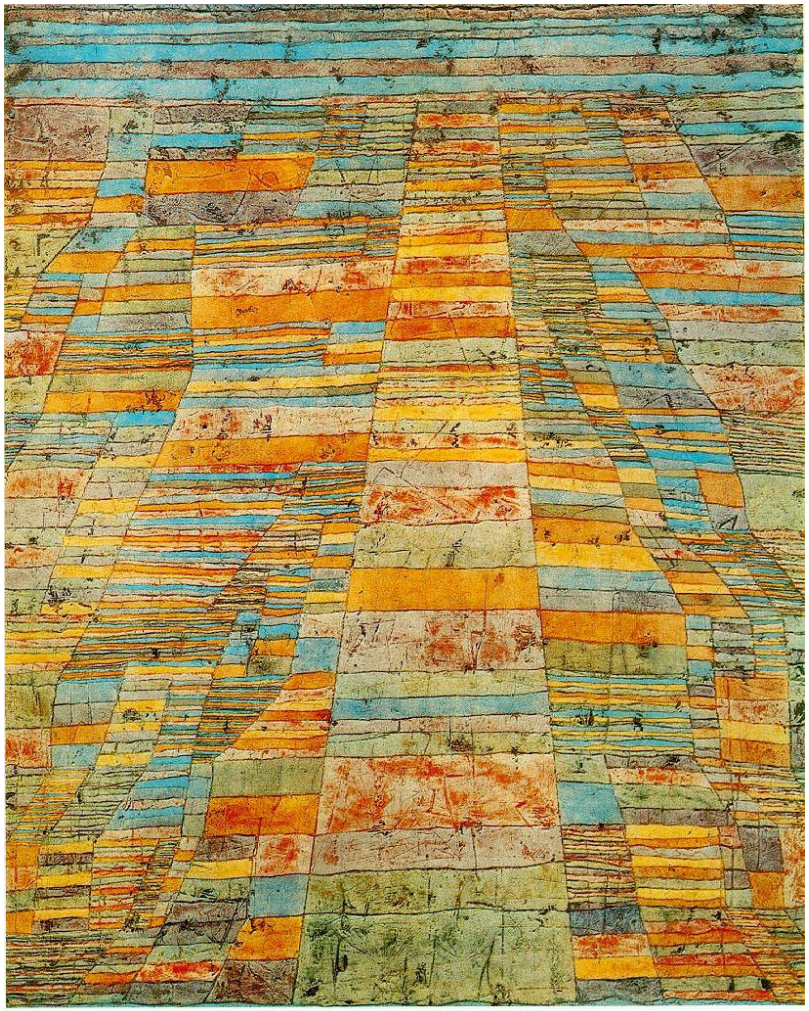
Highway and Byways

Paul Klee, 1929



Coarse-grained Models

Workshop

SU

October 1, 2009

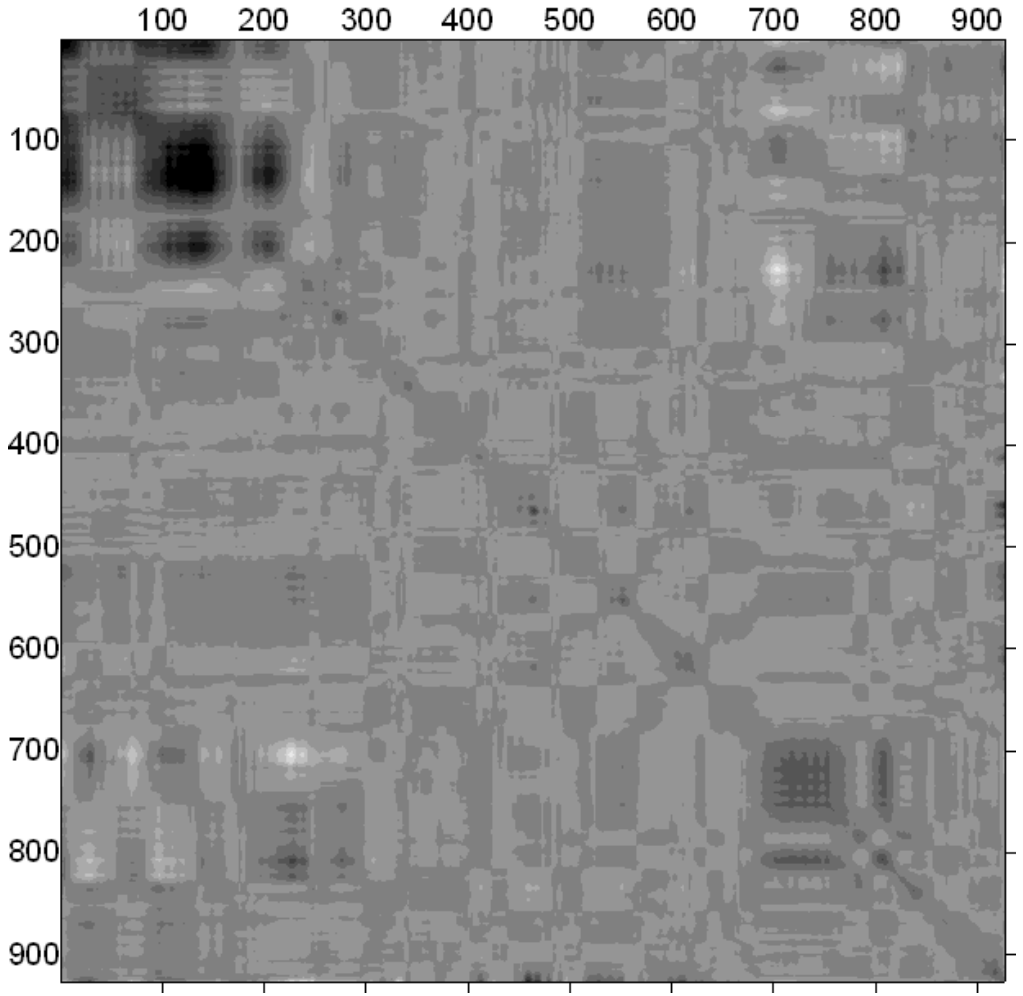
AR Atilgan, D Turgut, OB Okan, AO Aykut, I Inanc, M Mulayim, and C Atilgan

Sabanci University, Istanbul

haemophilus influenzae *ferric-binding protein*

(**1D9V** – unbound; **1MRP** – bound)

Ikeguchi *et al.*, *Phys Rev Lett* ***94*** 078102 (2005)

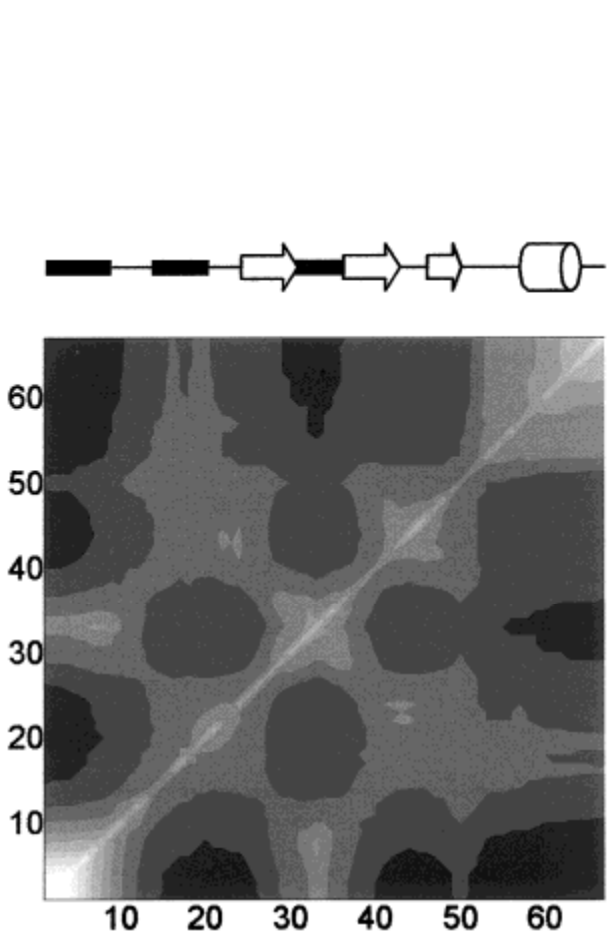


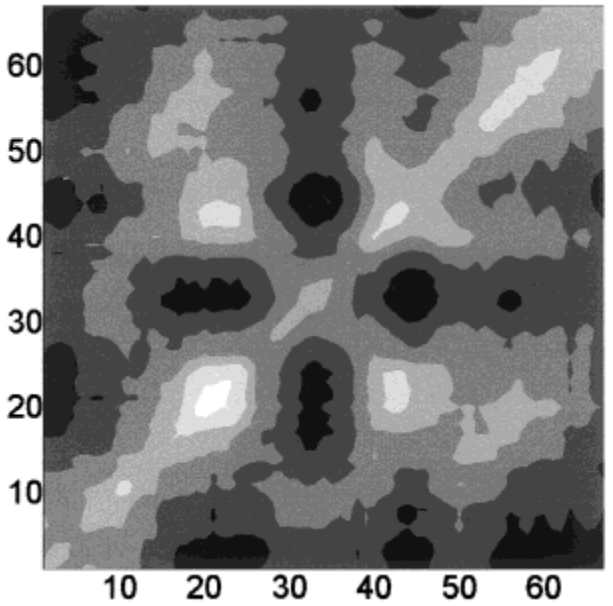
10 ns long **NPT ensemble runs**

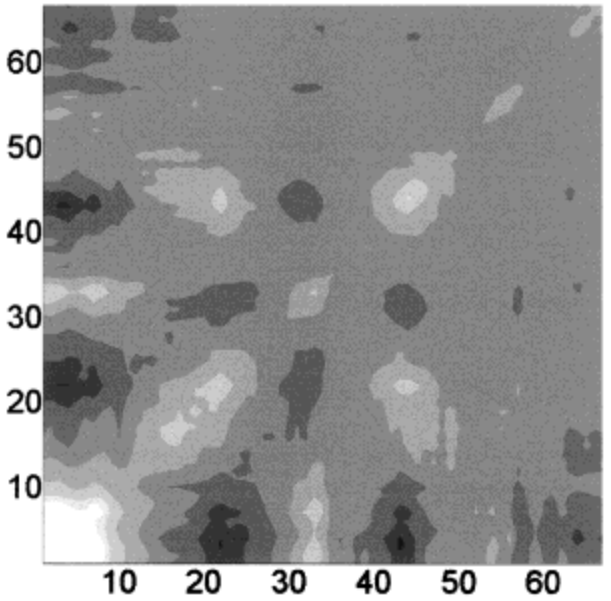
The coordinate sets at ***2*** **ps** intervals

Averages are taken over the

T = 5000 recorded snapshots







PROTEINS: Structure, Function, and Genetics 43:150–160 (2001)

Elucidating the Structural Mechanisms for Biological Activity of the Chemokine Family

Canan Baysal[1]* and Ali Rana Atilgan[2]
[1]*Faculty of Engineering and Natural Sciences, Sabanci University, Tuzla, Istanbul, Turkey*
[2]*Polymer Research Center and School of Engineering, Bogazici University, Bebek, Istanbul, Turkey*

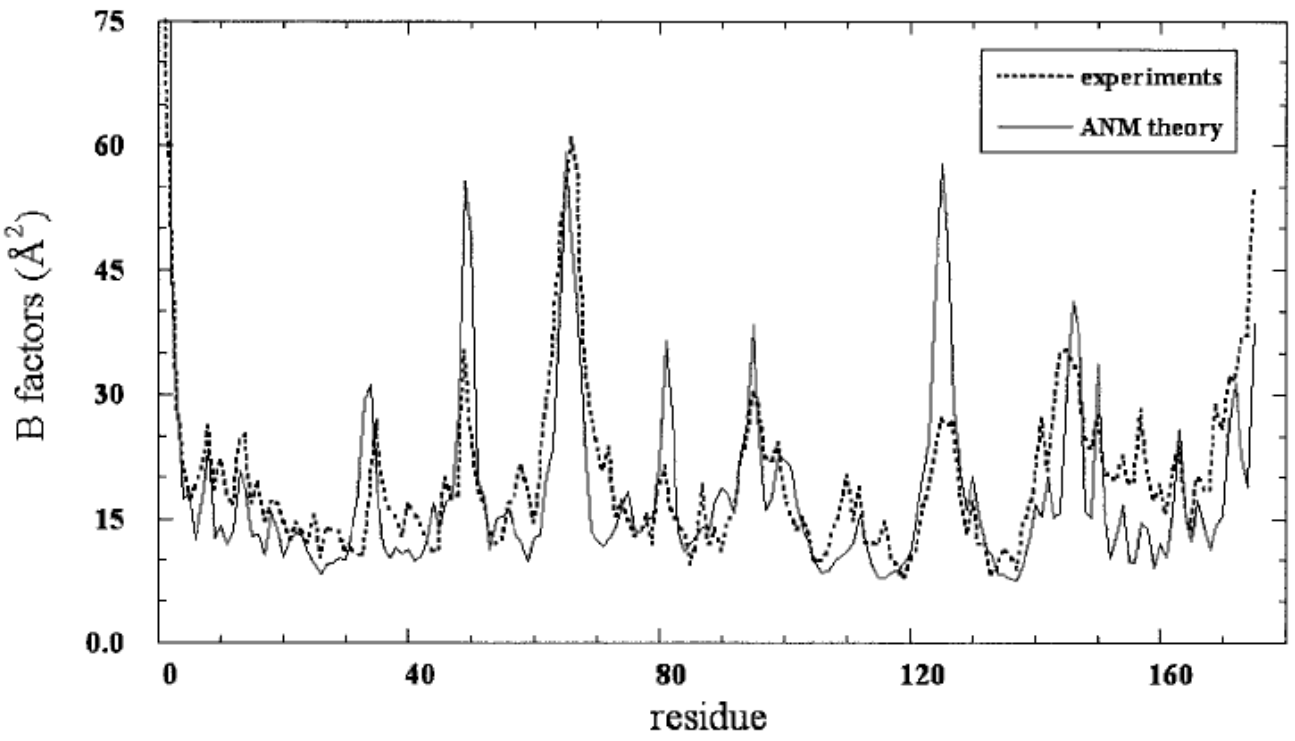


FIGURE 5 Comparison of experimental (Zanotti et al., 1998) and theoretical (ANM) temperature factors obtained for RBP α-carbons, as a function of residue index.

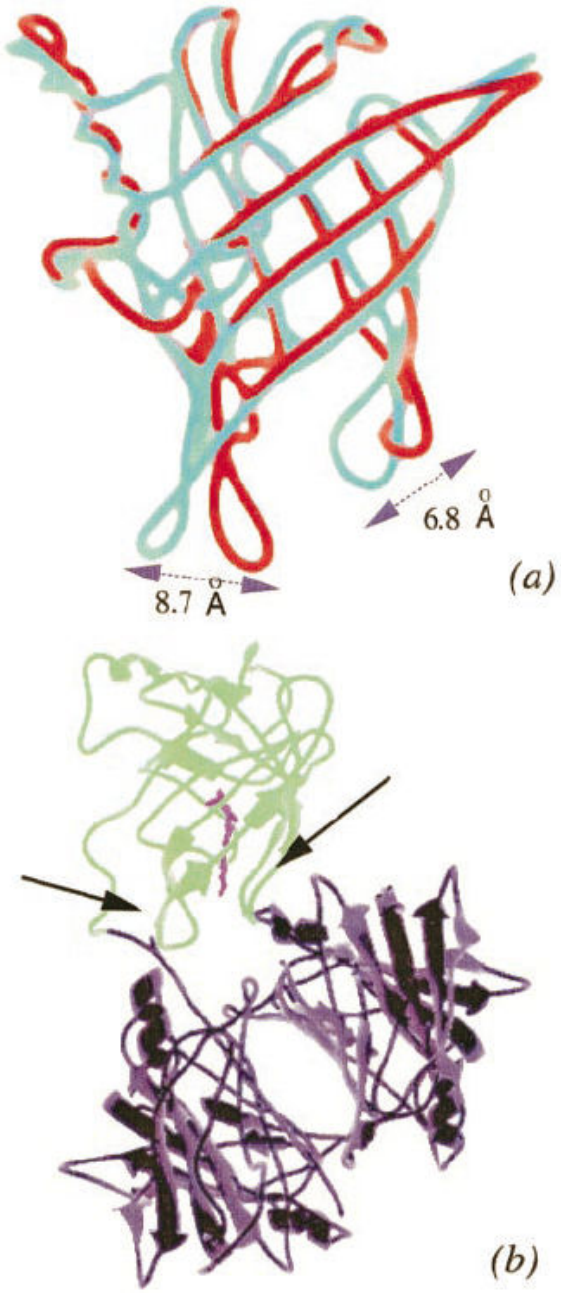
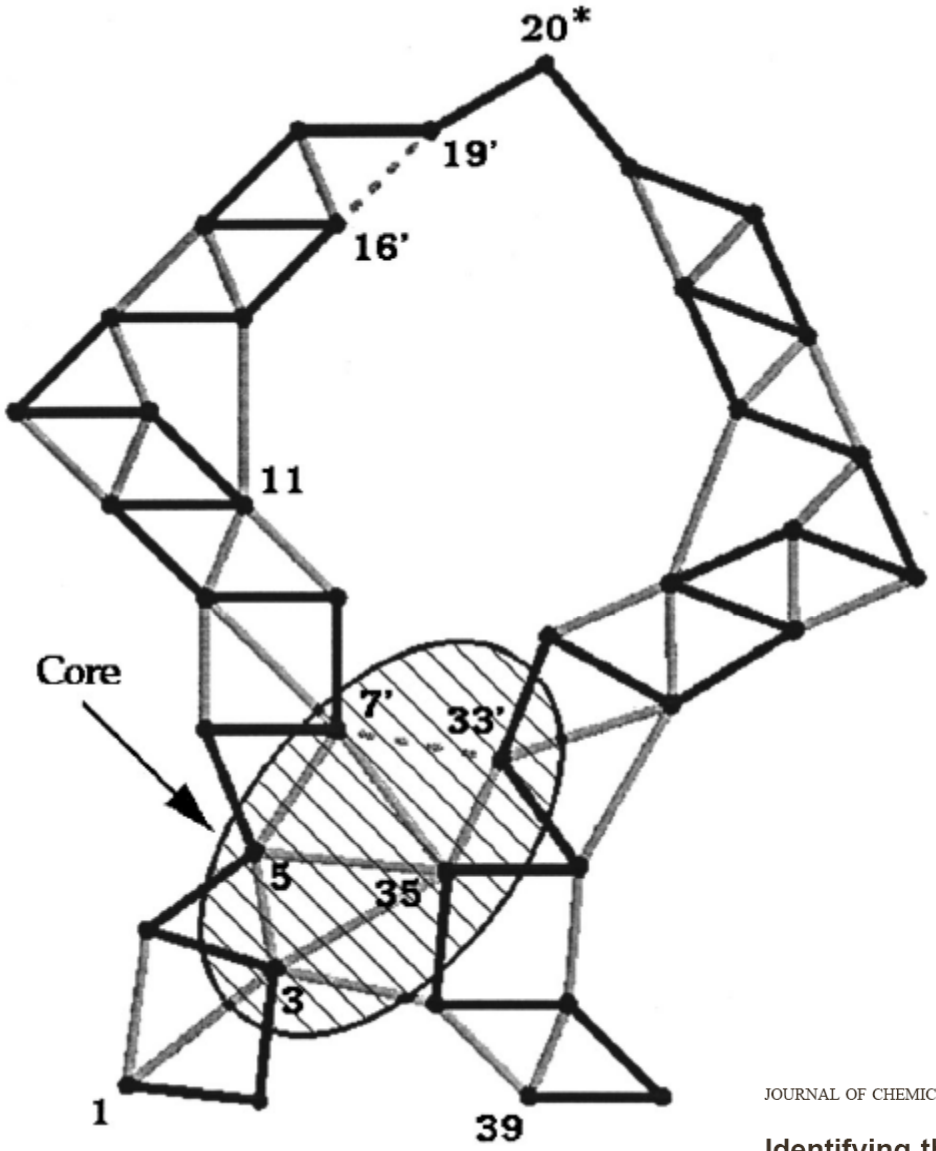


FIGURE 10 (*a*) RBP conformations visited by the action of the slowest mode (2), illustrating the large amplitude fluctuations of the loops at the entrance of the β-barrel. The fluctuation sizes refer to those of residues 65 (*left*) and 95 (*right*) on the loops. (*b*) Complexation of RBP (*green*) with its carrier protein transthyretin (*purple*), as observed in the complex crystallized by Naylor and Newcomer (1999). Note that the most flexible parts in (*a*) are involved in the recognition of the transthyretin, along with the C-terminus. Arrows indicate protein-protein close contacts. Vitamin A is shown in *magenta*.

