints placed yet.



The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

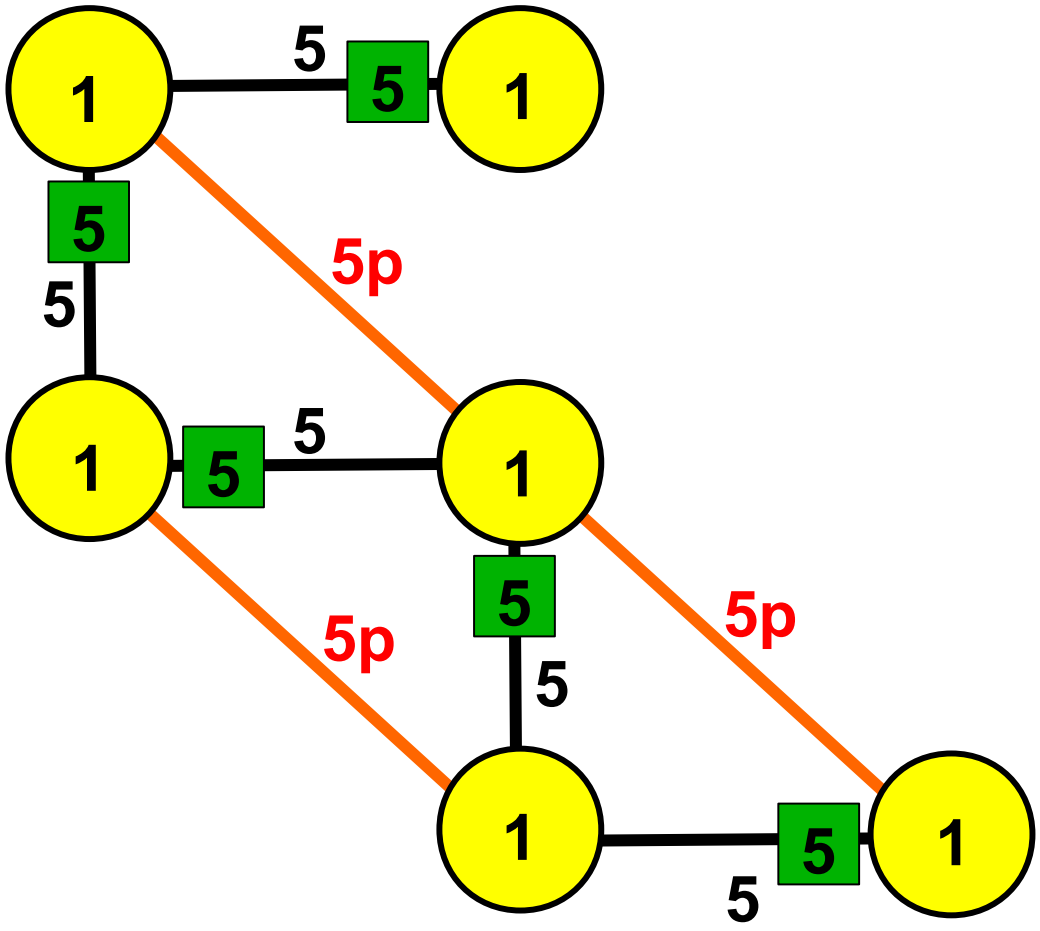
Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
average state □ **5p + 0(1-p)**
Each edge is assigned **5p** bars
Suppress intra-edge fluctuation

All covalent bonds placed.



The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

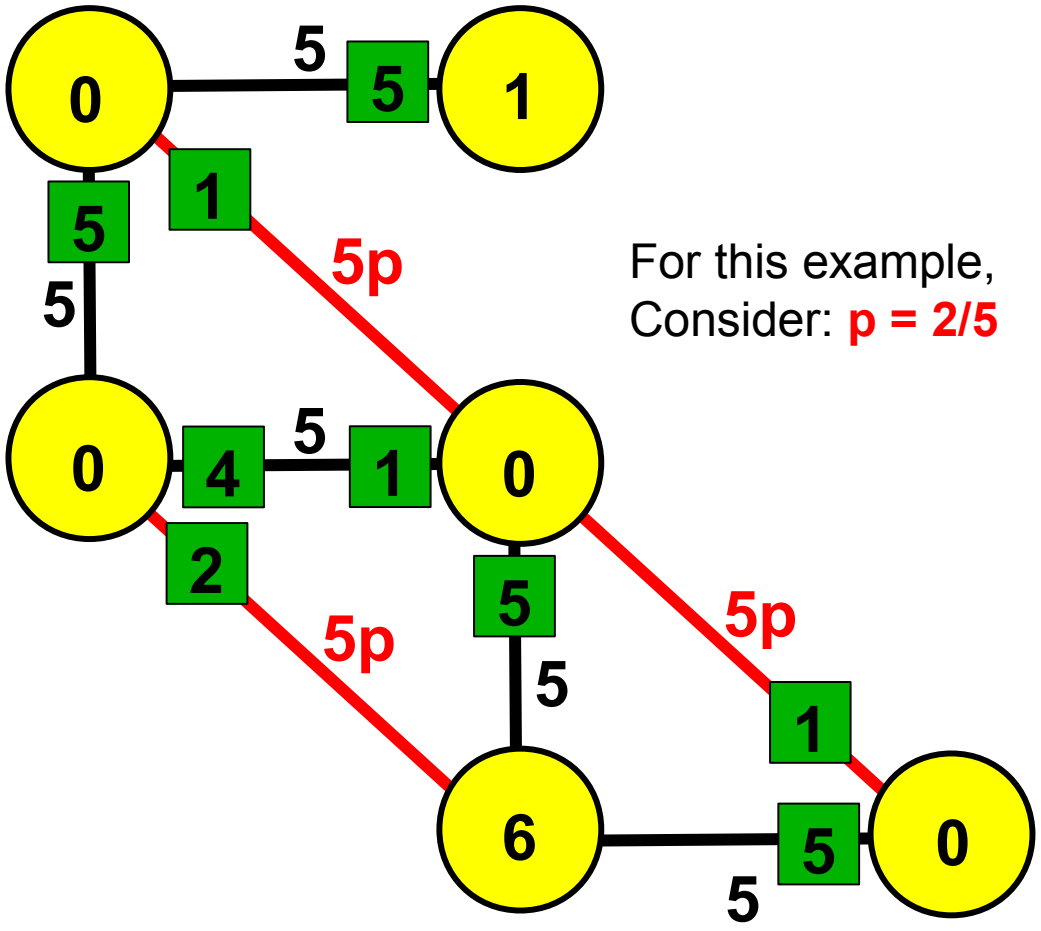
Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
average state □ **5p + 0(1-p)**
Each edge is assigned **5p** bars
Suppress intra-edge fluctuation

All covalent bonds placed.
H-bonds are placed at average capacity.



