

GLASSINESS AND COUPLING OF TIME SCALES IN FUNCTIONAL PROTEINS

by

OSMAN BURAK OKAN

**Submitted to the Graduate School of Engineering and Natural Sciences in partial
fulfillment of the requirements for the degree of Master of Science**

Sabanci University

Summer 2005

GLASSINESS AND COUPLING OF TIME-SCALES
IN FUNCTIONAL PROTEINS

APPROVED BY:

Assoc. Prof. Canan BAYSAL
(Thesis Supervisor)

Prof. Ali Rana ATILGAN

Prof. Pemra DORUKER

Assist. Prof. Melih PAPİLA

Assist. Prof. Osman Uğur SEZERMAN

DATE OF APPROVAL: 09.08.2005

ACKNOWLEDGEMENTS

It is not possible to fully express my gratitude to several people whom I give more credit in completing my thesis than myself. However, in this little section, I will endeavour to do my best.

I would first like to thank my advisor Canan Baysal for her diligence in proofreading the whole text several times to point out numerous mistakes that has far exceeded my most pessimistic expectations. Moreover, she has been a great fountain of inspiration since I got to know her in my sophomore year. Her relentless pursuit of perfection has deeply effected me albeit I failed to comply most of the time!

My deepest appreciation goes to Ali Rana Atılgan, who has proved to be an excellent co-advisor by incessantly exposing me to his bombardment of fresh ideas on almost everything; from the mathematical tools we have used, to the most basic physical questions like the importance of olfaction in proper tasting of wine.

I am also indebted to my two friends; Irmak Sırer and Deniz Turgut, who has provided me with the priviledge of sharing a lab space with two distinguished individuals with superabundant intelligence. However, no theoretical scientist could achieve much without resorting to the experimentalists. With that said, I express my gratitude to my friends Eren Şimşek, Kerem Gören and Sinan Yördem for their utterly delightful company throughout my fast paced one year masters work.

I can't make a dedication to my parents since they own this thesis as much as I do. However, I

hope that this work will prove to be a means of thanking albeit being in a very humble way considering what they have done for me so far.

At last but not least, I dedicate my part in this thesis to three people to whom I feel extremely grateful. First one is Türker Küçük, from whom I learned the nature of human nature as well as some extra credits like the D-branes in string theory. I also would like to make a dedication to Özge Akbulut, who in my eyes, has always been an epitome of perseverance and perfection. And the last person is Öznur Tastan, who has never left me alone at any moment longer than the Planck-time, notwithstanding the ocean dividing us.

ABSTRACT

GLASSINESS AND COUPLING OF TIME-SCALES IN FUNCTIONAL PROTEINS

Folded proteins are functional at relatively elevated temperatures. Below ca. 190–220 K, proteins may still display the same average structure, but lack function due to the absence of large size fluctuations that increase nonlinearly with temperature. In this state, proteins are similar to polymers displaying glassy behavior, with their disordered, amorphous character and heterogeneous dynamics. We provide evidence that the onset of the relevant fluctuations at physiological temperatures occurs with the residue-wise alteration of the slow –nanosecond time scale– motions due to the activity along the envelope of the energy surface defining the folded protein, and the fast –pico second time scale– motions of the activity along the pockets decorating the folded-state envelope. We investigate this time window with spectral analysis methods to map all the relevant modes of fluctuations. Moreover, the temperature dependence of molecular motions are treated within the context of Fractional Brownian Dynamics. The analysis is based on data describing the relaxation phenomena governing the backbone dynamics derived from molecular dynamics simulations of three proteins in the temperature region 140–330K. The shifts in the weights of fundamental dynamical processes are displayed. Implications on controllability of function are discussed.