Biophysical Journal Volume 80 January 2001 505–515

Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model

A. R. Atilgan,* S. R. Durell,† R. L. Jernigan,† M. C. Demirel,* O. Keskin,† and I. Bahar*†



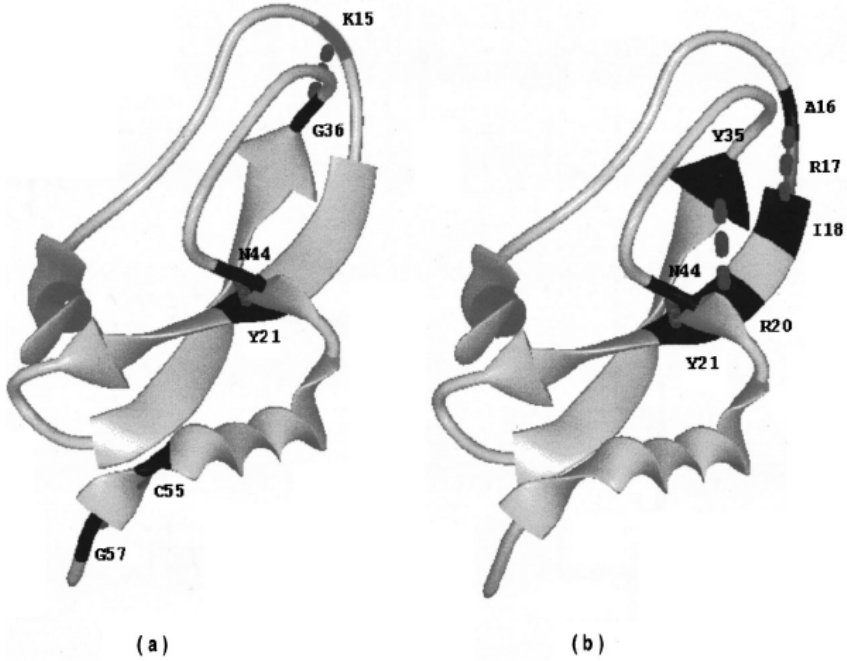


FIG. 10. Control architecture for the BPTI. In (a) the sensory residue Lys15 shown in gray. Residues displayed in black make the actuating contacts (marked by dashed lines) that control the Lys15. In (b) Arg17 is taken as a sensor and colored in gray. The controlling residues are displayed in black and actuating contacts are indicated by the dashed lines.

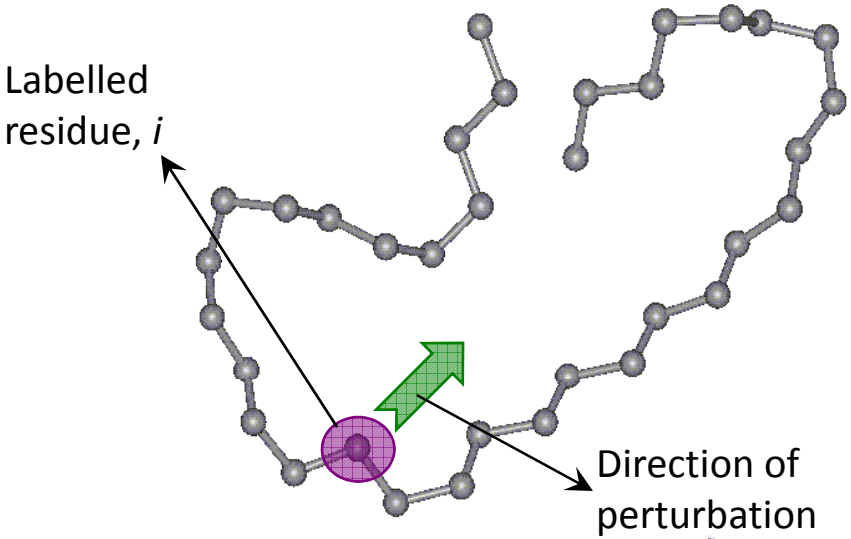
JOURNAL OF CHEMICAL PHYSICS VOLUME 113, NUMBER 10 8 SEPTEMBER 2000

Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts

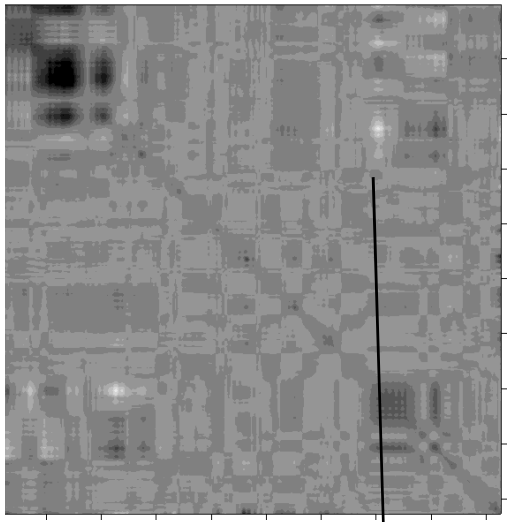
Lutfu Safak Yilmaz and Ali Rana Atilgan[a)]
Polymer Research Center and School of Engineering, and TUBITAK Advanced Polymeric Materials Research Center, Bogazici University, Bebek 80815, Istanbul, Turkey

linear response theory

uses cross correlations as a *kernel*



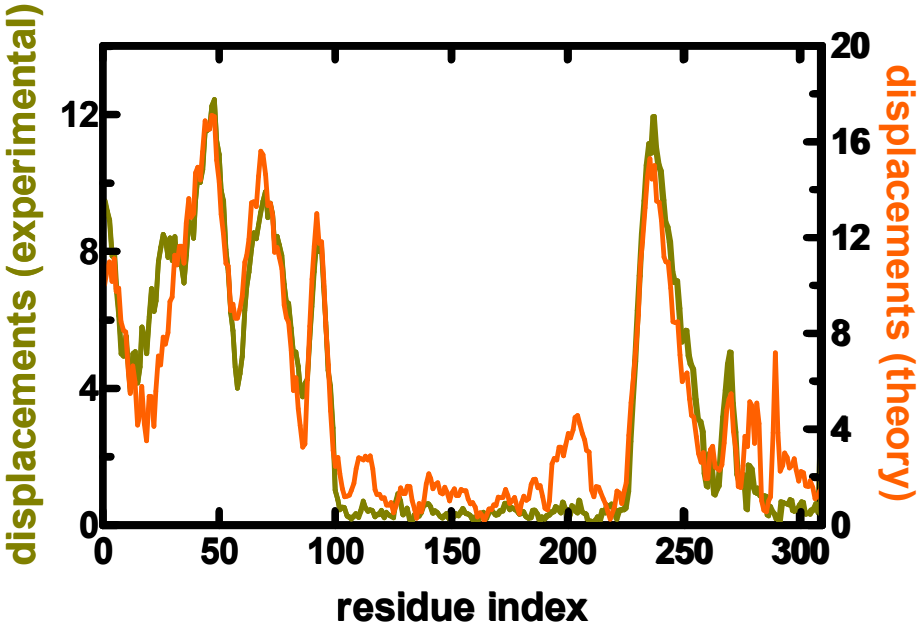
$$\begin{bmatrix} \Delta F_x^i \\ \Delta F_y^i \\ \Delta F_z^i \end{bmatrix}$$

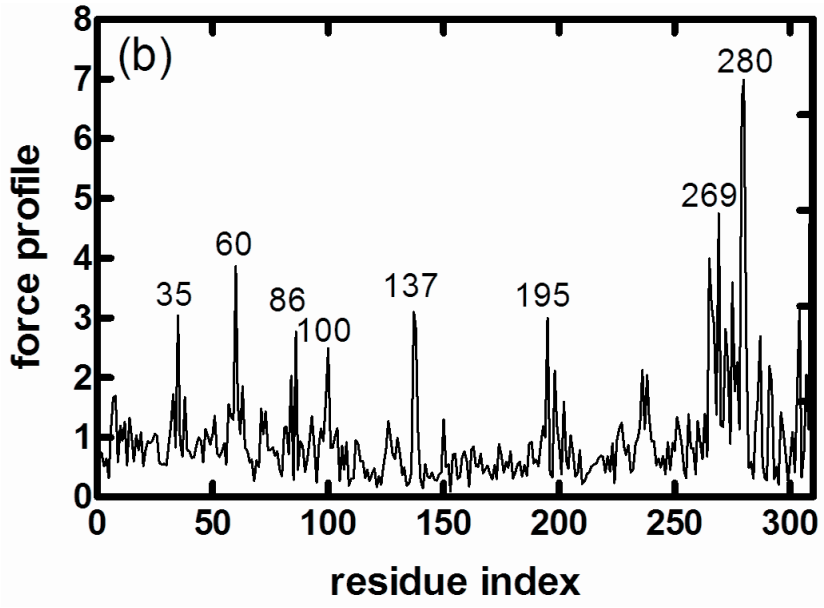


$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Delta F_x^i \\ \Delta F_y^i \\ \Delta F_z^i \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{3N\times 1} = \begin{bmatrix} \Delta R_x^1 \\ \Delta R_y^1 \\ \Delta R_z^1 \\ \vdots \\ \Delta R_x^N \\ \Delta R_y^N \\ \Delta R_z^N \end{bmatrix}_{3N\times 1}$$

to calculate the response

$$\Delta \mathbf{R}^{ki} = \begin{bmatrix} g_{x_k x_i} & g_{x_k y_i} & g_{x_k z_i} \\ g_{y_k x_i} & g_{y_k y_i} & g_{y_k z_i} \\ g_{z_k x_i} & g_{z_k y_i} & g_{z_k z_i} \end{bmatrix} \begin{bmatrix} \Delta F_x^i \\ \Delta F_y^i \\ \Delta F_z^i \end{bmatrix}$$



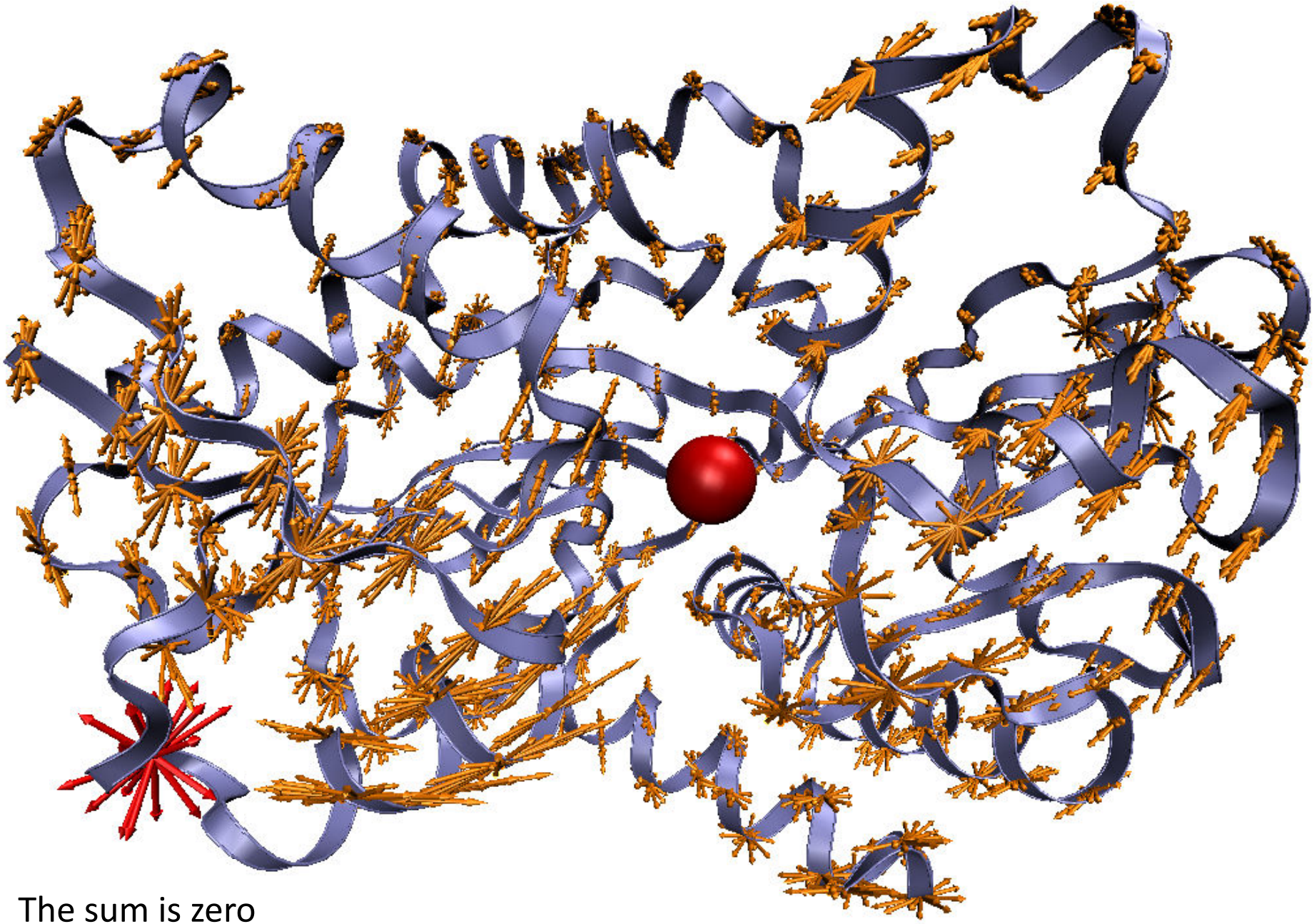


$$\langle (\Delta \mathbf{R}^{ki})^2 \rangle = \langle (\Delta F_x^i)^2 \rangle \left[(g_{x_k x_i})^2 + (g_{x_k y_i})^2 + (g_{x_k z_i})^2 \right]$$

$$\langle (\Delta \mathbf{R}^{ki})^2 \rangle = \frac{\bar{F}^2}{3} \sum_{l=x_k,y_k,z_k} \sum_{m=x_i,y_i,z_i} (g_{lm})^2$$

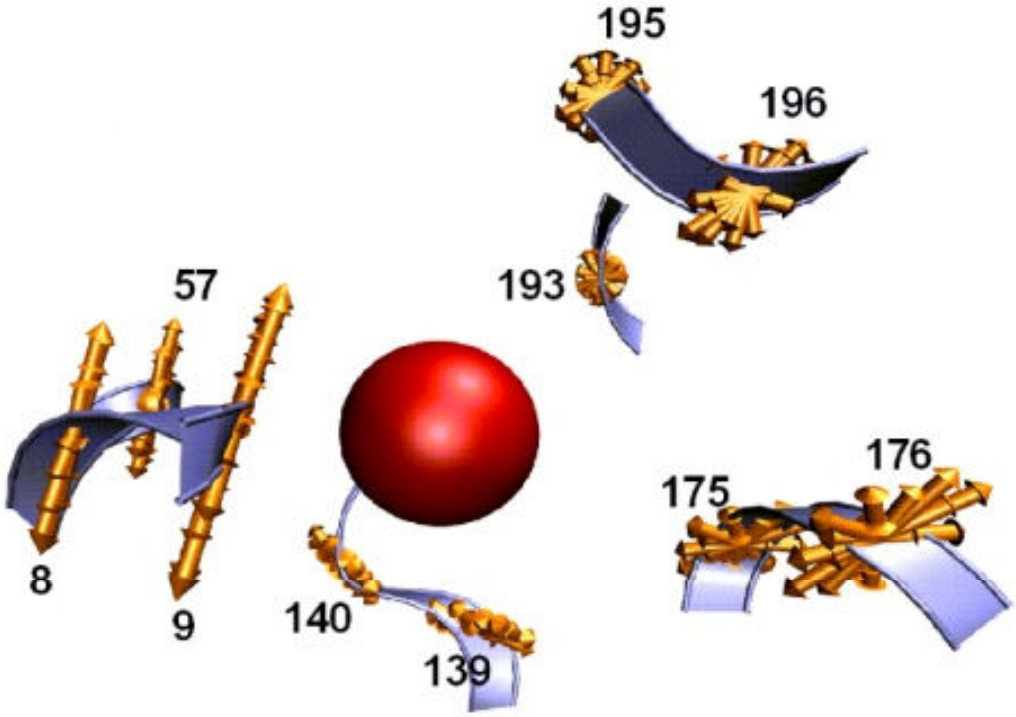
mode dominant behavior

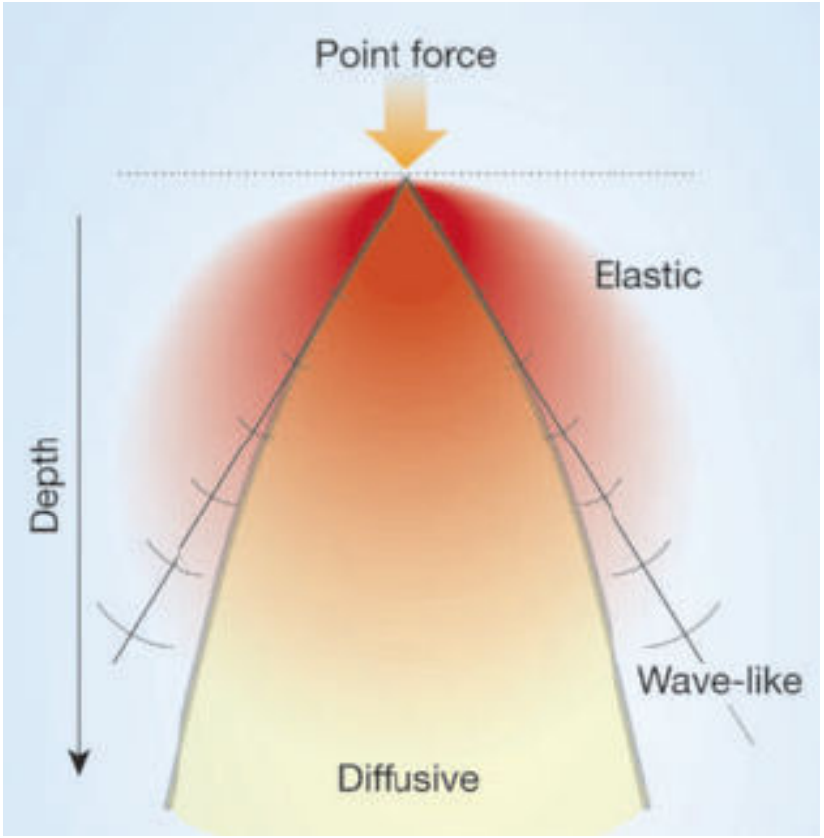
CA & ARA, ***PLOS Comp Biol***, to appear 2009



The sum is zero

how does **it** *propagate*?