For this example,
Consider: **p = 2/5**

The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

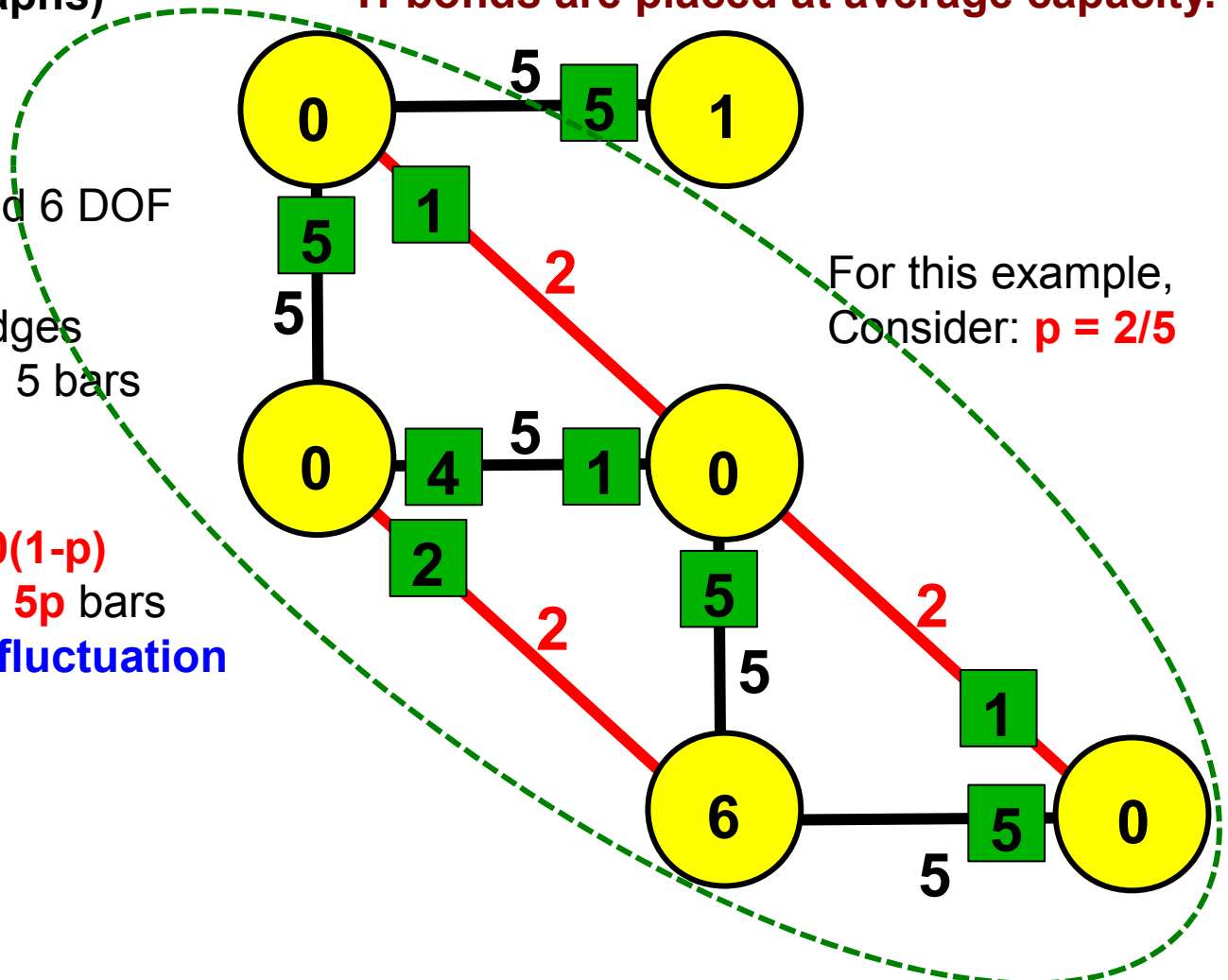
Virtual Pebble Game Rules
(Body-bar graphs)

Atoms □ vertices
Each vertex is assigned 6 DOF

Covalent bonds □ edges
Each edge is assigned 5 bars

H-bonds
average state □ **5p + 0(1-p)**
Each edge is assigned **5p** bars
Suppress intra-edge fluctuation

All covalent bonds placed.
H-bonds are placed at average capacity.



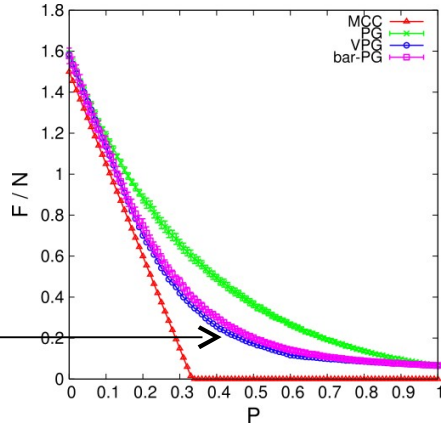
For this example,
Consider: **p = 2/5**

Virtual Pebble Game Results

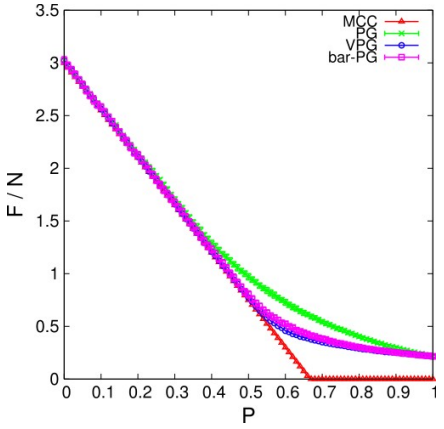
Three dimensional diluted lattices

Showing VPG results for four different lattice models as typical representative examples.

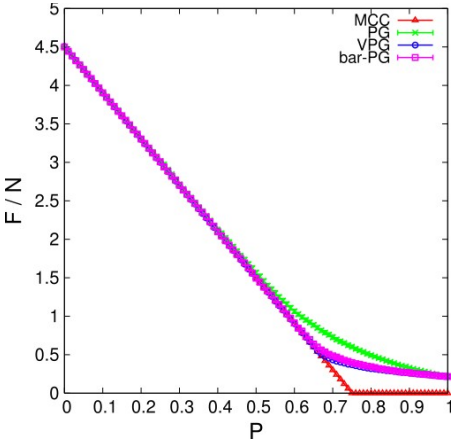
bar-PG considers all 5 bars within an edge as independent so that it is possible to have 0,1,2,3,4,5 constraints in contrast to the cooperative case of (0 or 5).