ÖZET

İŞLEVSEL PROTEİNLERDE CAMSI HAL VE ZAMAN ÖLÇEKLERİ ARASI ETKİLEŞİMLER

Katlanmış proteinler görece yüksek sıcaklıklarda da işlevseldir. 190-220 K altında ise, proteinler aynı ortalama yapılarını koruyabildikleri halde yüksek sıcaklıklarda gözelenen doğrusal olmayan büyük ölçekli dalgalanmaların olmaması yüzünden işlevlerini yerine getiremezler. Böyle bir durumda proteinler düzensiz ve amorf özellikleri ve heterojen dinamikleri ile camsı polimerlere benzerler. Bu çalışmada, işlevle ilgili dalgalanmaların, nano ve pikosaniye ölçeğindeki moleküler hareketlerin fizyolojik sıcaklıklarda amino asit bazında farklılaşmasıyla ortaya çıktığı gösterilmektedir. Bu zaman penceresinde fiziksel olarak işlevle ilgili tüm hareketler spektral metodlar kullanılarak incelenmektedir. Moleküler hareketlerin sıcaklığa bağlı değişimi de kesirli Brownian hareketi çerçevesinde çalışılmaktadır. Yapılan analizler, moleküler dinamik benzetişimleri sonucunda üç farklı protein için 140-330 K aralığında elde edilen protein ana omurgasının dinamiği ve yapısal hafıza kaybından gelen bulgulara dayanmaktadır. Temel moleküler hareketlerin, toplam dinamiğe olan katkılarındaki kaymalar gösterilmekte, bunların kontrol ve işlevle ilgili olası sonuçları tartışılmaktadır.

TABLE OF CONTENTS

ACKNOWLEDGEMENT.....	iii
ABSTRACT.....	v
ÖZET.....	vi
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
LIST OF SYMBOLS/ABBREVIATIONS.....	xii
1. INTRODUCTION.....	1
2. THEORETICAL BACKGROUND.....	6
2.1. Molecular Dynamics.....	6
2.2. Brownian Motion and Fluctuations.....	8
2.3. Hurst Exponents.....	10
2.4. The Fast Fourier Transform.....	13
2.5. Heat Capacities.....	16
2.6. Relaxations and Autocorrelation Functions.....	17
3. MOLECULAR SYSTEMS.....	20
3.1. Bio-molecular Systems Studied.....	20
3.2. Equilibration and Simulations.....	21
4. RESULTS AND DISCUSSION.....	23

4.1. Phase Transitions.....	23
4.2. Deviating from Random Walk: Dihedral Fluctuations of Surface Residues.....	26
4.3. Spectral Analysis of Trajectories.....	28
4.4. A Three Mode Model for Stretch Exponential Decays.....	32
5. CONCLUSIONS AND FUTURE PROSPECTS.....	35
APPENDIX A: RELATIONSHIP BETWEEN HEAT CAPACITY AT CONSTANT TEMPERATURE AND THE VARIANCE OF ENERGY.....	39
APPENDIX B: FOURIER AND EIGENVALUE SPECTRUMS OF 5PTI, 193L AND 1G6P OVER DIFFERENT TEMPERATURES OF INTEREST.....	40
REFERENCES.....	46

LIST OF FIGURES

Figure 1.1	Rugged topology decorating a hypothetical parabolic potential energy envelope (Generated with function $U(r^1, r^2) = r_1^2 + r_2^2 + \text{random number}$, using Surfer 8.0).....	2
Figure 2.2.1	Non-bonded close contacts and the formation of residue network with a generic potential.....	10
Figure 2.6.1	Two molecules at different times t_1 and t_2 along a trajectory is overlapped with the quaternionic molecule fit.....	18
Figure 3.1.1	CA wire representations of a)193L, b)1G6P and 5PTI. Color coded in secondary structure and Van der Waals surface superimposed.....	20
Figure 4.1.1	(a) Neutron scattering experiments on Rnase A powders (Tsai et al., 2001), (b) Temperature (B)-factors ($= 8\pi^2/3 \langle \Delta R_i \cdot \Delta R_j \rangle$) for the three proteins studied	23
Figure 4.1.2	Temperature dependence of β exponents.....	24
Figure 4.1.3	Temperature dependence of heat capacities.....	25
Figure 4.2.1	79 th residue of 193L is shown along with the side-chain and back-bone dihedrals that are plotted in Figure 4.2.2.....	27
Figure 4.2.2	Dihedral fluctuations for Arg 79 residue of 193L 1)Back-bone 2)Side-chain.....	28
Figure 4.3.1	Evolution of the eigenvalue spectrum of 193L at different temperatures...	29
Figure 4.3.2	Evolution of packing density in time and temperature domains.....	30
Figure 4.3.3	Power spectrum of the displacements of C_α atoms pertaining to Lysozyme at 158 K (a) and (b) 298 K.....	31