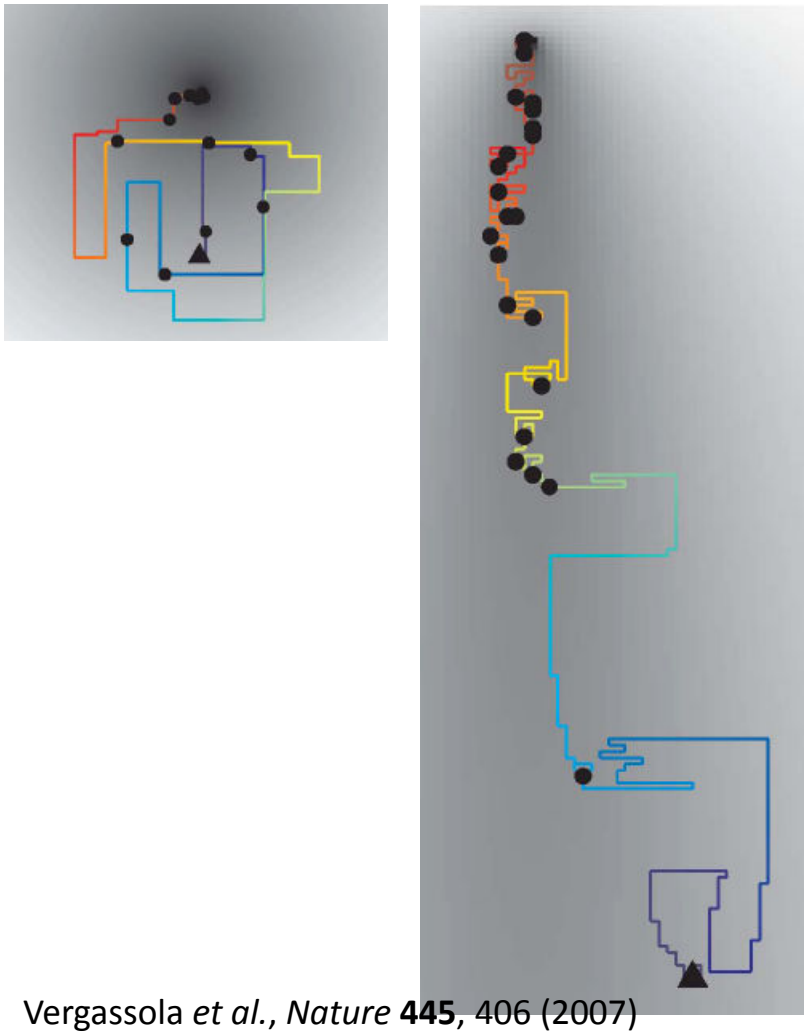




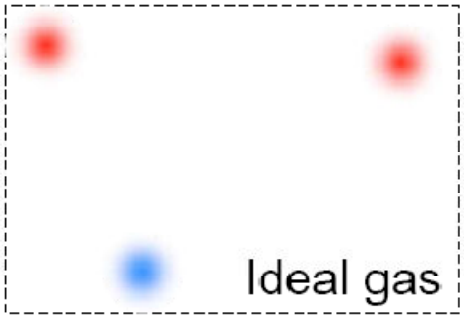
for short distances, forces in granular systems propagate much like waves, but at longer distances an applied force causes an elastic-like deformation

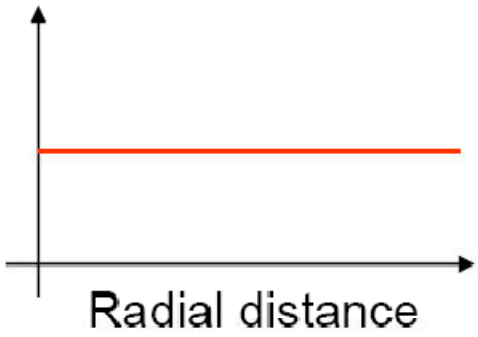
Goldenberg and Goldhirsch, *Nature* **435**, 188 (2005)

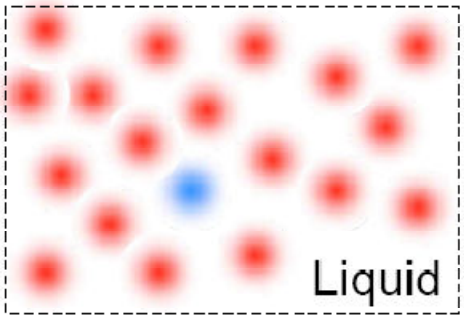
chemotactic bacteria rely on local concentration gradients to guide them towards the source of a nutrient

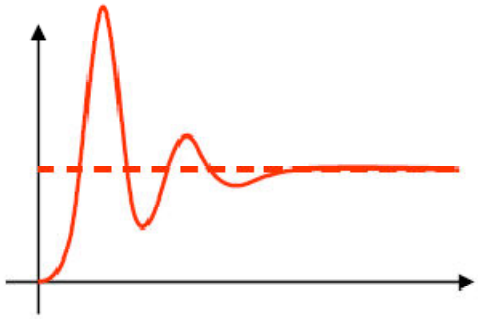


Vergassola *et al.*, *Nature* **445**, 406 (2007)

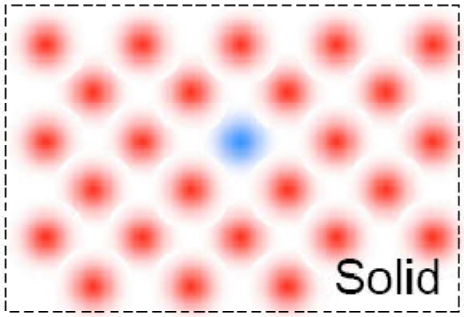


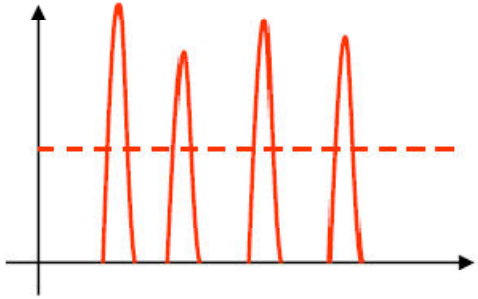




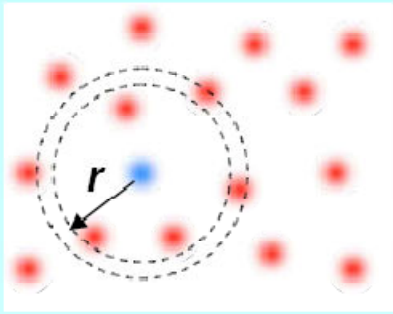


Soft matter





Radial distribution function



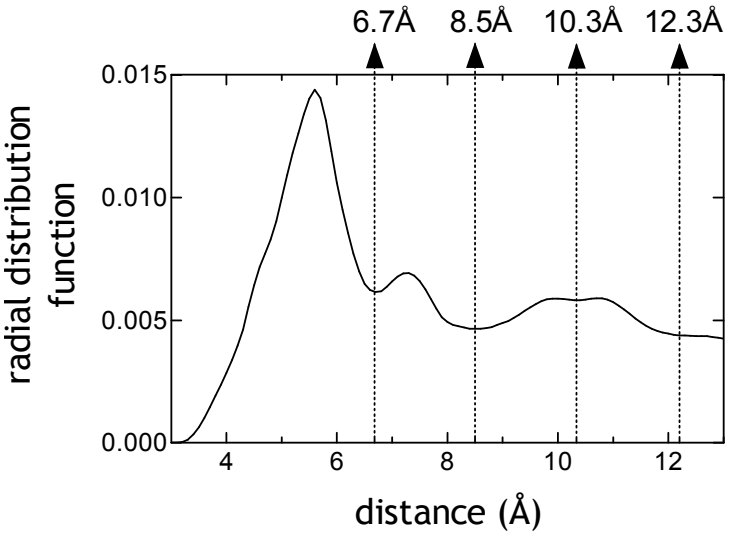
Calculate the number of particles inside a radial segement $(r, r+\Delta r)$ from one specific particle in the system.

Average the distribution over all particles in the system.

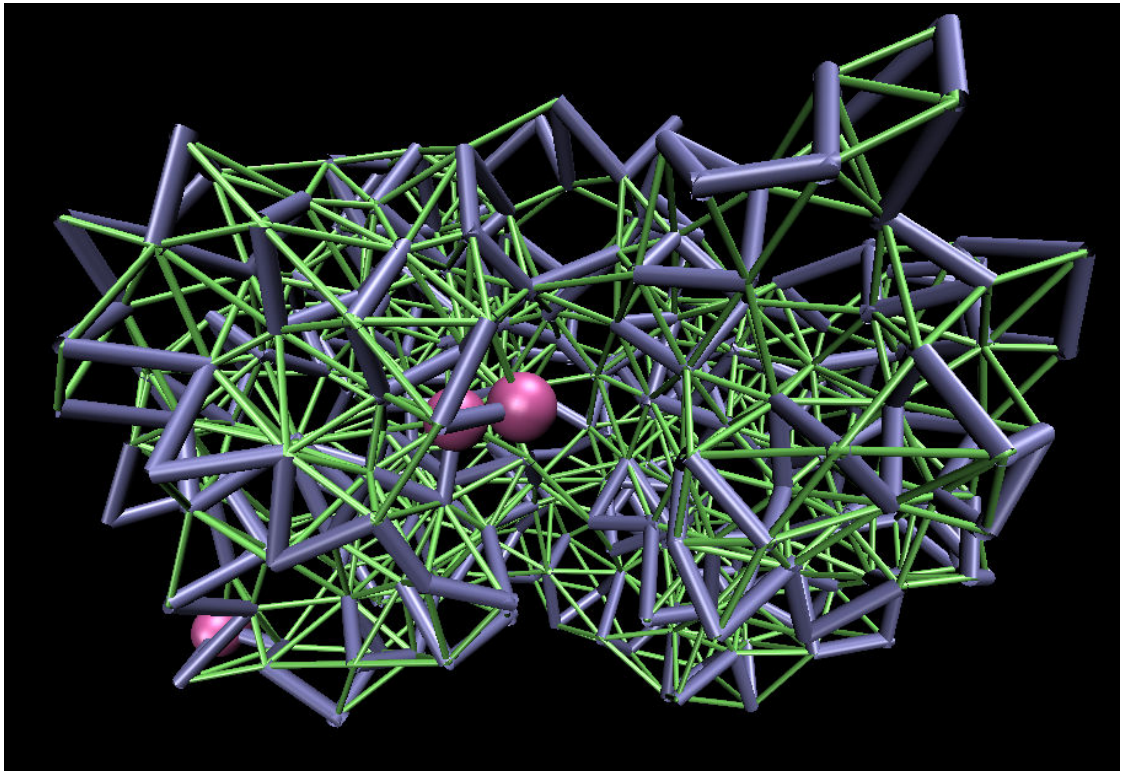
Divide the (average) number of particles in a segment by the segment volume.

Finally, normalize thus obtained radial density profile with the bulk density N/V of the system.

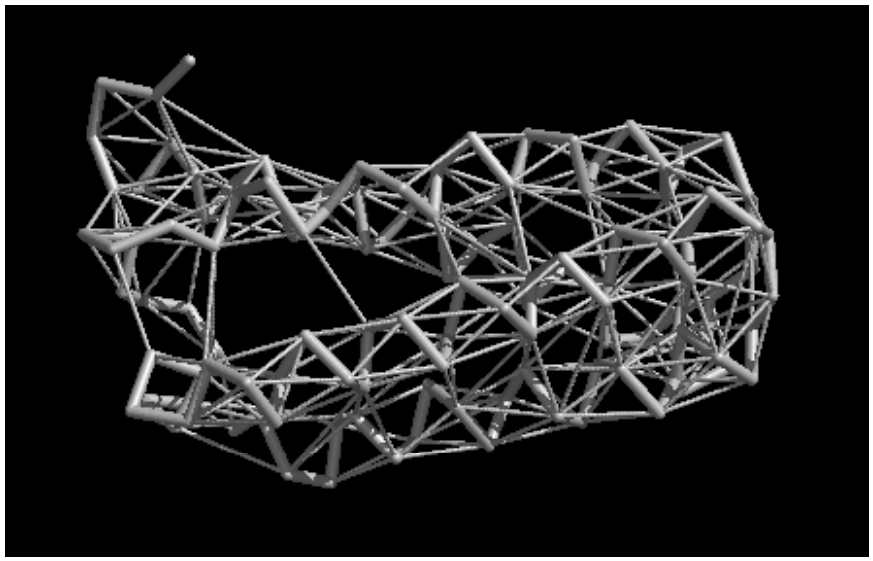
http://www.eng.buffalo.edu/~kofke/applets/rdfSW.html



This architecture is capable of **organizing short average path lengths** between any two nodes in a structure, but it cannot warrant a high clustering similar to regular packing



Adjacency and connectivity

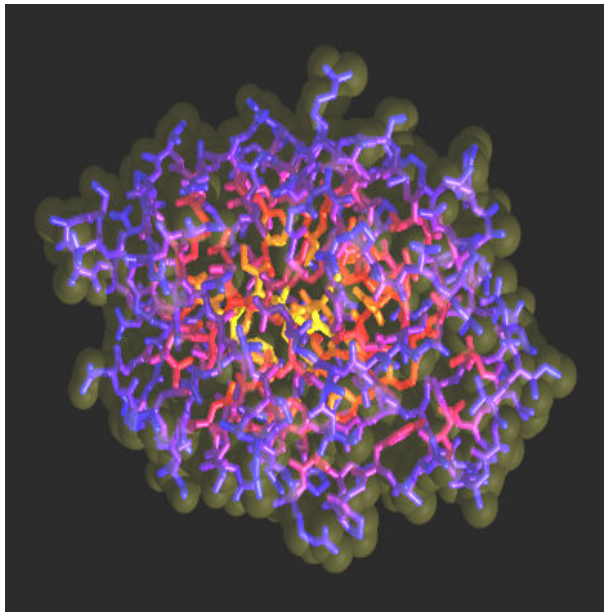


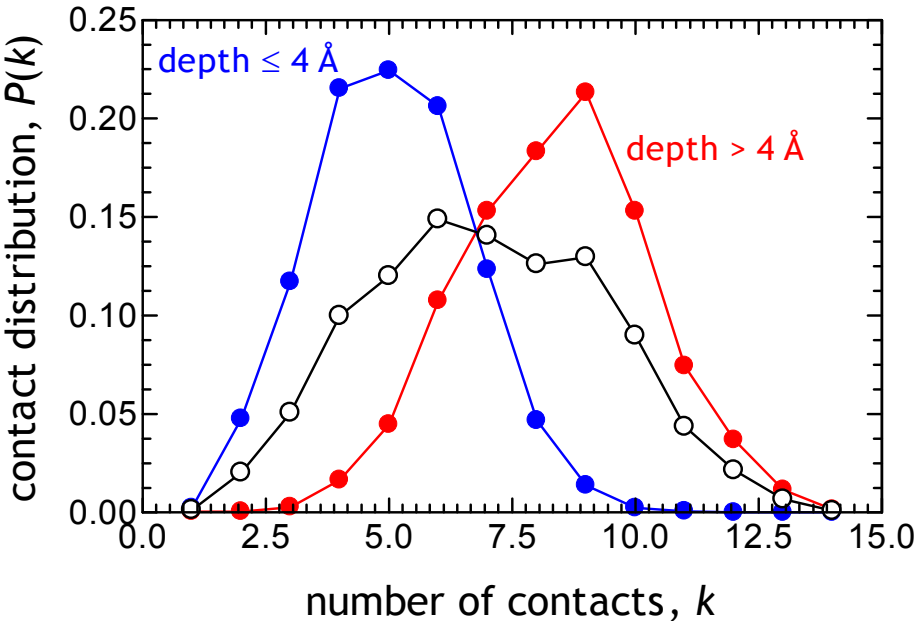
$$A_{ij} = \begin{cases} 1 & if\ nodes\ i\ and\ j\ are\ connected \\ 0 & if\ nodes\ i\ and\ j\ are\ not\ connected \end{cases}$$

$$k_i = \sum_{j=1}^{N} A_{ij}$$

Contact distribution

surface-molten solids – Zhou *et al.*, JMB, 285, 1371 (1999)
random packing of hard spheres – Soyer *et al.*, PRL, 85, 3532, (2000)





the interiors of proteins are more like randomly packed spheres near their percolation threshold and that larger proteins are packed more loosely than smaller proteins – Liang and Dill, BJ, 81, 751, (2001)