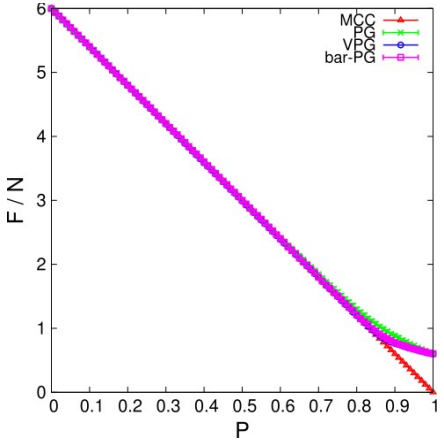
(a) $q_{fluct} = 0.3$, $q_{fix} = 0.3$



(b) $q_{fluct} = 0.3$, $q_{fix} = 0.2$



(c) $q_{fluct} = 0.4$, $q_{fix} = 0.1$

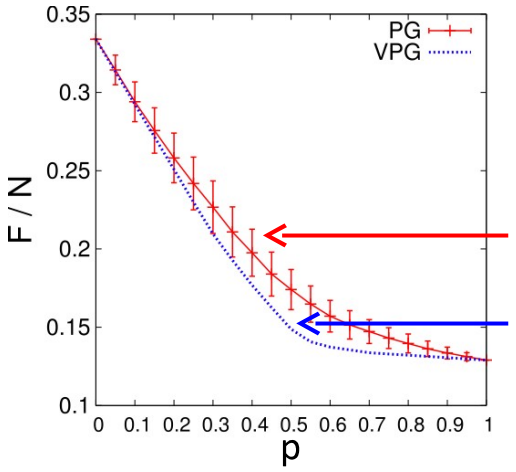


(d) $q_{fluct} = 0.4$, $q_{fix} = 0.0$

Virtual Pebble Game Results

Based on a non-redundant dataset of 272 proteins

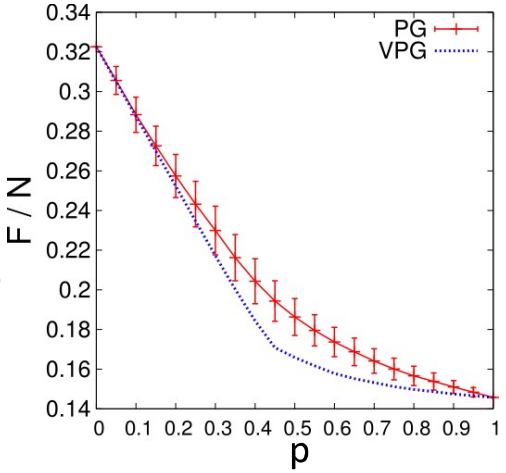
Showing VPG results for four different proteins as typical representative examples