Figure 4.4.1	Stretch exponential functions as described in Table 4.4.1 are drawn together.....	33
Figure B.1	Eigenvalue Spectra for 5PTI at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.2 increment in contour levels.....	40
Figure B.2	Eigenvalue Spectra for 193L at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.2 increment in contour levels.....	41
Figure B.3	Eigenvalue Spectra for 1G6P at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.2 increment in contour levels.....	42
Figure B.4	Power Spectra for 5PTI at five different temperatures. Color coded for real eigenvalues (Green (Smaller) to White (Bigger)) with 0.03 increment in contour levels.....	43
Figure B.5	Power Spectra for 193L at five different temperatures. Color coded for real eigenvalues (Green (Smaller) to White (Bigger)) with 0.03 increment in contour levels.....	44
Figure B.6	Power Spectra for 1G6P at five different temperatures. Color coded for real eigenvalues (Green (Smaller) to White (Bigger)) with 0.03 increment in contour levels.....	45

LIST OF TABLES

Table 2.4.1	Properties of the Fourier Transform.....	13
Table 4.1.1	Transition temperatures obtained from three methods.....	26
Table 4.2.1	Hurst exponents for four surface residues of 193L at different temperatures.....	28
Table 4.4.1	Three models according to the first equality in equation 4.4.1, and the best fitted KWW parameters to these models due to the second equality in equation 4.4.1.....	33

LIST OF SYMBOLS/ABBREVIATIONS

A_i	Dimensionless multiplicative coefficient
a	Acceleration
β	Exponent for stretch exponential function
C	Heat Capacity
C_n	n^{th} Fourier coefficient
$C(t)$	Normalized autocorrelation function of a dynamical entity
D_{fps}	Dimension of the probability space
D_{fts}	Fractal dimension of time series
E	Energy
F	Force on a rigid body object
\mathbf{F}	Random force matrix
Γ	Kirchoff (connectivity) matrix
H	Hurst exponent
\mathbf{H}	Mass matrix dor hydroynamics interactions
k	Multiplicative factor
m	Mass of a rigid body
R	Distance covered (range)
\mathbf{R}	Trajectory matrix of best fitted structures
ΔR	Displacement from the equilibrium position
r	Generalized coordinate
S	Standard deviation
T	Temperature

t	Time
τ	Characteristic relaxation time
U	Potential energy function
x	Distance
v	Velocity

1. Introduction

Temperature characteristics of internal motions and the structural relaxations in folded proteins are crucial for unveiling the nature of function at the molecular level. Moreover, a functional protein assumes the concerted motion of its residues, thanks to the structural flexibility at ambient temperatures. Therefore, the determination of time-scales and pertinent frequency windows is of utmost importance before embarking upon devising plausible models for protein dynamics. Both experimental (Daniel et al., 1998; Zaccai, 2000; Tsai et al., 2001; Fenimore et al., 2002; Pal et al., 2002) and theoretical (Vitkup et al., 2000; Dvorsky et al., 2000; Baysal and Atilgan, 2002; Tarek and Tobias, 2002; Tournier et al., 2003; Baysal and Atilgan, 2005) effort which have been devoted to the glassiness, function and relaxation phenomena of those polypeptides are abound. Of particular interest is the coupling of time-scales approach, which conjectures that a functional protein at physiological temperatures owes its functionality to the coupling of concurrent structural changes occurring at different time-scales (Baysal and Atilgan, 2005). This coupling brings out the feature that the structural relaxation of a protein molecule has Kohlrausch-Williams-Watts (KWW) type (stretch exponential) dependency (Kohlrausch, 1874; Williams and Watts, 1970). This functional form can be seen as a manifestation of the superposition of single exponential decays with different characteristic decay times;

$$\exp \left[\left[-\frac{t}{\tau} \right]^\beta \right] = \sum_n A_i \exp \left[-\frac{t}{\tau_i} \right] \quad (1.1)$$

However, given a stretch exponential decay, the corresponding single exponential decomposition is prohibitively tedious for large n . Moreover, it is not possible to determine n precisely by just examining the stretch exponential function.