DEPTH. Chakravarty and R. Varadarajan, Structure, 7, 723 (1999)

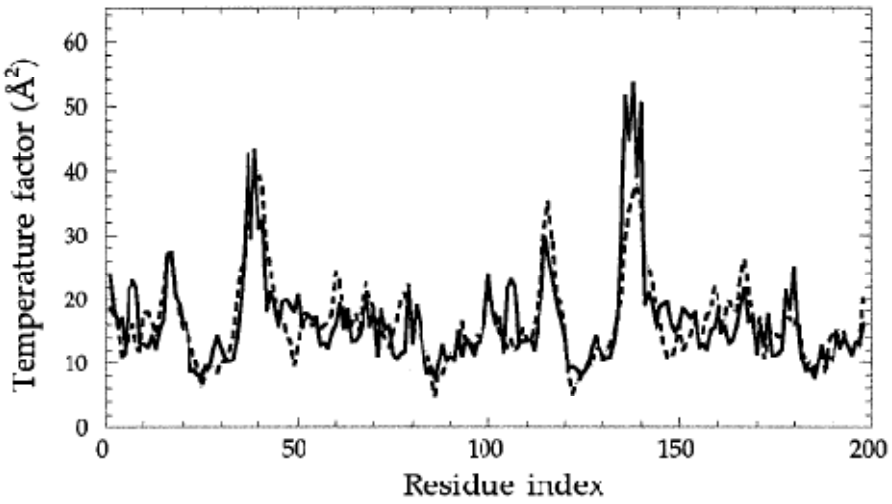
Slow and fast motions

Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential

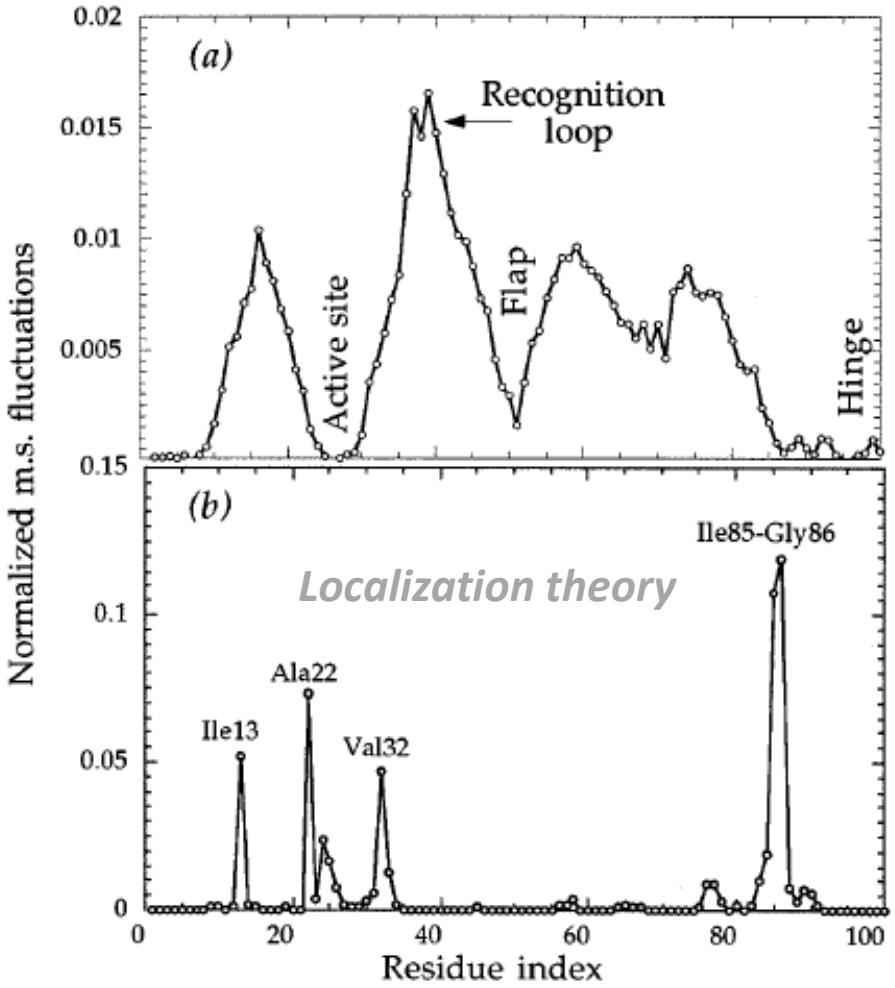
Ivet Bahar, Ali Rana Atilgan and Burak Erman

Folding & Design 07 May 1997, 2:173–181

X-ray crystallographic temperature factors



$$B_i = 8\pi^2 \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle / 3$$



VOLUME 80, NUMBER 12 PHYSICAL REVIEW LETTERS 23 MARCH 1998

Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability

Ivet Bahar, Ali Rana Atilgan, Melik C. Demirel, and Burak Erman

Polymer Research Center, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey

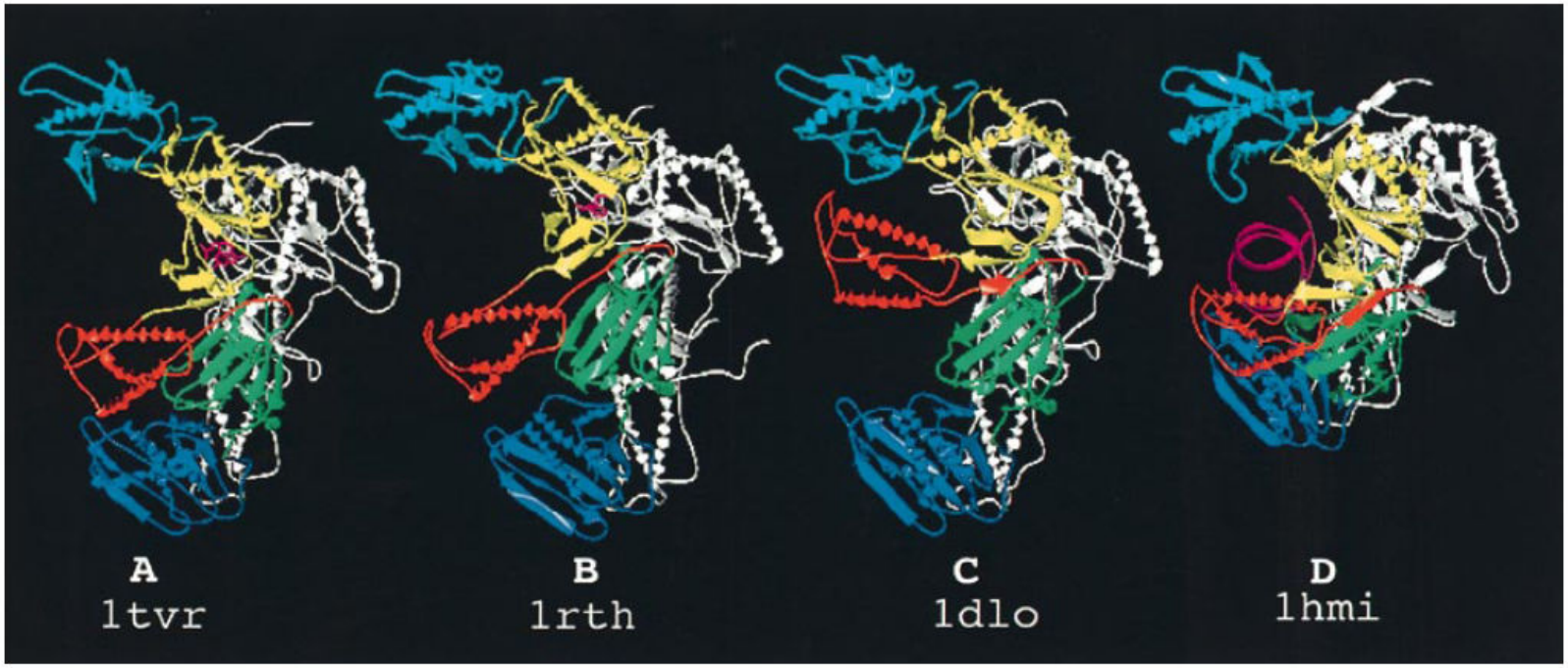


Figure 4. Ribbon diagrams of the four RT structures considered in Figure 3. The PDB codes of the structures are indicated. The thumb, palm, fingers and connection subdomains of the p66 polymerase are colored in red, yellow, cyan, and green, respectively. The RNase H domain is shown in blue, and the p51 subunit is white. The NNRTIs in 1tvr and 1rth and the DNA double-strands in 1hmi are shown in magenta.

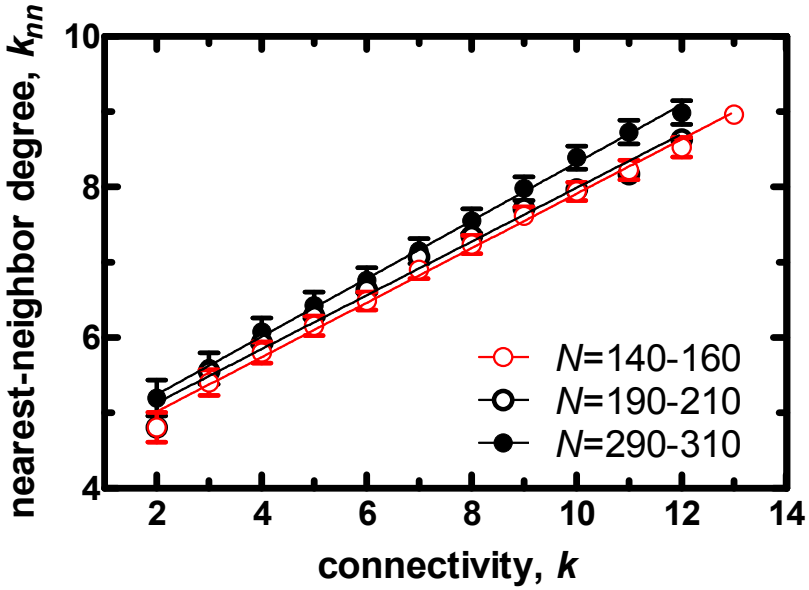
Average degree of the nearest neighbors

exploration of the connectivity correlations

The conditional probability that a link belonging to a node with connectivity *k* points to a node with connectivity *k'*

$$\overline{k}_{nn}(k) = \sum_{k'} k' P(k'|k)$$

$$k_{nn,i} = \frac{\sum_{j=1}^{N}\sum_{m=1}^{N} A_{ij}A_{jm}}{\sum_{j=1}^{N} A_{ij}} = \frac{\sum_{j=1}^{N} A_{ij}k_j}{k_i}$$

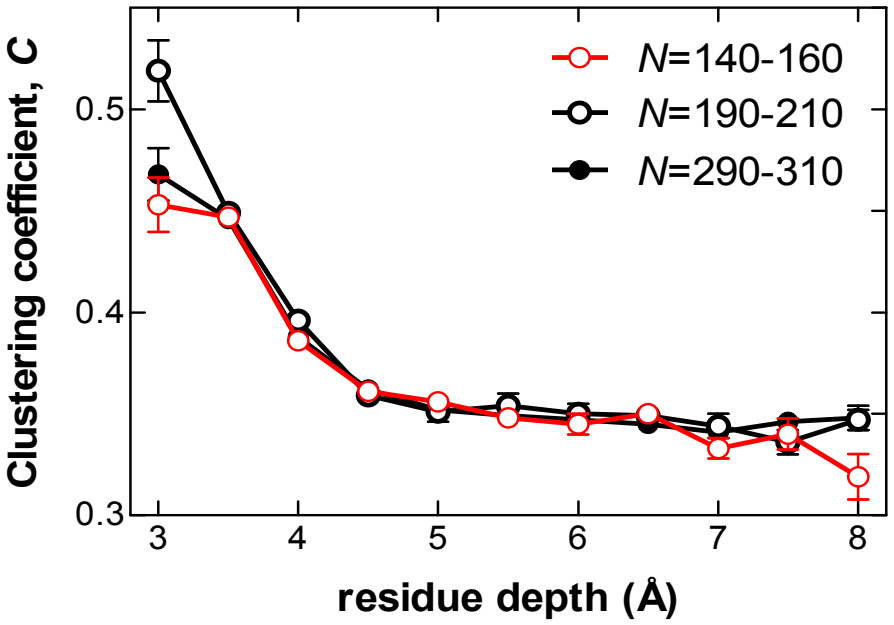


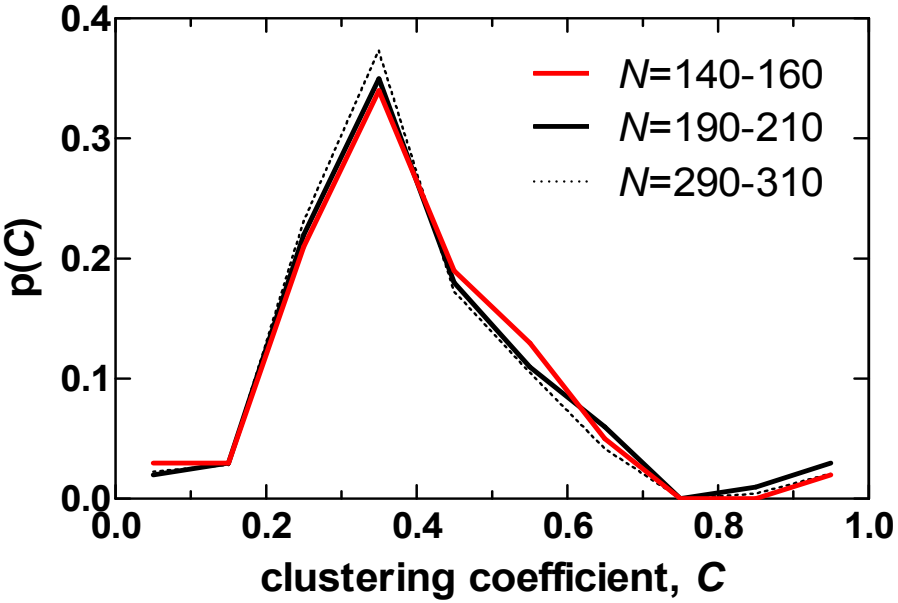
the explicit dependence on *k* is a signature of nontrivial correlations among the nodes' connectivity, and the possible presence of a hierarchical structure in the network topology

Romualdo Pastor-Satorras, *et al.*, PRL, 87, 258701 (2001)

$$C_i = \frac{\frac{1}{2}\sum_{j=1}^{N}\sum_{m=1}^{N} A_{ij} A_{jm} A_{mi}}{\frac{k_i(k_i-1)}{2}}$$

The probability that two vertices with a common neighbor are also connected to each other is called the *clustering coefficient* of the common vertex



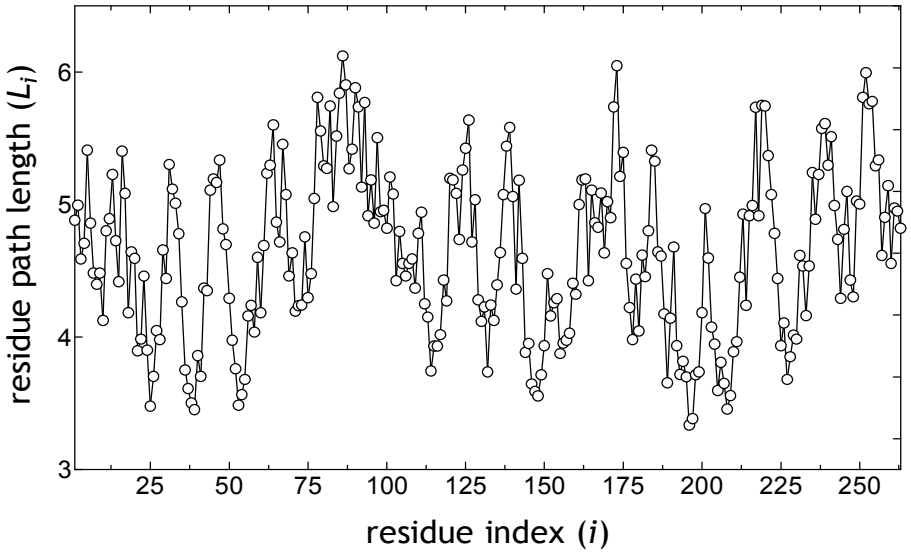


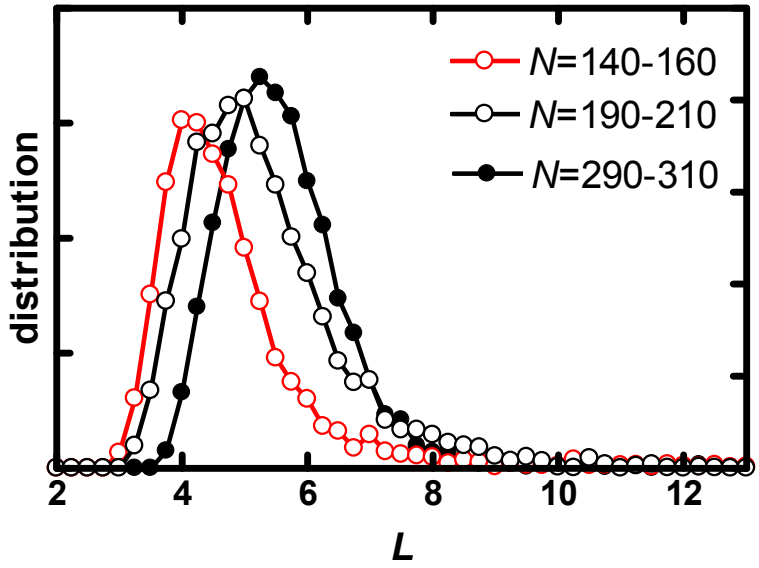
Biophysical Journal Volume 86 January 2004 85–91

Small-World Communication of Residues and Significance for Protein Dynamics

Ali Rana Atilgan,* Pelin Akan,† and Canan Baysal†
*School of Engineering, Bogazici University, Bebek 34342, Istanbul, Turkey; and †Laboratory of Computational Biology, Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 34956, Tuzla, Istanbul, Turkey