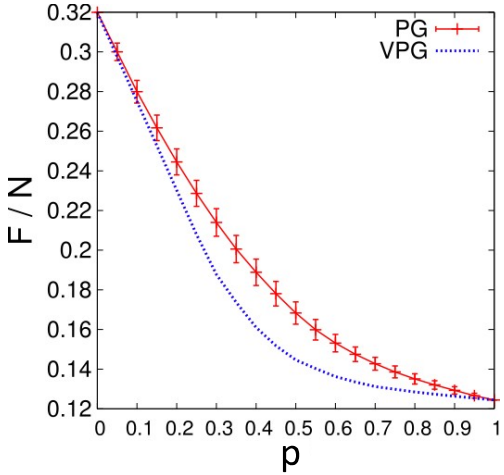
statistical error bars

precise

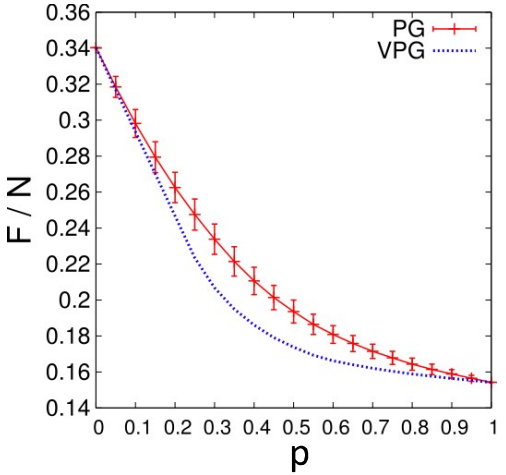
(a) Scorpion protein toxin



(b) Oncogene MTCP-1



(c) FLAP endonuclease

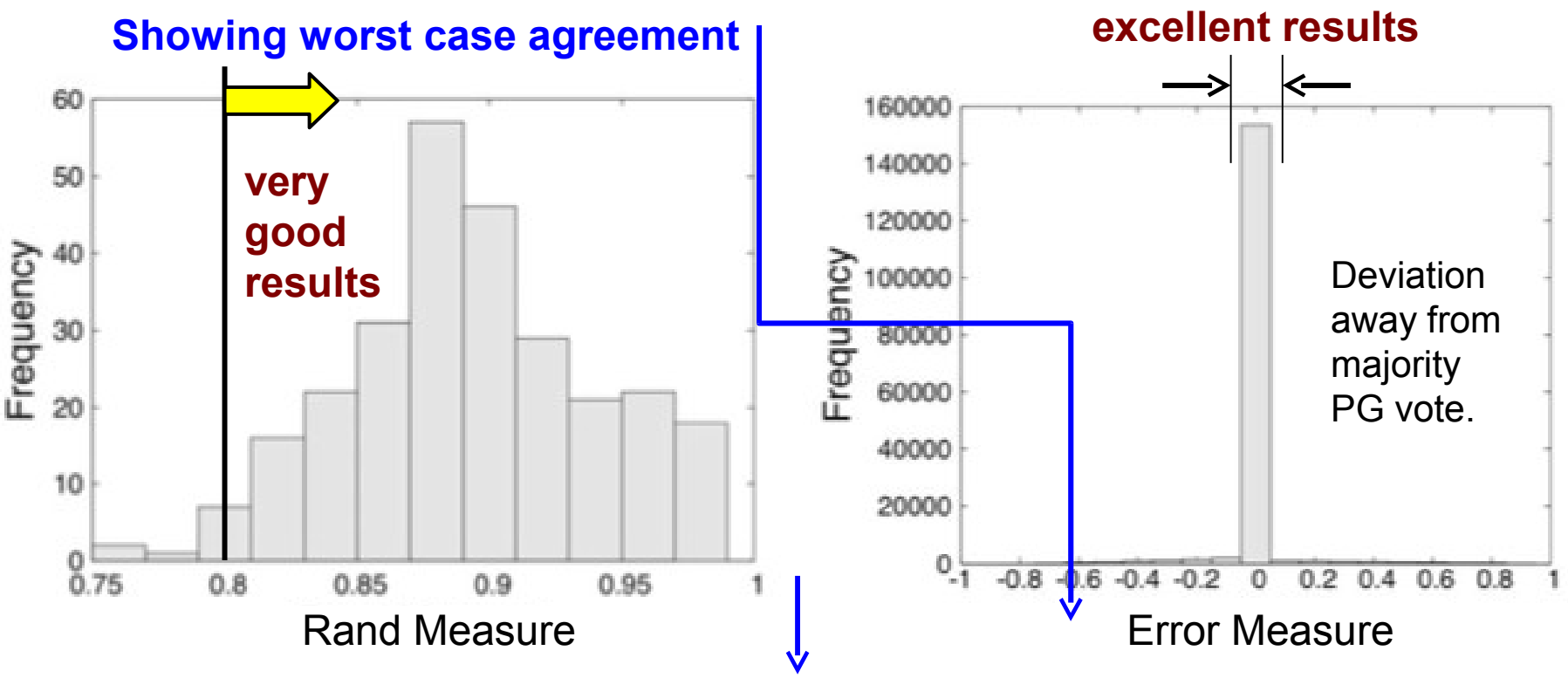


(d) Transcription / DNA

Gonzalez et. al. (2012)
PLoS ONE 7(2): e29176.

Virtual Pebble Game Results

Based on a non-redundant dataset of 272 proteins



For a given protein, the # of H-bonds that lead to the greatest disagreement is used to benchmark the worse case errors in terms of the Rand and error measures

Gonzalez LC, Wang H, Livesay DR, Jacobs DJ (2012) Calculating Ensemble Averaged Descriptions of Protein Rigidity without Sampling. PLoS ONE 7(2): e29176.

Virtual Pebble Game Results

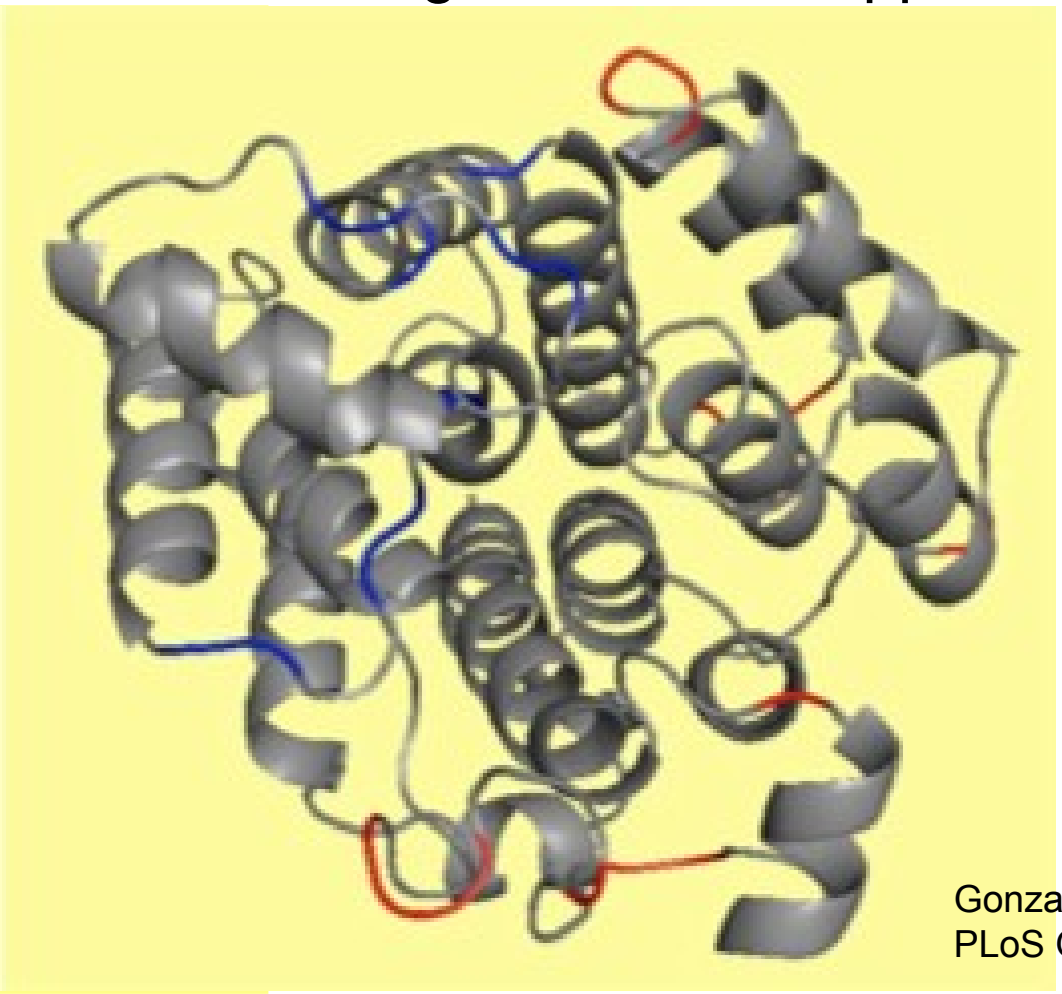
Based on a non-redundant dataset of 272 proteins

Example protein case showing where errors between VPG and average PG results appear.