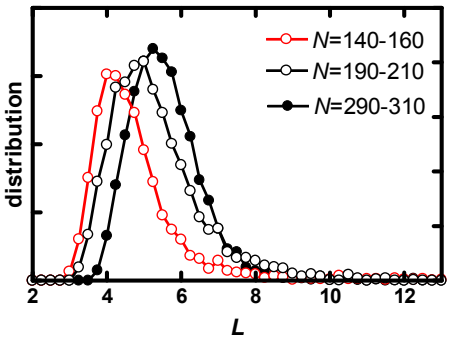
The characteristic path length is defined as the number of edges in the shortest path between two vertices, averaged over all pairs of vertices

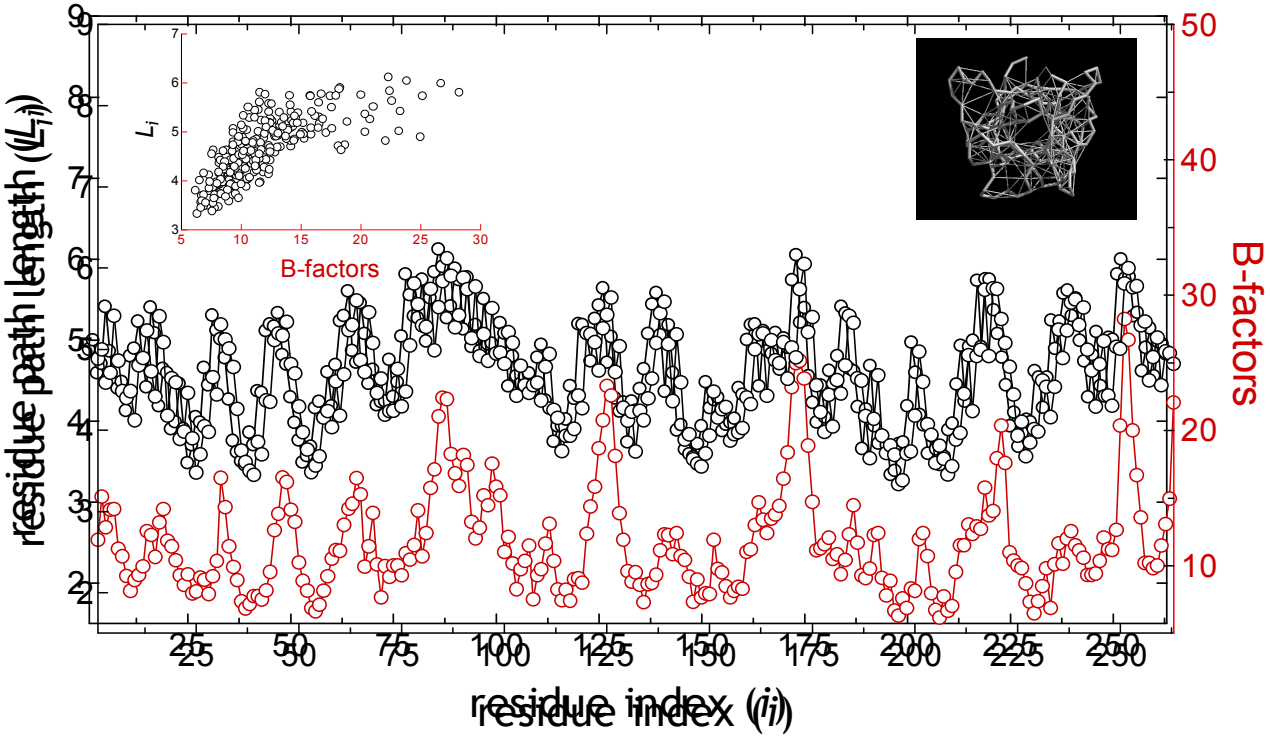




Average path length

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = (k_B T/\gamma)[\Gamma^{-1}]_{ij}$$



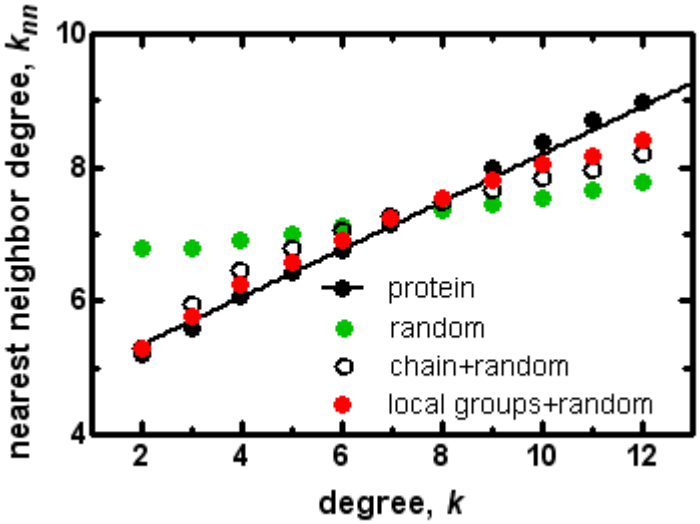


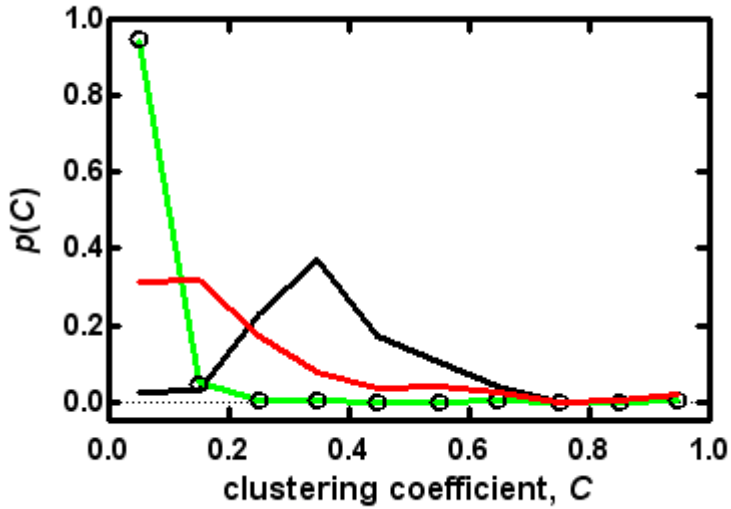
One would expect an indirect correlation between the fluctuations and shortest path lengths: The former are smaller for highly connected residues, which are in turn connected to the rest of the molecule, on average, in a shorter number of steps

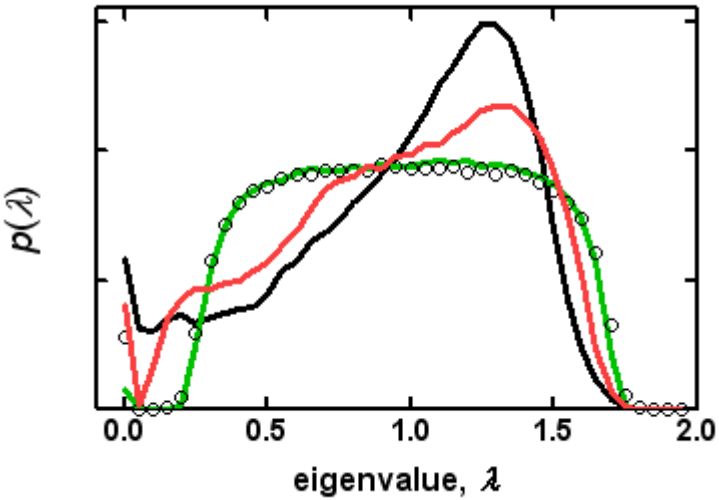
Biophysical Journal Volume 86 January 2004 85–91

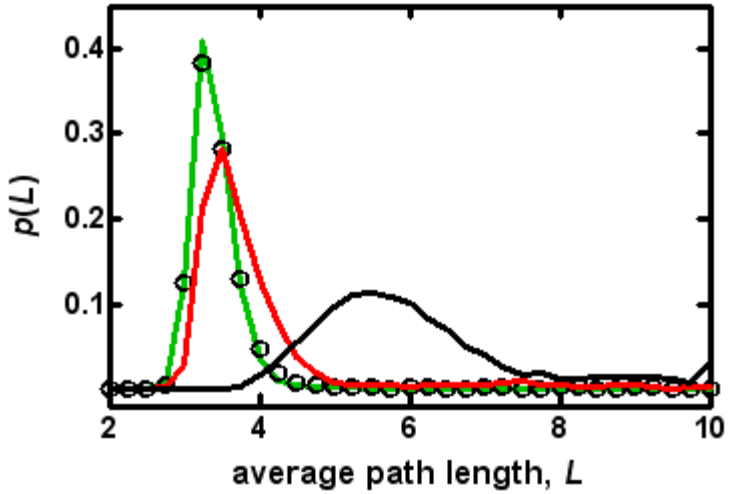
Small-World Communication of Residues and Significance for Protein Dynamics

Ali Rana Atilgan,* Pelin Akan,† and Canan Baysal†
*School of Engineering, Bogazici University, Bebek 34342, Istanbul, Turkey; and †Laboratory of Computational Biology, Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli 34956, Tuzla, Istanbul, Turkey

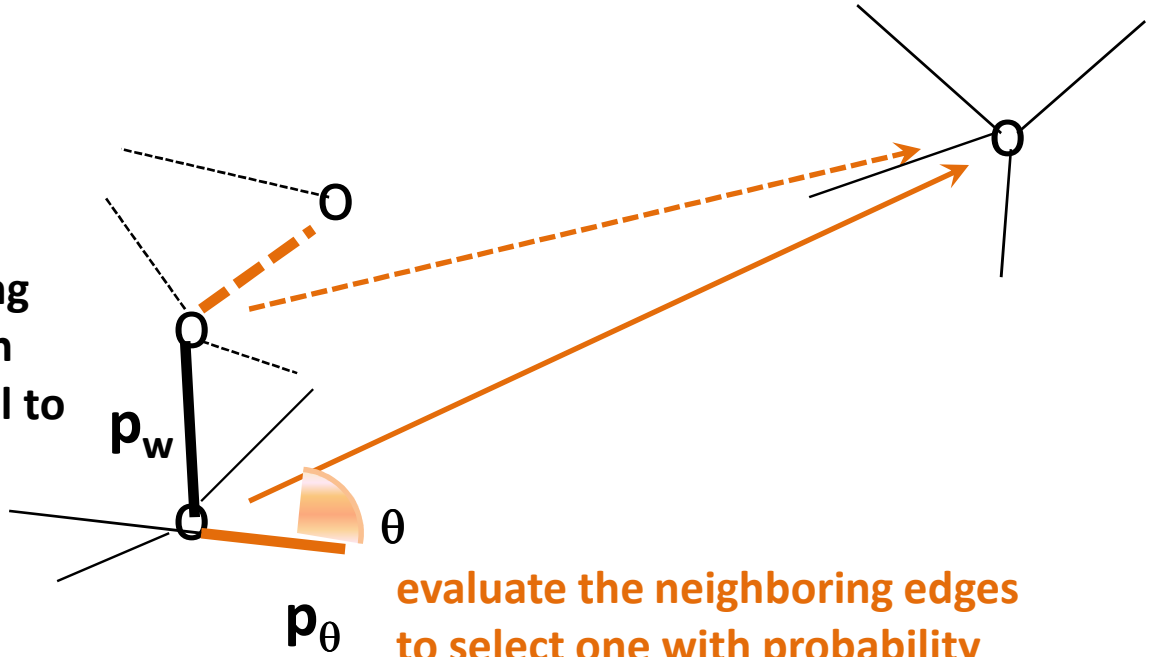

nearest neighbor degree, k_{nn}
degree, k
protein
random
chain+random
local groups+random


$p(C)$
clustering coefficient, C


$p(\lambda)$
eigenvalue, λ


$p(L)$
average path length, L

evaluate the neighboring edges to select one with probability proportional to their weights



evaluate the neighboring edges to select one with probability proportional to cos θ

f() := sum of weights → weak path length

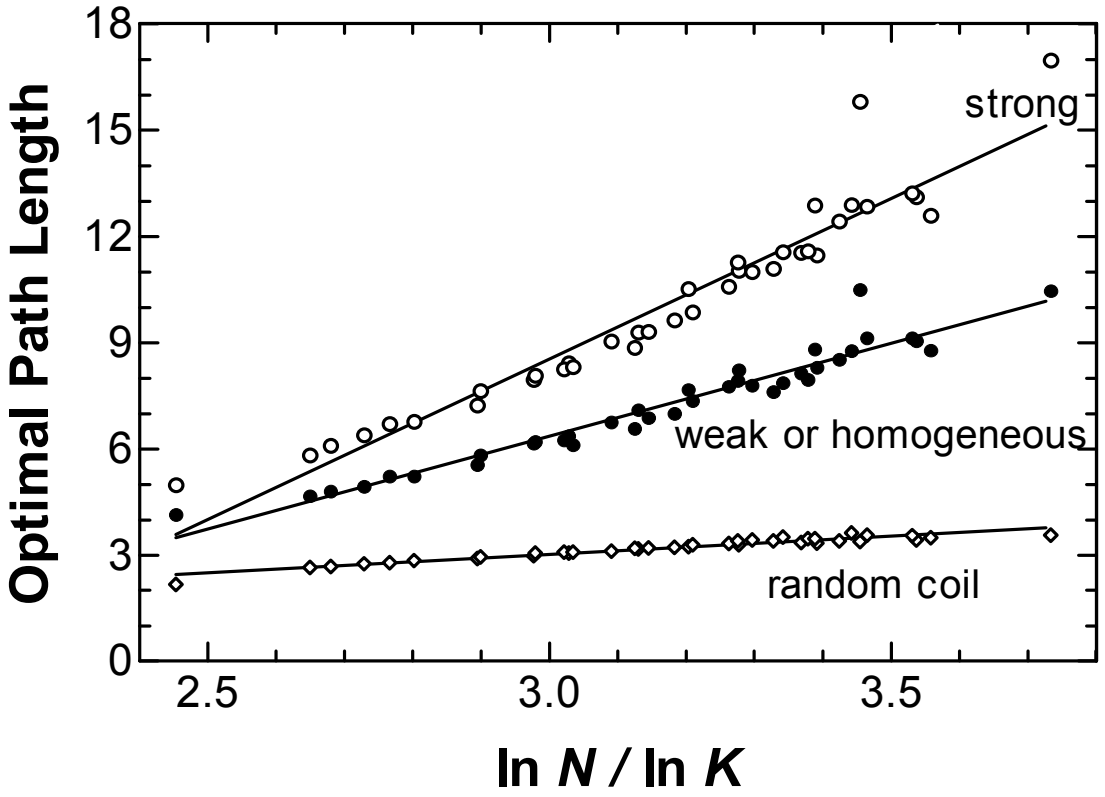
f() := max (weigths) → strong path length

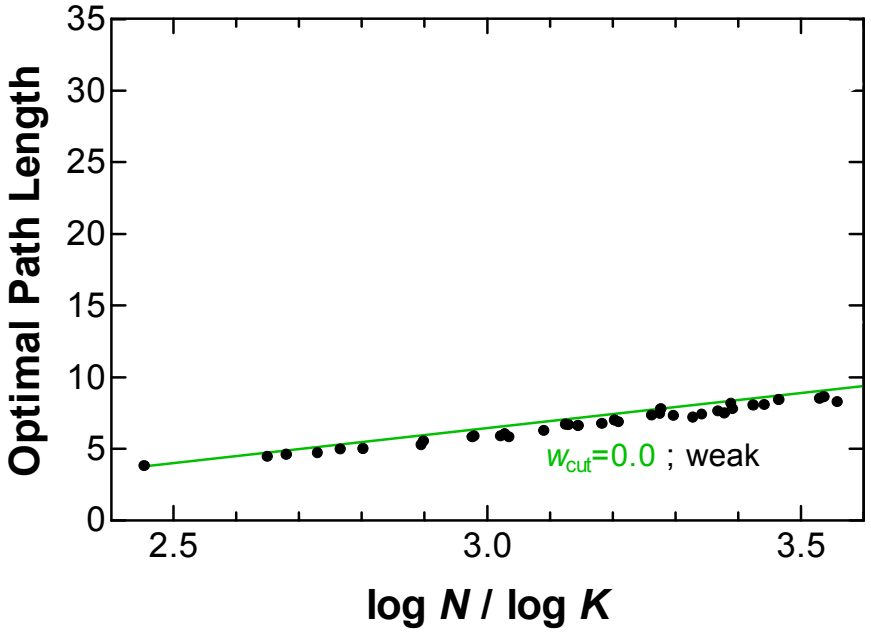
$$L_{ij} := \text{number of steps } (argmin_k (f(path_{ij}))_k)$$

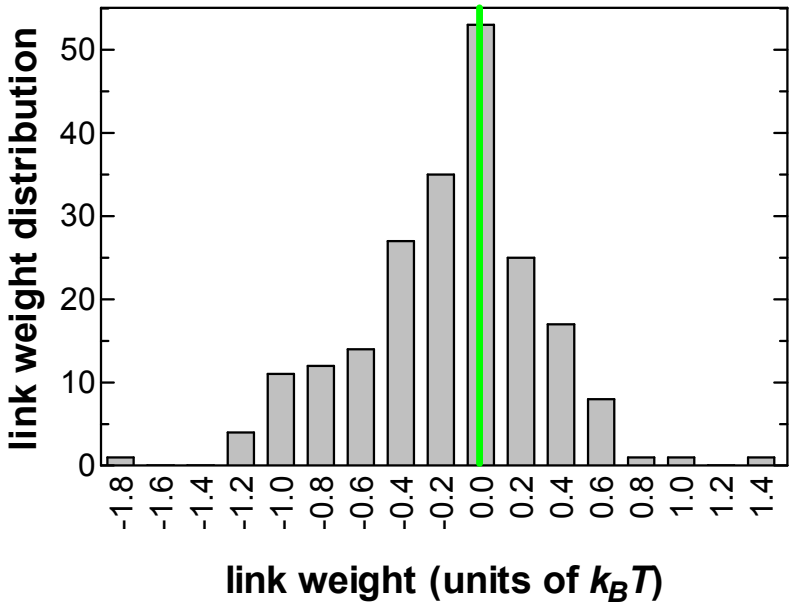
3052 Biophysical Journal Volume 92 May 2007 3052–3062

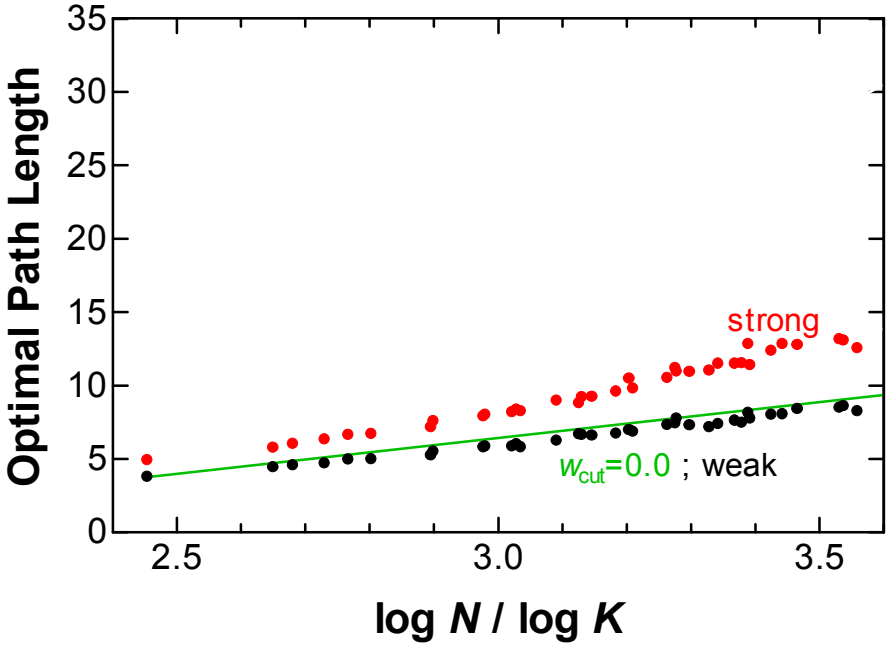
Screened Nonbonded Interactions in Native Proteins Manipulate Optimal Paths for Robust Residue Communication

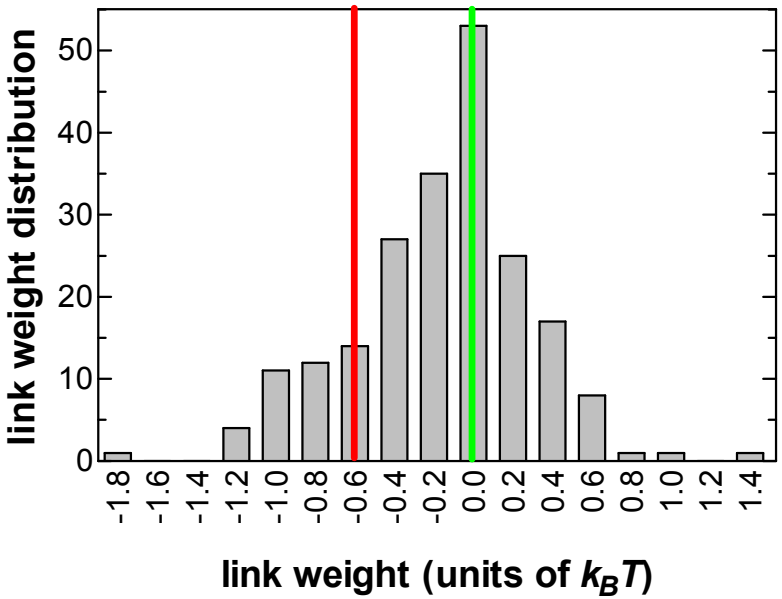
Ali Rana Atilgan, Deniz Turgut, and Canan Atilgan
Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey

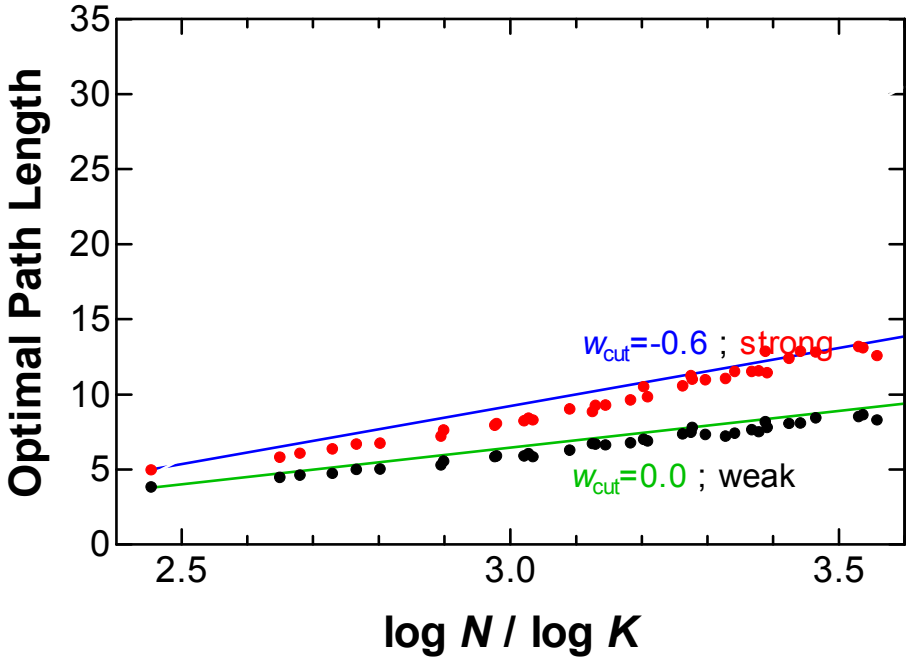

Optimal Path Length
18
15
12
9
6
3
0
2.5
3.0
3.5
ln N / ln K
strong
weak or homogeneous
random coil

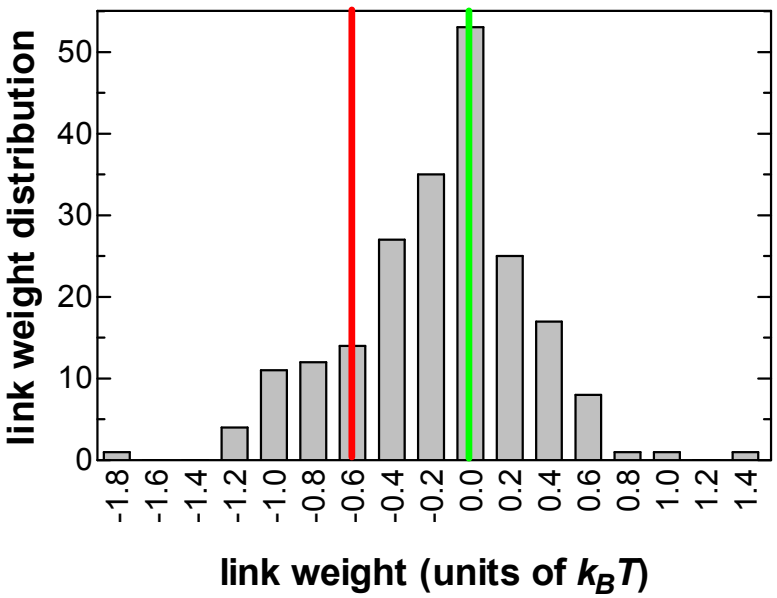

Optimal Path Length
35
30
25
20
15
10
5
0
2.5
3.0
3.5
log N / log K
w_{cut}=0.0 ; weak

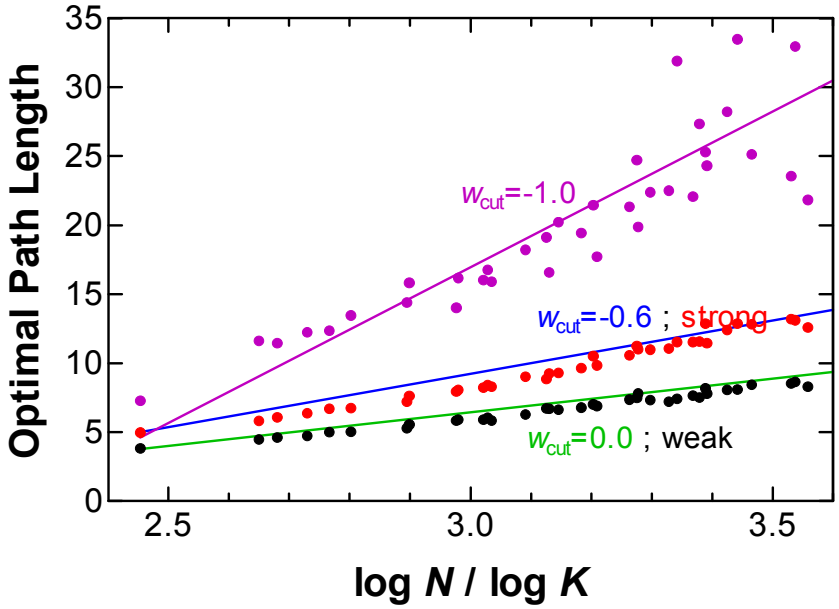

link weight distribution
50
40
30
20
10
0
-1.8
-1.6
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
link weight (units of k_BT)

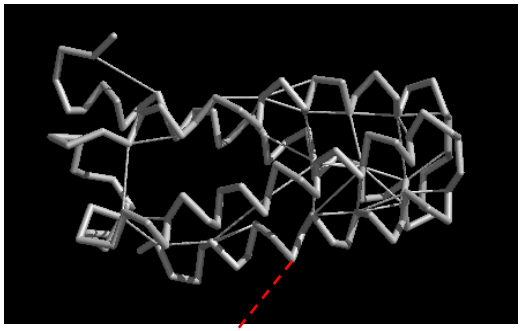

Optimal Path Length
log N / log K
strong
w_{cut}=0.0 ; weak
0
5
10
15
20
25
30
35
2.5
3.0
3.5

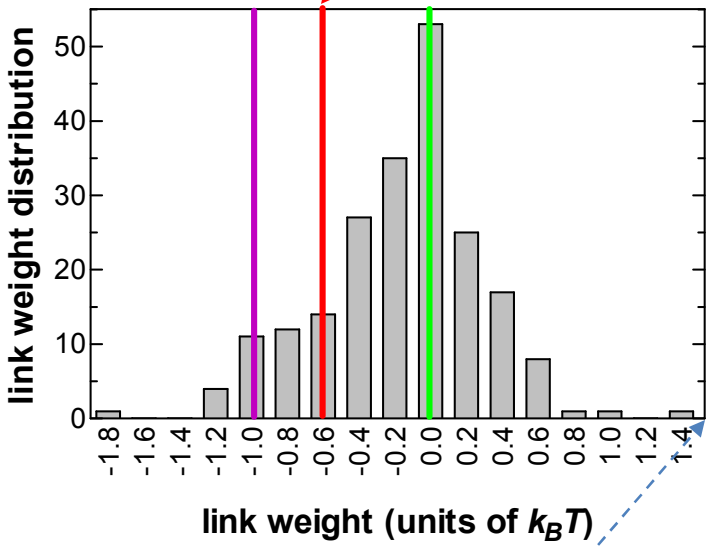

link weight distribution
link weight (units of k_BT)
0
10
20
30
40
50
-1.8
-1.6
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

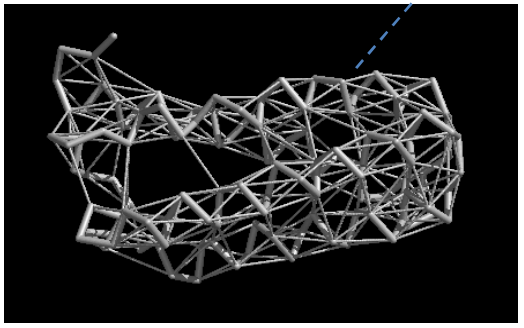

Optimal Path Length
log N / log K
w_cut=-0.6 ; strong
w_cut=0.0 ; weak
35
30
25
20
15
10
5
0
2.5
3.0
3.5

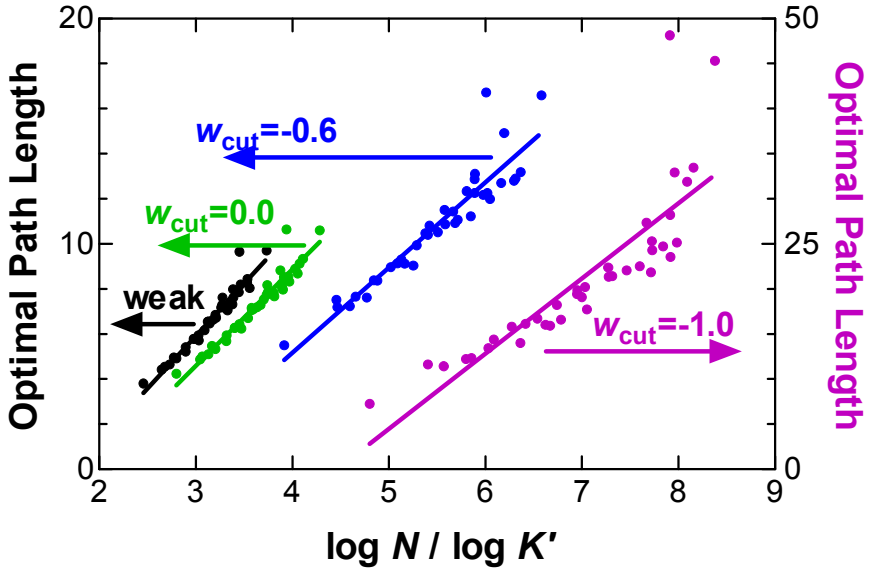

link weight distribution
link weight (units of k_BT)
50
40
30
20
10
0
-1.8
-1.6
-1.4
-1.2
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

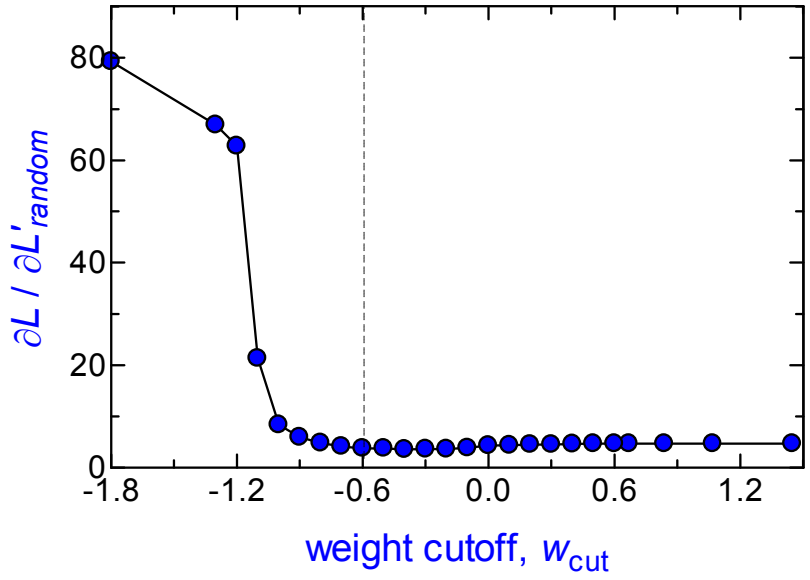

Optimal Path Length
log N / log K
w_cut=-1.0
w_cut=-0.6 ; strong
w_cut=0.0 ; weak

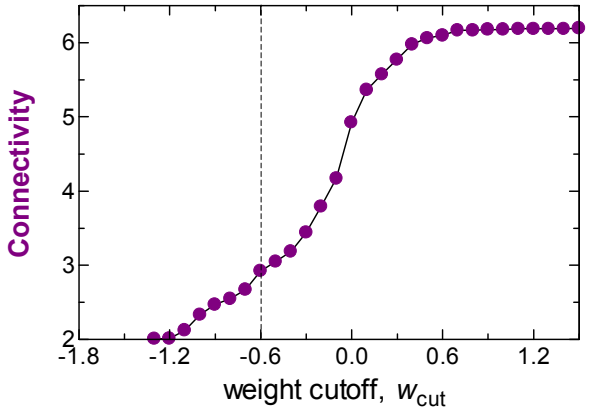


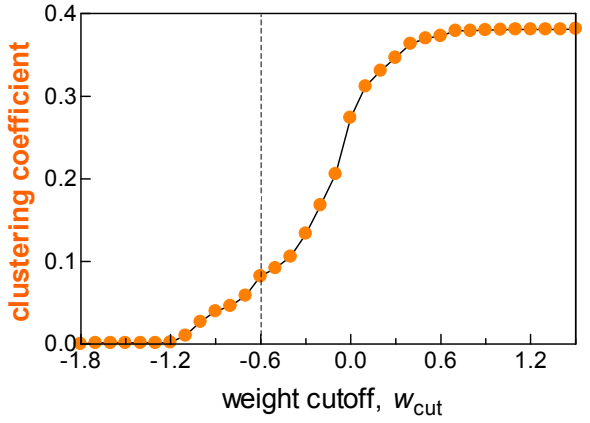

link weight distribution
link weight (units of k_BT)

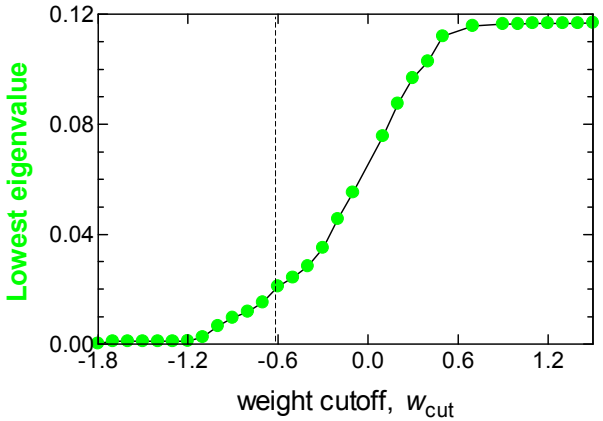



Optimal Path Length
Optimal Path Length
log N / log K'
w_cut=-0.6
w_cut=0.0
weak
w_cut=-1.0
20
10
0
50
25
0
2
3
4
5
6
7
8
9