Blue VPG over estimates rigidity

RED VPG over estimates flexibility

GREY VPG agrees with

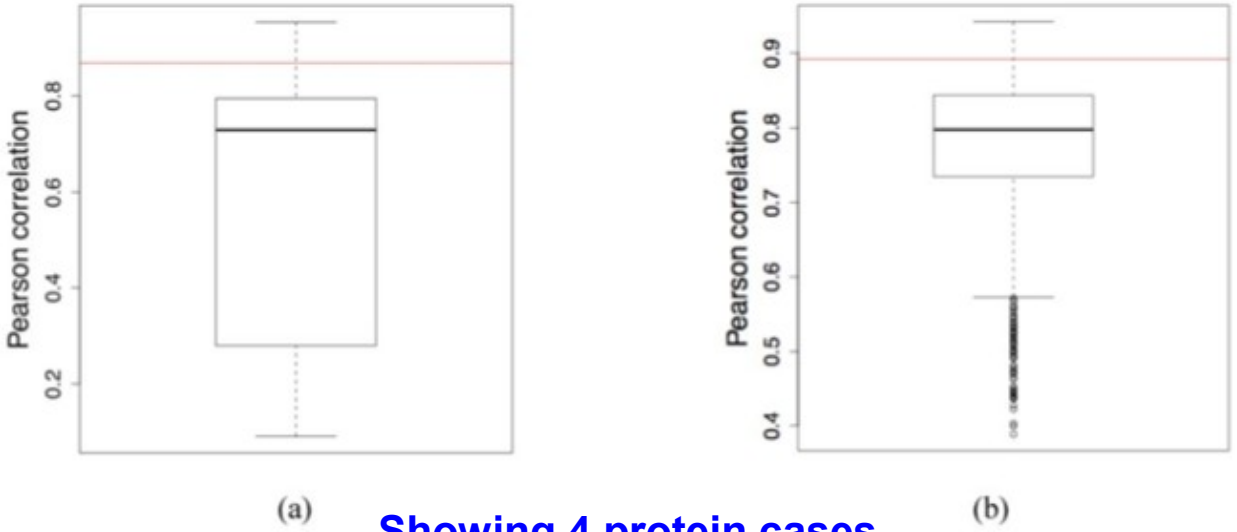


Gonzalez et. al. (2012) PLoS ONE 7(2): e29176.

Virtual Pebble Game Results

Based on a non-redundant dataset of 272 proteins

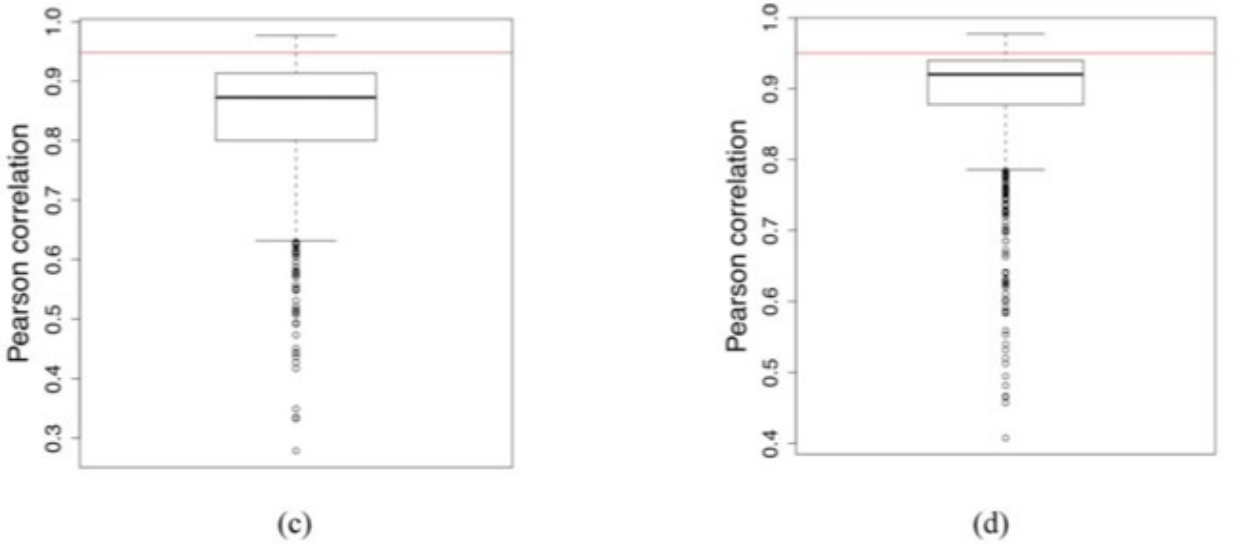
Correlations between residue to residue mechanical couplings.



Showing 4 protein cases

RED LINE

VPG results correlated to average PG results. Box plot created based on correlating all 1000 PG results to the average PG result given by **black line**.



Gonzalez et. al. (2012)
PLoS ONE 7(2): e29176.

Conclusions

Virtual Pebble Game Characteristics

Calculation of the number of DOF in a network is much more accurate than Maxwell Constraint Counting.

Provides detailed information about network rigidity.

Provides precise results without statistical error bars.

Less accurate than sampled averages over 1000 PGs but only requires one run, instead of 1000 runs. Thus, a dramatic speedup, with highest possible precision for the price of little systematic error.

Remark: The VPG provides the necessary precision for advanced free energy calculations in proteins **[1,2]**. The precision and speedup enhances *in silico* high throughput screening applications in protein design and drug discovery.

[1] D.J. Jacobs, *Computer Implemented System for Quantifying Stability and Flexibility Relationships in Macromolecules*, ***U.S. Patent No. 8,244,504*** (2012).

[2] D.J. Jacobs, An Interfacial Thermodynamics Model for Protein Stability, **Biophysics**, Ed: A.N. Misra, Intech publishers, pages 91-132, ISBN 978-953-51-0376-9 (2012).

Open Questions

Empirical observation 1:

The number of DOF predicted by the virtual pebble game is always less than the sampled average number of DOF over a large number of pebble games.

Can it be proved that the VPG gives a lower bound for the population ensemble average DOF from the pebble game?

Open Questions

Empirical observation 1:

The number of DOF predicted by the virtual pebble game is always less than the sampled average number of DOF over a large number of pebble games.

Can it be proved that the VPG gives a lower bound for the population ensemble average DOF from the pebble game?

Empirical observation 2:

The virtual pebble game can over or under predict the degree of rigidity in localized regions compared to sampled ensemble average results using the PG. However, the VPG over predicts rigidity much more frequently, hence the total predicted number of DOF is less.

Can the rigid cluster decomposition or the determination of where DOF are located in a network provide bounds on the respective population ensemble average results?

BMPG

Summer 2013

Newtonian Mechanics

BMPG

Boltzmann Statistics

Collaborators

Dennis Livesay
Irina Nesmolova
Jerry Troutman
Chris Yengo (PSU)
Jörg Rösgen (PSU)

Current Students

Jenny Farmer
Chris Singer
Aaron Brettin
Alex Rosenthal
*Samantha Dodbele

Post Doctoral Fellows

Amit Srivastava

Former:

Research Associates

Oleg Varsy
Jim Mottonen

Former:

Postdoctoral Fellows

Tong Li
Chuanbin Du
Hui Wang
Andrei Istomin
Greg Wood
Sargis Dallakayan

High School student: Sunny Potharaju

Former:

Students

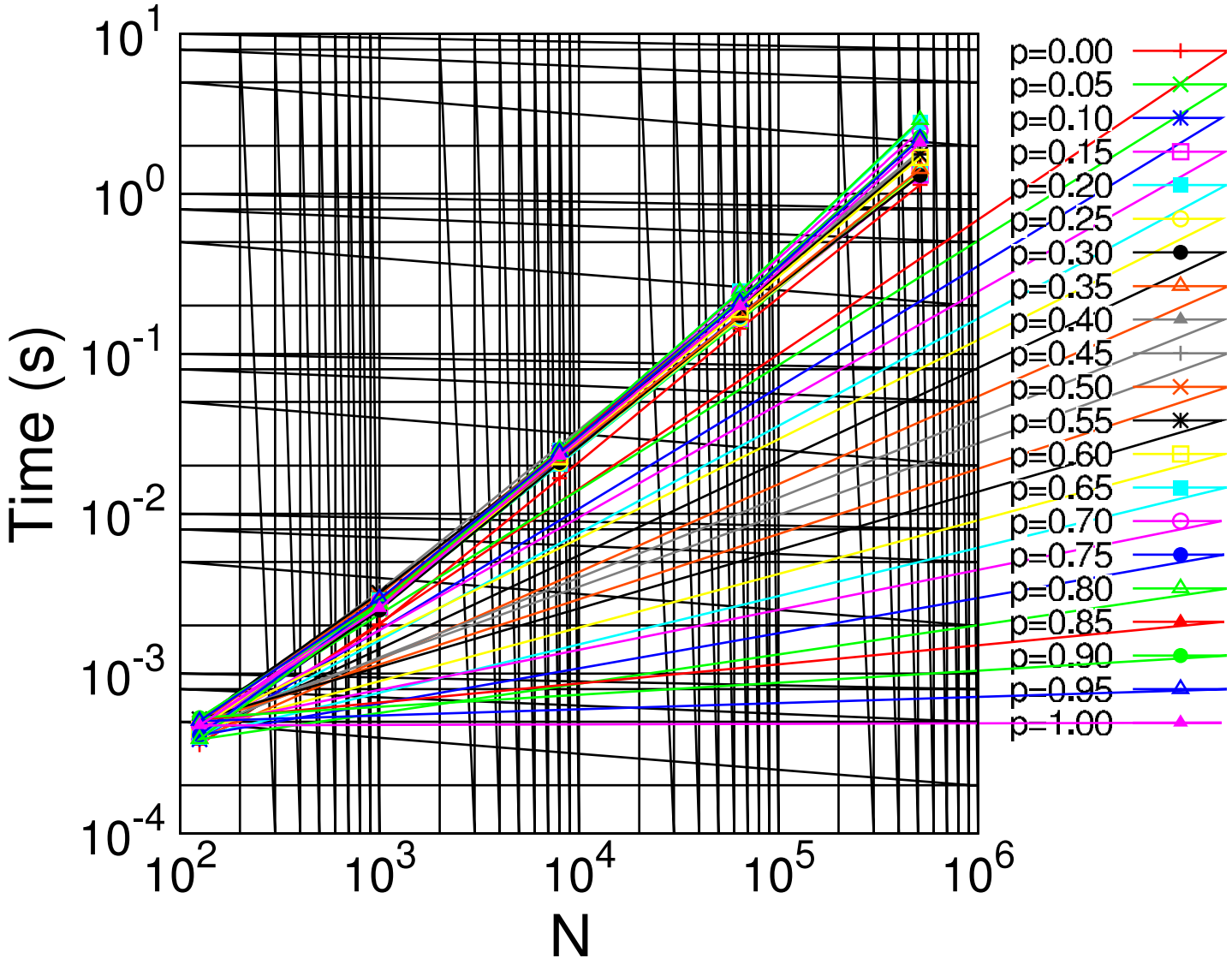
Alicia Heckathorne
Dang Huynh
Jeremy Hules
Moon Lee
Shelley Green
Mike Fairchild
Shira Stav
*Chuck Herrin
*Luis Gonzalez
Charles David
*Deeptak Verma
Wei Song

Work supported by:

NIH R01 GM 073082, S10 SRR026514
MedImmune, Inc.
Charlotte Research Institute (CRI)
Center for Biomedical Engineering and Science (CBES)

Virtual Pebble Game Performance characteristics

VPG scales linearly with size of system at ~20% faster than PG



The VPG is Isomorphic to the PG

Suppress fluctuations at the edge level

Virtual Pebble Game Implementation
(Body-bar graphs)

DOF (or pebbles) are now continuous variables (floats).

Approximate floats using a finite level of precision.

Let 1 be represented using 100,000,000.

All edge capacities are rounded to the nearest 10-8.

Discrete operations are preserved.

Virtual Pebble Game Results

Based on a non-redundant dataset of 272 proteins

Compare VPG rigid cluster decomposition to the ensemble of PG results.

1) Comparison is made against PG results on 1000 independent samples.

2) Use standard **Rand measure** to determine how different two graphs are partitioned into rigid clusters over entire range of H-bond sub-ensembles.

3) For each of the 1000 VPG to PG comparisons, use the number of H-bonds that leads to the **greatest deviation** when calculating the **Error Measure**.

4) Define **Error Measure**, EM, over all PHI and PSI backbone edges.
For each PHI and PSI backbone edge,

EM = 0 if VPG agrees with the majority of the 1000 PG results.
otherwise disagreement with majority implies wrong prediction.

EM = – (Nwrong – Nagree)/Ntotal when VPG predicts edge as part of a rigid cluster.

EM = + (Nwrong – Nagree)/Ntotal when VPG predicts edge as being flexible.

5) Combine all the worse case EM values found for all backbone edges over all 272 proteins, and plot these as a histogram.

minimal Distance Constraint Model (mDCM)

Macrostates are defined within constraint space

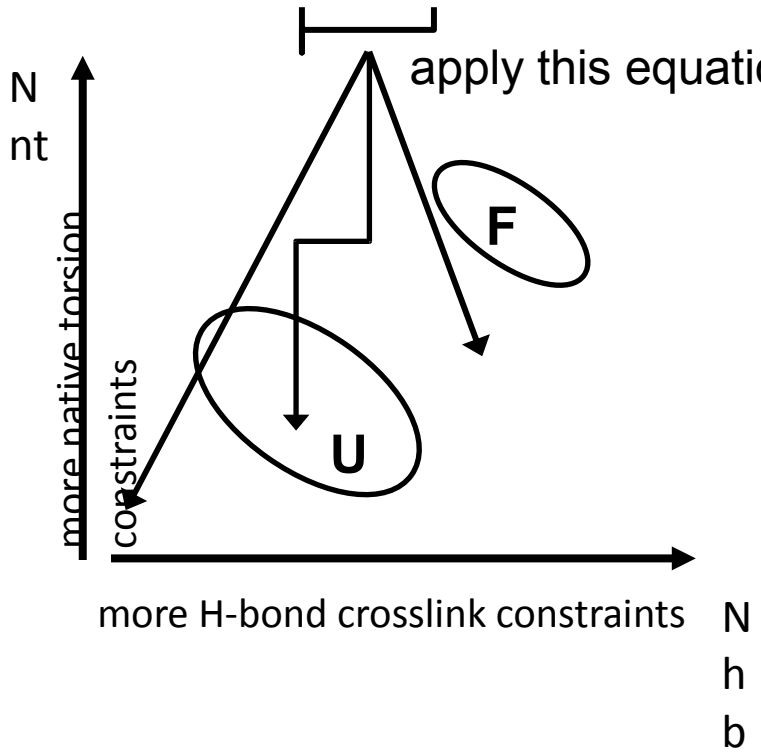
Statistical Mechanical Model

Jacobs & Dallakyan (2005) *Biophysical J.* 88:903

Livesay et al. (2004) *FEBS Letters* 576:468

Free energy function: G(Nhb,Nnt) = UIHB – u Nhb + v Nnt – T(Sc(nat) + Smix)

for given template structure



better atomic packing

more stable secondary structure

FEL and Ensembles of Constraint Topologies

Macrostates are defined within constraint space

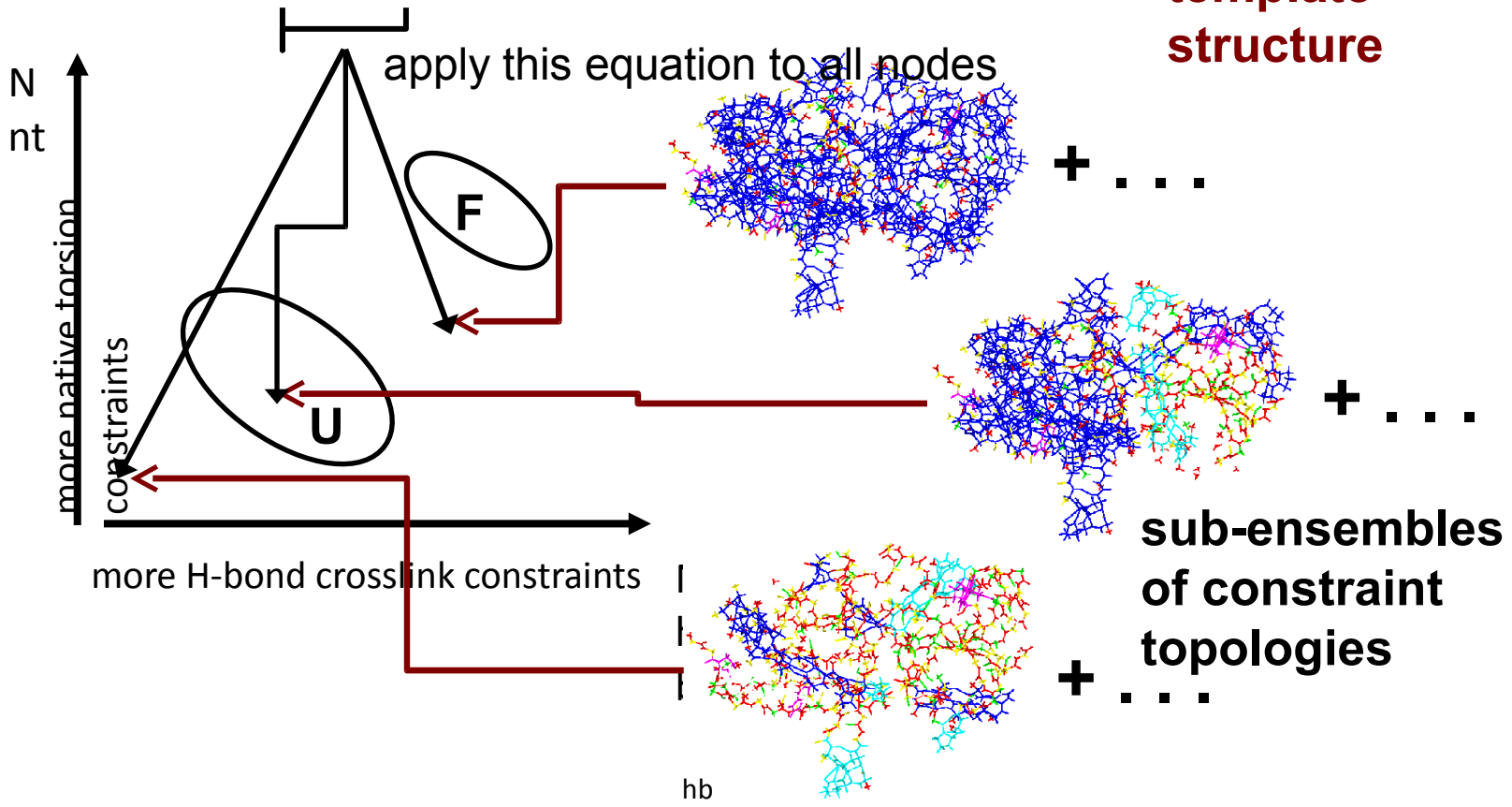
Statistical Mechanical Model

Jacobs & Dallakyan (2005) *Biophysical J.* 88:903

Livesay et al. (2004) *FEBS Letters* 576:468

Free energy function: G(Nhb,Nnt) = UIHB – u Nhb + v Nnt – T(Sc(nat) + Smix)

for given template structure



sub-ensembles of constraint topologies

better atomic packing

more stable secondary structure

Protein Thermodynamics

Empirical model with 3 fitting parameters: mDCM

Statistical Mechanical Model

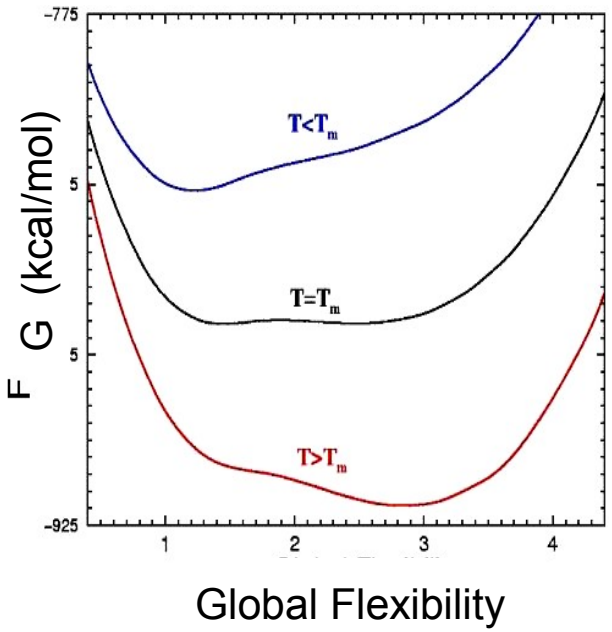
Jacobs & Dallakyan (2005) *Biophysical J.* 88:903
Livesay et al. (2004) *FEBS Letters* 576:468