∂L / ∂L'random
weight cutoff, w_cut
80
60
40
20
0
-1.8
-1.2
-0.6
0.0
0.6
1.2


Connectivity
weight cutoff, w_cut
6
5
4
3
2
-1.8
-1.2
-0.6
0.0
0.6
1.2

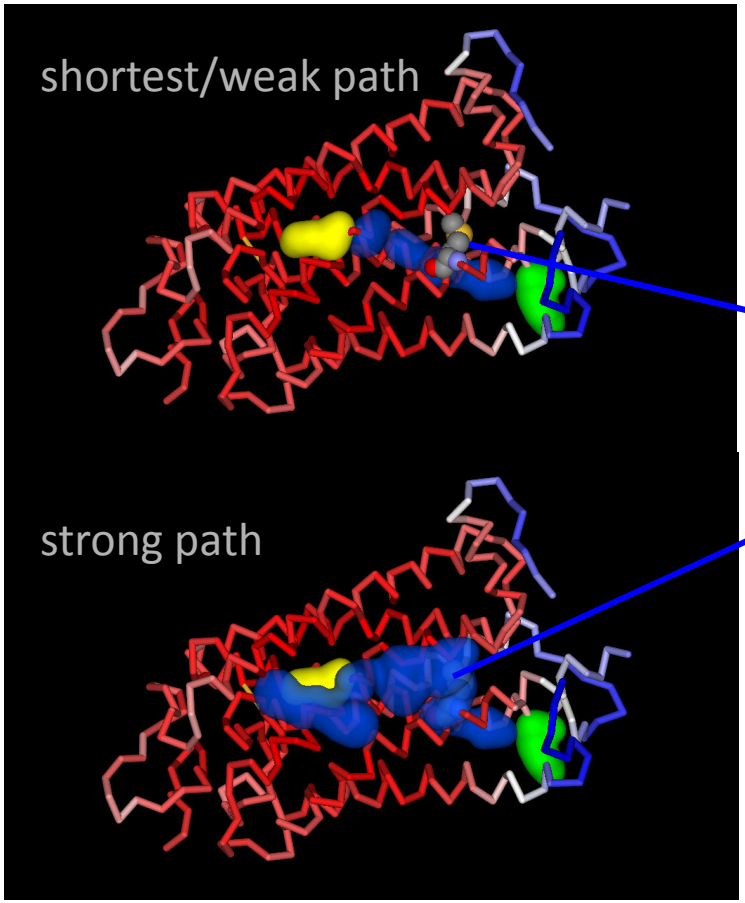

clustering coefficient
weight cutoff, w_cut
0.4
0.3
0.2
0.1
0.0
-1.8
-1.2
-0.6
0.0
0.6
1.2


Lowest eigenvalue
weight cutoff, w_cut
0.12
0.08
0.04
0.00
-1.8
-1.2
-0.6
0.0
0.6
1.2

Conserved residues regulating allostery identified

Lockless, SW and R Ranganathan, *Science* **286**, 295 (1999)

GPCR family representative (Rhodopsin – 1l9h)

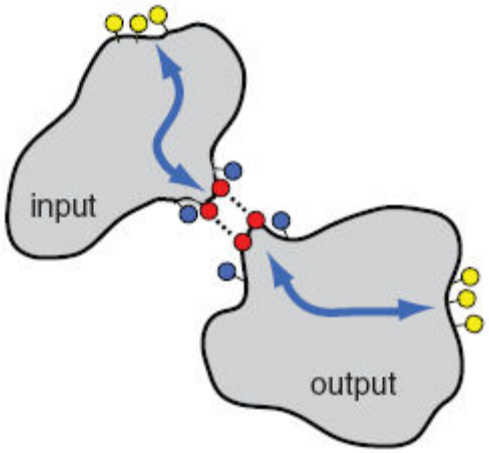


We study the **paths between the active site residue LYS296 (yellow) and TYR136** that participates in one of the regions that undergoes a structural change upon light activation in rhodopsin **(green)**

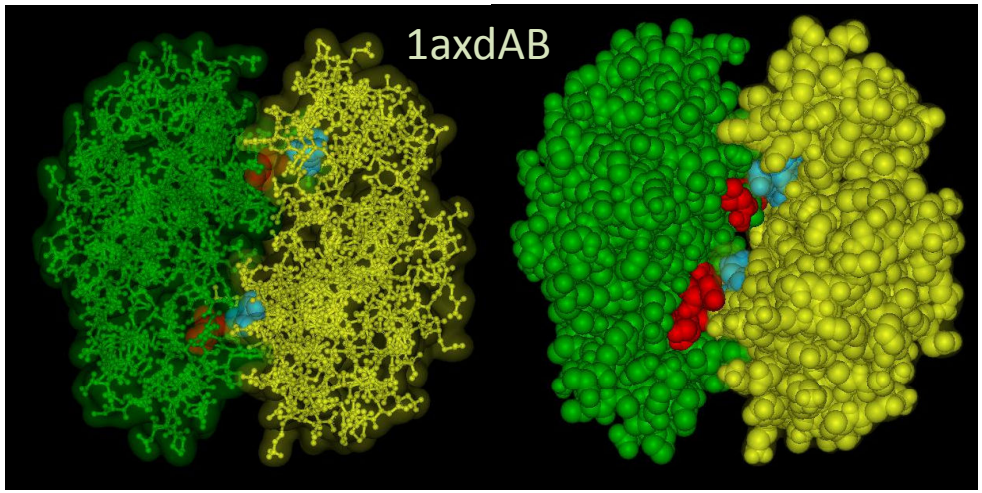
MET 257

known to participate in constitutive activity of the molecule, possibly through interactions with the conserved motif involving residues 302-306 on helix 7. *It is bypassed in the weak path, but detected by the strong path.*

Key interactions on binding surface identified



Lee J *et al.*, *Science* **322**, 438 (2008)



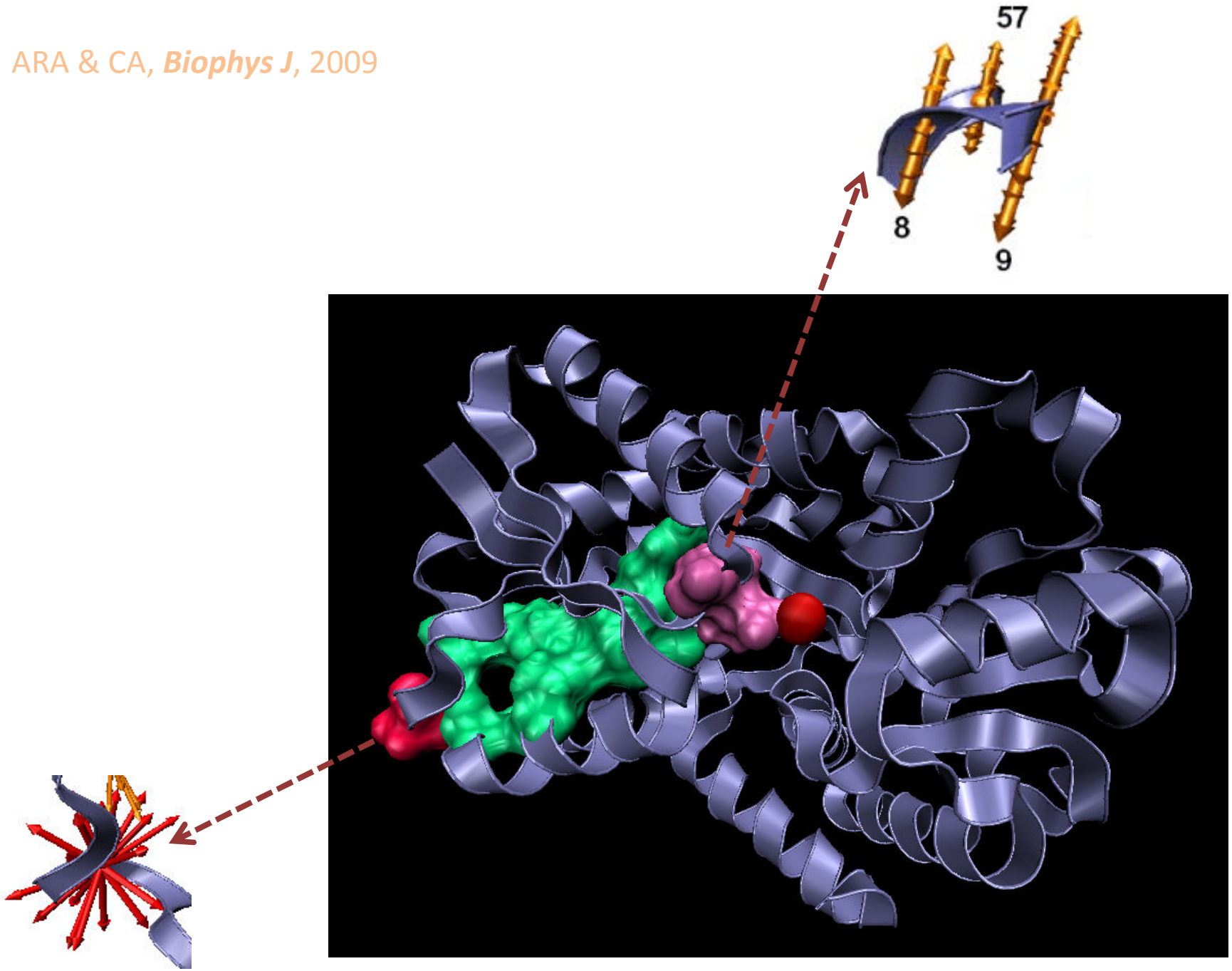
Strong paths
A64-B97 (33.5 %)
A97-B64 (37.8 %)

Frequently used pairs in the interface among all possible pathways between two proteins (44100) locate the hot spots

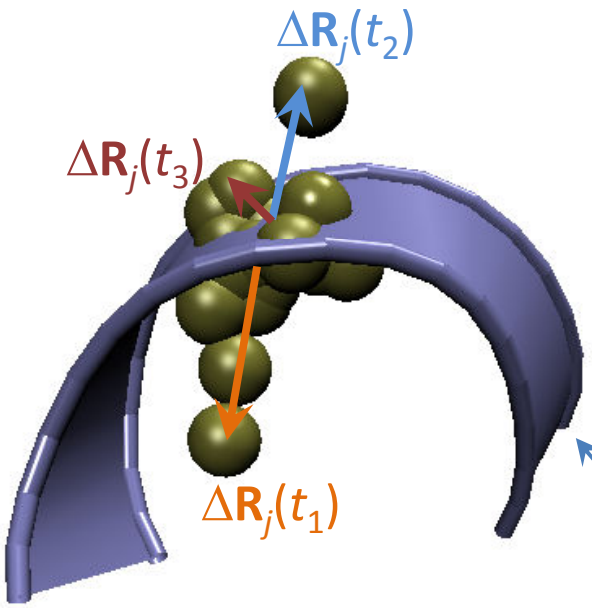
Hot regions
(structural alignment and conservation)
A64-B96
A96-B64

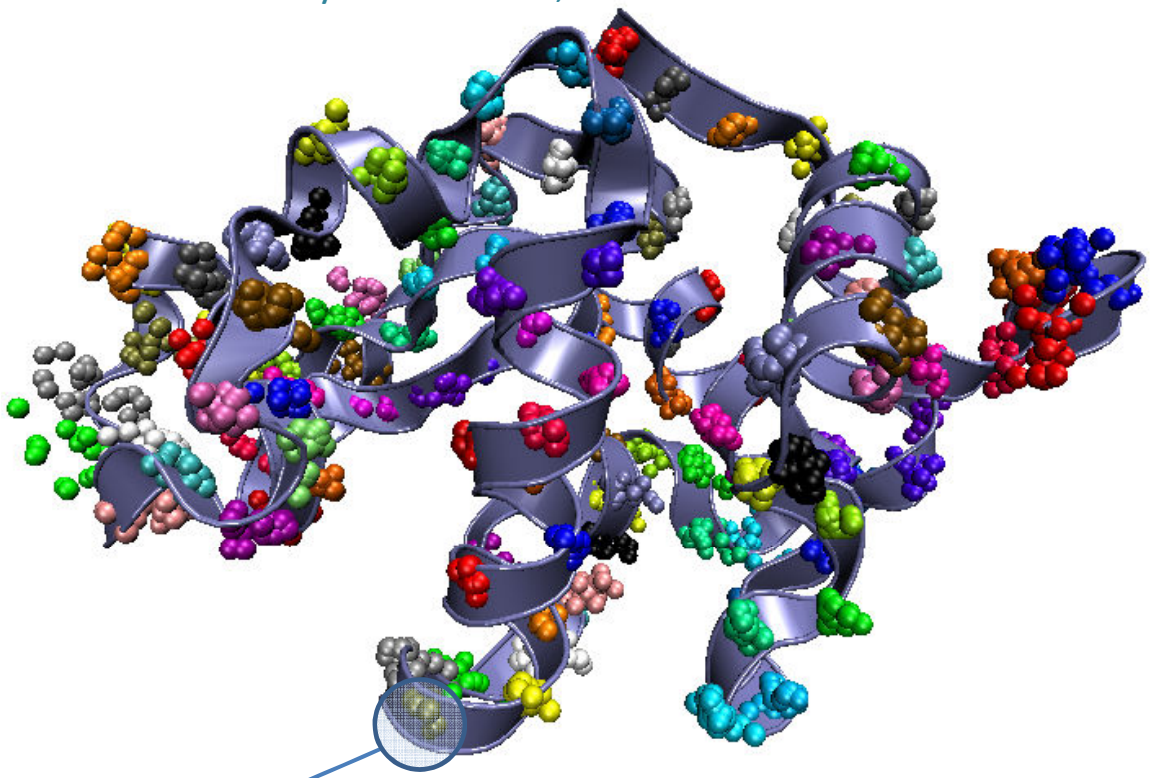
Keskin O, B Ma, R Nussinov, *J Mol Biol* **345**, 1281 (2005).

ARA & CA, ***Biophys J***, 2009



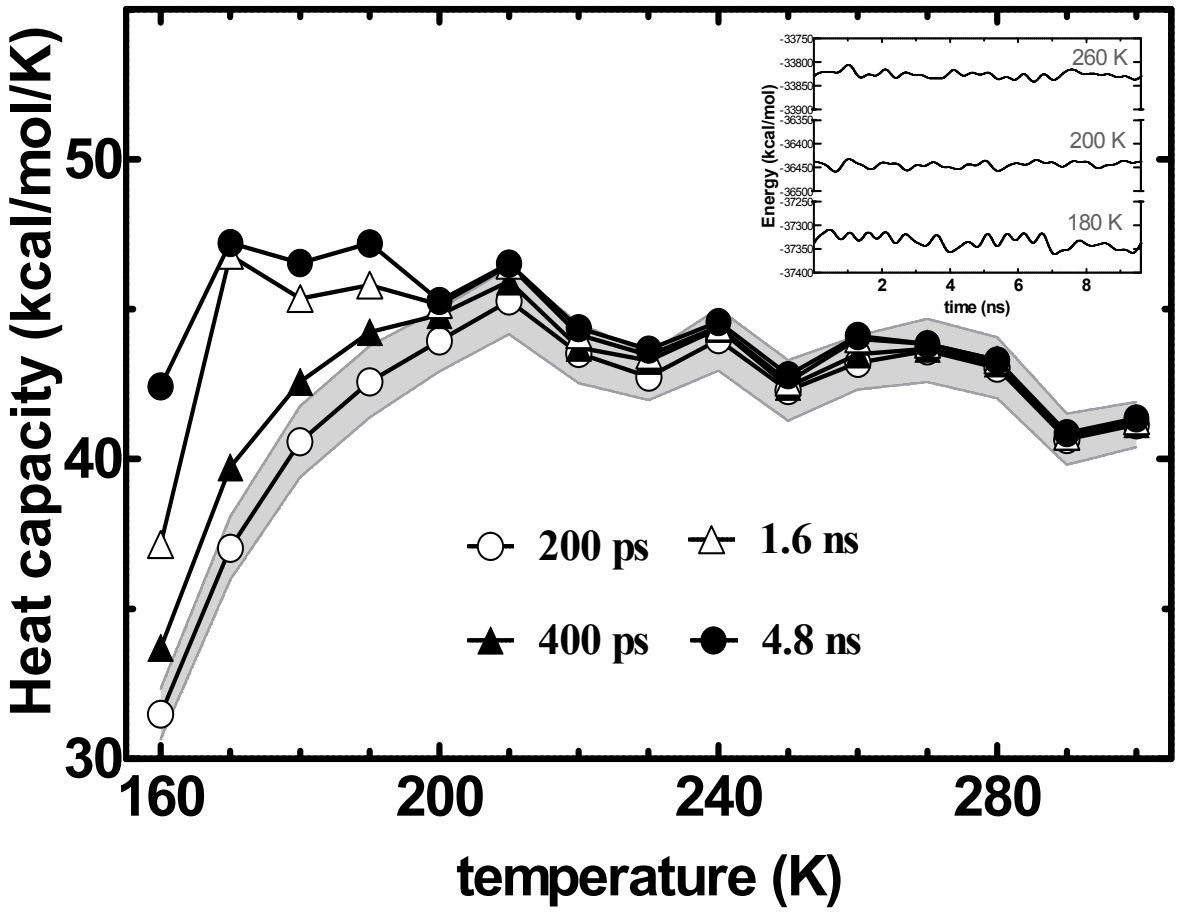
10 to 15 ns long MD trajectories are produced,
spanning the temperature range 160 – 300 K with 10 K increments.
In addition, 100 ns long trajectories are calculated for systems at 160, 180 and 300 K.