Free energy function: G(Nhb,Nnt) = UIHB – u Nhb + v Nnt – T(Sc(□nat) + Smix)

Experimental input data
NATIVE STRUCTURE + HEAT CAPACITY

→

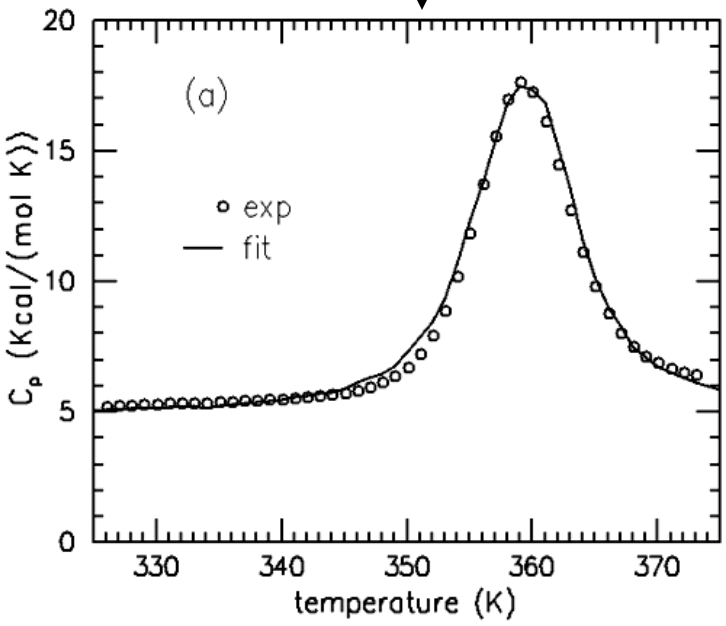
Simulated annealing over
u, v, □nat



←

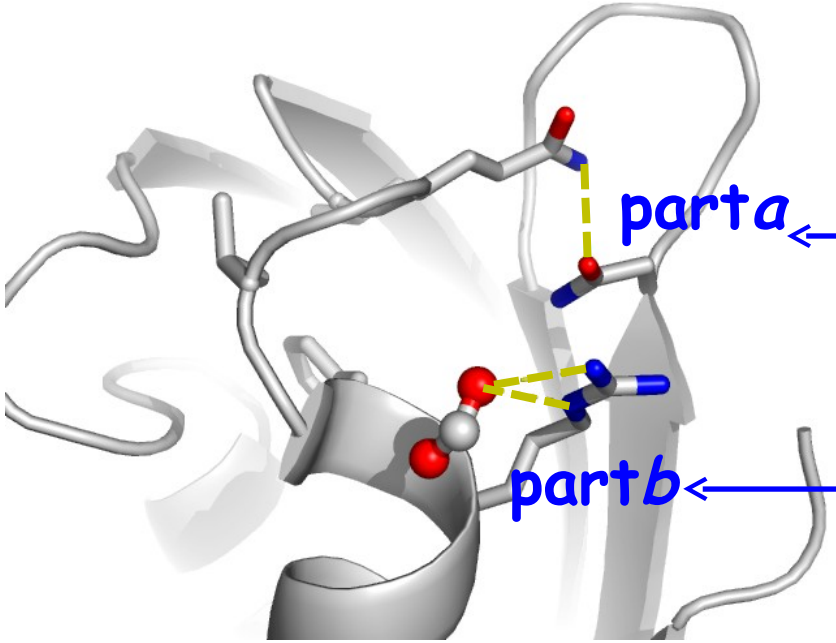
DCM
OUTPUT

includes
FREE ENERGY
LANDSCAPES



Free Energy Decomposition and its Reconstitution

Non-additivity derives from molecular cooperativity



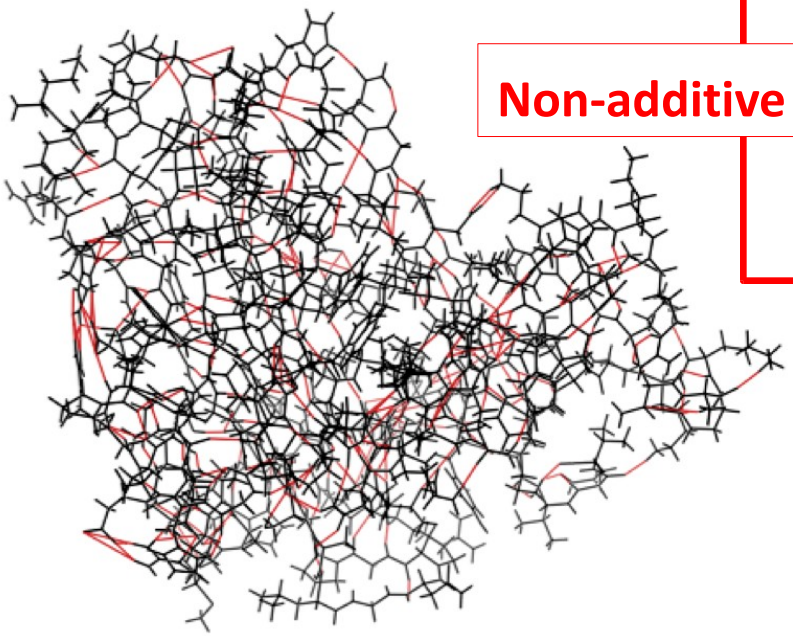
Ledger of free energy contributions from specific interactions

$$G^{FED}_{part}$$

Free Energy Decomposition and its Reconstitution

Non-additivity derives from molecular cooperativity

Free Energy Reconstitution (FER)



Non-additive parts

$$G_{total} \neq \sum G_{part}^{FED}$$

Hidden Thermodynamics
Prevents additive rules from reproducing macromolecular free energies.

Hidden Thermodynamics of Mutant Proteins, Gao, Kuczera, Tidor and Karplus, Science 244, 1069 (1989).

Additivity Principles in Biochemistry,

Dill, J Biol Chem 272, 701-704 (1997).

Decomposition of the Free Energy of a System in Terms of Specific Interactions, Mark and van Gunsteren, *J Mol Biol* 240, 167 (1994).

"In regard to the detailed separation of free energy components, we must acknowledge that the hidden thermodynamics of a protein will, unfortunately, remain hidden."