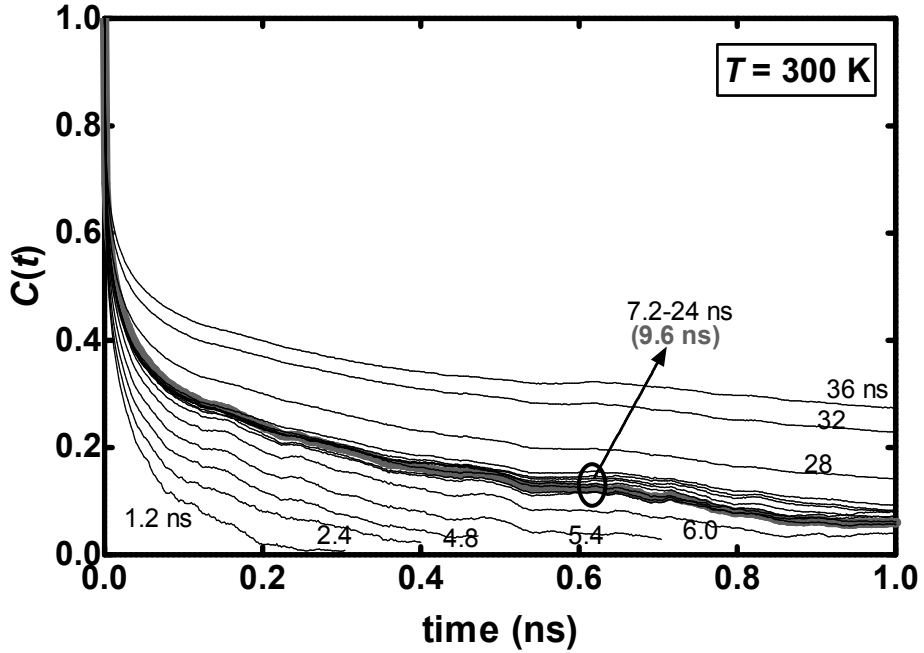


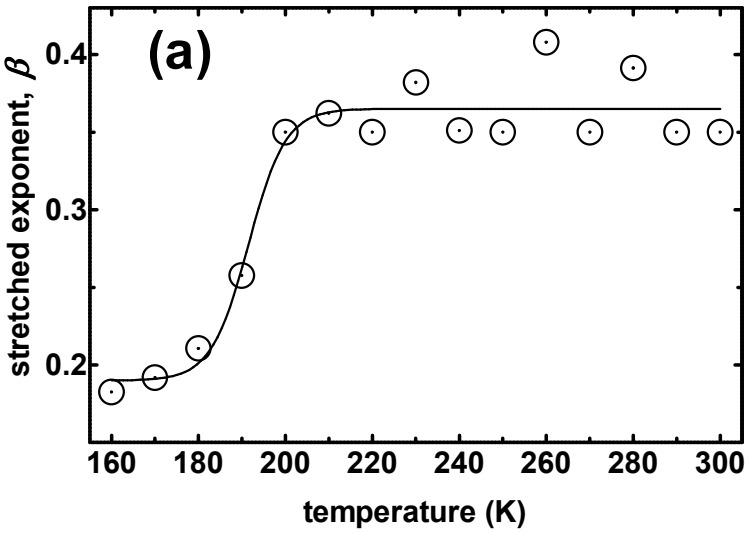
The simulated protein-water complex is comprised of 2271 TIP3P water molecules soaking a single hen egg white lysozyme molecule, a 129 residue protein. Eight chlorine ions are added for charge neutrality. The system has a total of 10281 atoms, prepared using the VMD 1.8.6 program with solvate plug-in version 1.2. Electrostatic interactions are calculated via a partial mesh Ewald method. The system is first subjected to energy minimization until the gradient tolerance is less than 10^{-2} kcal/mol/Å with the conjugate gradients method. The resulting system with the orthorhombic periodic cell of starting dimensions 48 × 45 × 54 Å is used as the initial structure in all MD simulations. All runs are made in the NPT ensemble at 1 atm with the NAMD software. Volumetric fluctuations are preset to be isotropic. Verlet algorithm is used for integrating the equations of motion with a time-step of 2 fs .

$(\langle E^2 \rangle - \langle E \rangle^2)/k_B T^2$

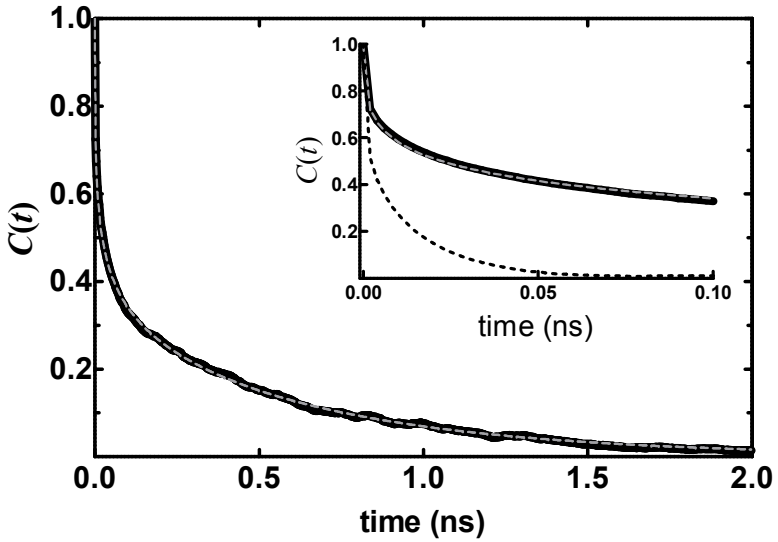


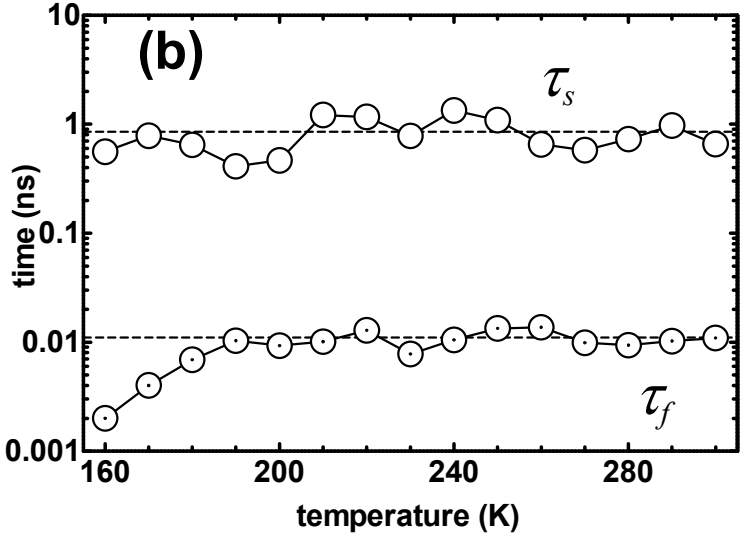
The heat capacity calculated from the whole trajectory (*i.e.*, 9600 ps), as well as the trajectory partitioned into lengths of 200 (48 sets), 400 (24 sets) and 1600 ps (6 sets). We find that at and above 200 K, the value of the heat capacity is the same for all trajectory lengths within the error bounds (as shown by the shaded area). Conversely, at lower temperatures, it depends strongly on the length of the trajectory used in the calculation, with larger fluctuations in longer trajectories, implying that the average energy is not independent of the length of observations.

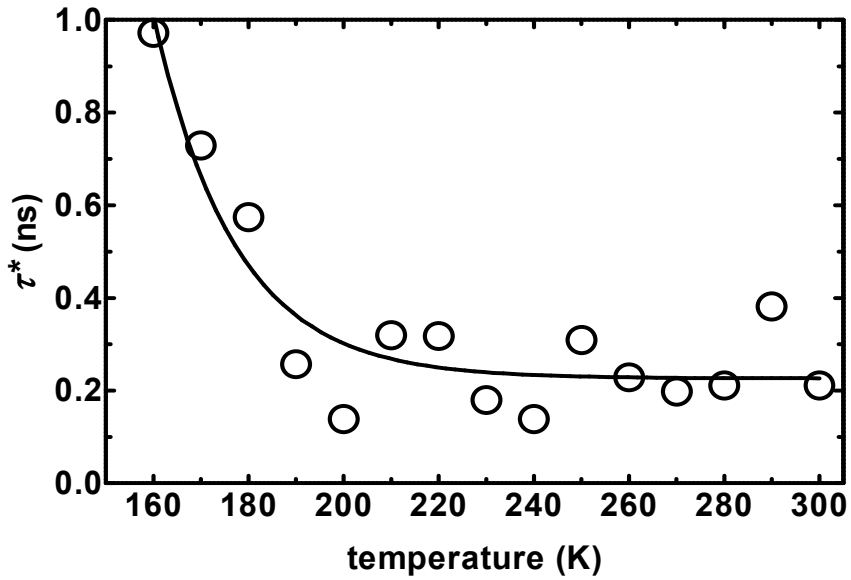


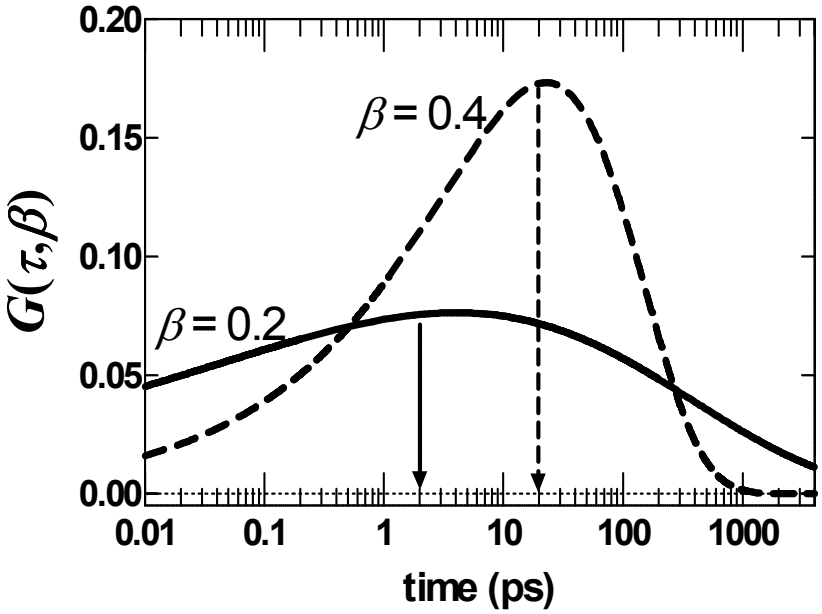


$$C(t) = a\exp\left[-\left(t/\tau_f\right)^{\beta}\right] + (1-a)\exp[-(t/\tau_s)]$$









$$\tau^* = \int_0^\infty C(t)\,dt = a\frac{\tau_f}{\beta}\Gamma\left(\frac{1}{\beta}\right) + (1-a)\tau_s$$

$$\exp\left(-t/\tau_f\right)^\beta = \int_0^\infty \exp(-t/\tau)\,\rho(\tau,\beta)\,d\tau$$

$$\rho(\tau,\beta) = -\frac{\tau_f}{\pi\tau^2}\sum_0^\infty \frac{(-1)^k}{k!}\sin(\pi\beta k)\Gamma(\beta k+1)\left(\frac{\tau}{\tau_f}\right)^{\beta k}$$

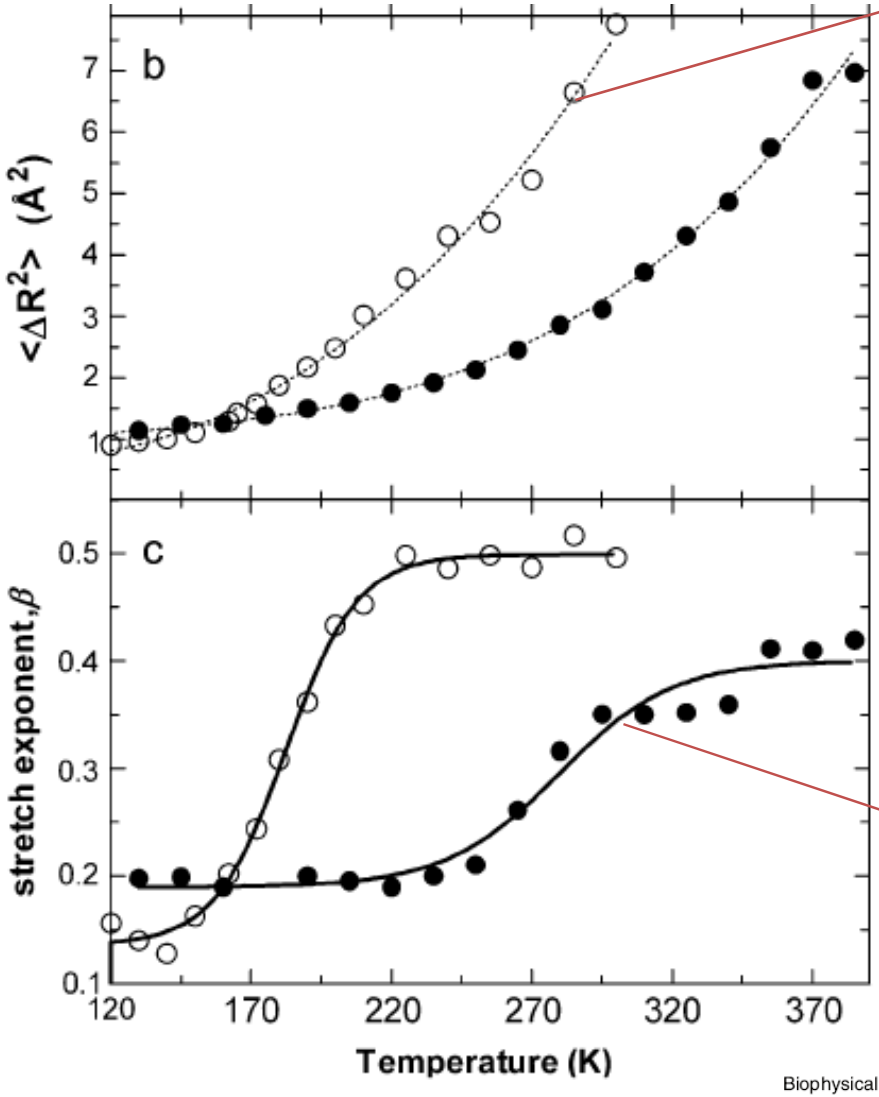
$$G(\tau,\beta) = \tau\,\rho(\tau,\beta)$$

we show that a slower, **nanosecond, time scale motion** in the protein **imposes bounds on** the distribution of the plethora of **fast**, local processes averaging on the picosecond range

2080 Biophysical Journal Volume 97 October 2009 2080–2088

Nanosecond Motions in Proteins Impose Bounds on the Timescale Distributions of Local Dynamics

Osman Burak Okan, Ali Rana Atilgan, and Canan Atilgan*
Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey



water

$$k_p' = \begin{cases} k_p & T < T* \\ k_p\left(1 - \dfrac{k_{pv}^2}{k_p k_v}\right) & T > T* \end{cases}$$

vicinal layer makes the protein more flexible

glycerol

Biophysical Journal Volume 94 January 2008 79–89

How a Vicinal Layer of Solvent Modulates the Dynamics of Proteins

Canan Atilgan, Ayse Ozlem Aykut, and Ali Rana Atilgan
Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

$$\zeta'_{\mathrm{p}} = \begin{cases} \zeta_{\mathrm{p}} & T < T* \\ \zeta_{\mathrm{p}}\left[1 - \dfrac{k_{\mathrm{pv}}^2}{k_{\mathrm{p}}k_{\mathrm{v}}}\left(2\dfrac{\zeta_{\mathrm{pv}}/k_{\mathrm{pv}}}{\zeta_{\mathrm{p}}/k_{\mathrm{p}}} - \dfrac{\zeta_{\mathrm{v}}/k_{\mathrm{v}}}{\zeta_{\mathrm{p}}/k_{\mathrm{p}}}\right)\right] & T > T* \end{cases}$$

vicinal layer slows down the relaxation

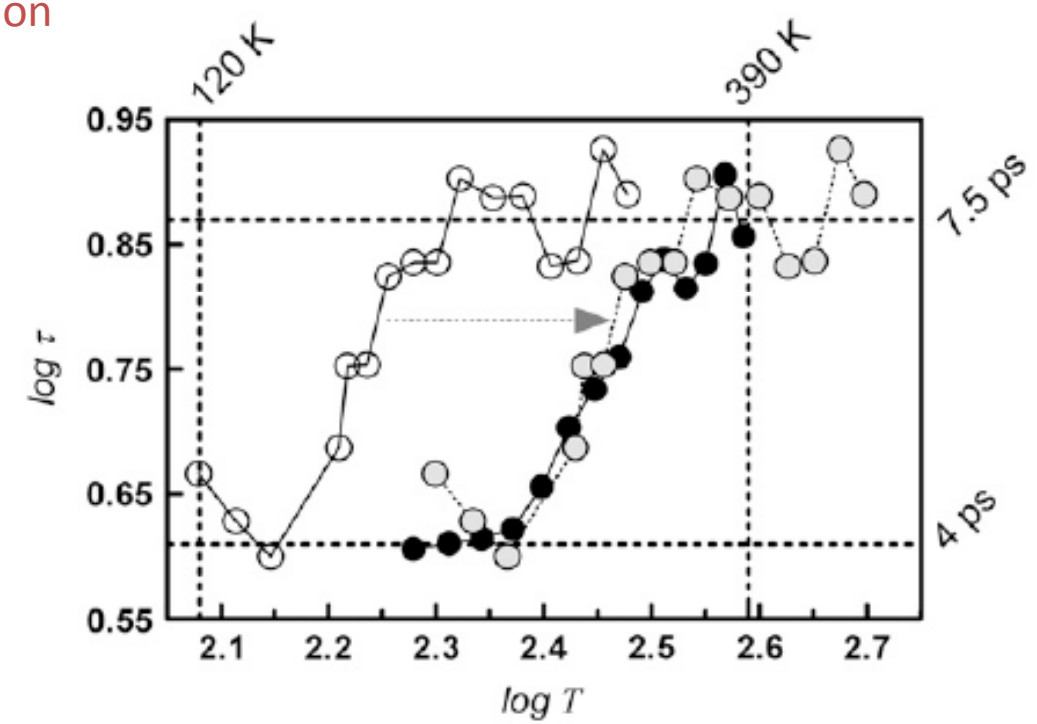


FIGURE 3 The C_α relaxation times as a function of temperature on a double logarithmic scale. (○) Protein-water; (●) protein-glycerol systems. The dotted curve shows the data for the protein-water system with the temperatures shifted by a factor of $T^*_{\mathrm{glycerol}}/T^*_{\mathrm{water}}$ (Eq. 11).

Protein transition temperatures obtained by various approaches

	Protein-water system	Protein-glycerol system
Heat capacity, c_v	164 ± 4 K	277 ± 5 K
Residue fluctuations, $\langle \Delta \mathbf{R}^2 \rangle$	166 ± 6 K	267 ± 3 K
Stretch exponent, β	182 ± 1 K	280 ± 5 K
Average	171 ± 11 K	275 ± 13 K

Accessible surface area and volume of solvent molecules

Molecule	Surface area, a (Å^2)	Volume, v (Å^3)	a/v (Å^{-1})*
Water	31	16	1.9
Glycerol	101	81	1.2
Trehalose†	268 ± 1	271 ± 6	1

*The a/v ratios are the same for probe sizes ranging from 1.4 to 1.6 Å; results displayed here are for probe radius of 1.5 Å.

†Values averaged over different conformers around the torsional angle connecting the rings.

$$\frac{(T^*)_{\text{glycerol}}}{(T^*)_{\text{water}}} = \frac{(a/v)_{\text{glycerol}}}{(a/v)_{\text{water}}}.$$

A **simple** scaling argument