Stretch exponential behavior is not the only function type that is ascribed to the complex physical processes that follow non-exponential trends as well. However, the basic question of how a folded protein loses its memory in the solvent entails a closer perusal of the energy landscape in the neighborhood of local minima and corresponding topographical features as shown in Figure 1.1

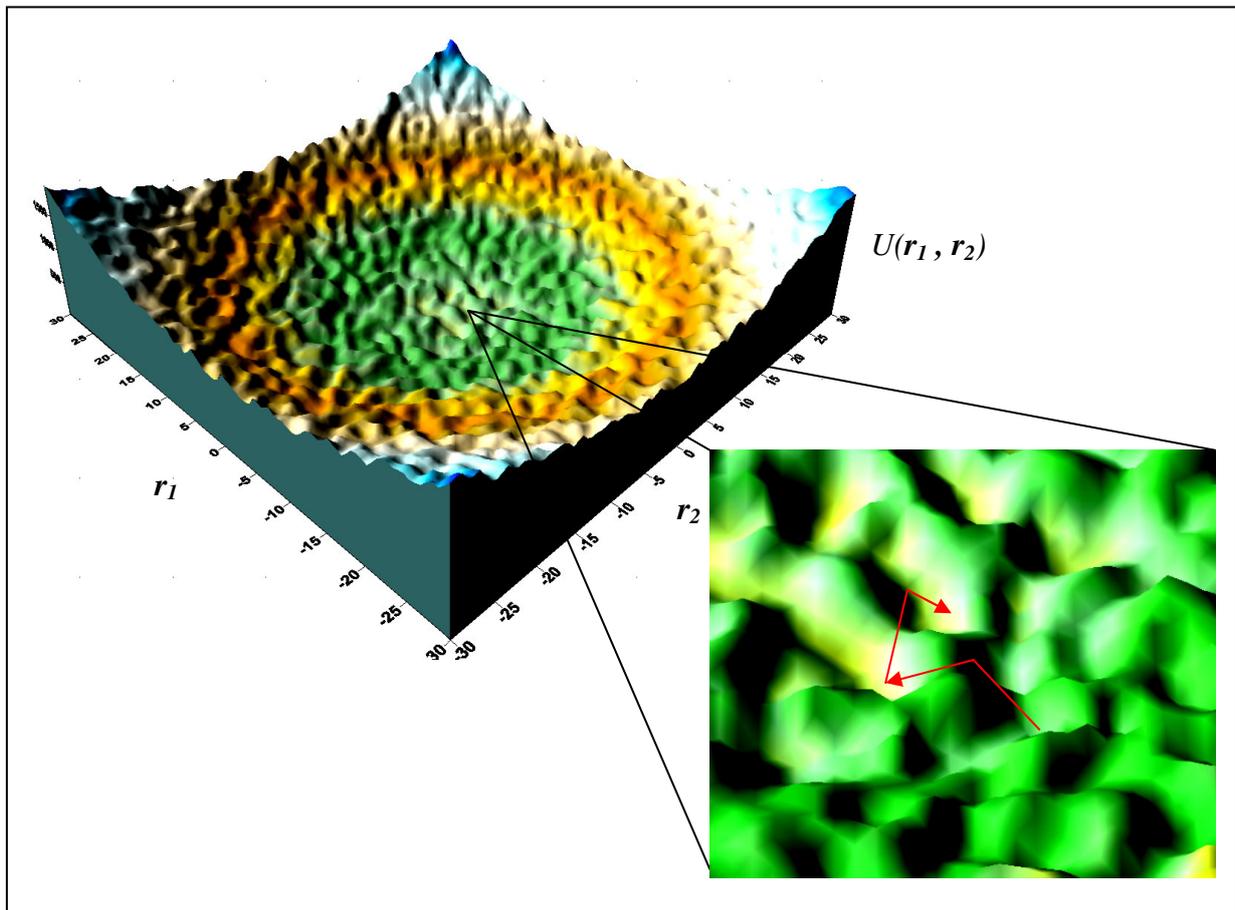


Figure 1.1 Rugged topology decorating a hypothetical parabolic potential energy envelope (Generated with function $U(r_1, r_2) = r_1^2 + r_2^2 + \text{random number}$, using Surfer 8.0)

It has been shown that the protein dynamics about the folded state assume nonlinear motion on a multi-basin landscape (Garcia, 1997). An illustration of this model is depicted above in Figure 1.1. With this approach, it is possible to construct a direct analogy between a weakly adsorbed diffusing particle and a protein as monitored in the configurational space, while establishing a direct link to the Fractional Brownian Motion (FBM). Moreover, frequent hopping between the local minima decorating the potential envelope is shown to be decisive in determining the time-scales of fluctuations around the equilibrium, which turns out to be occurring at picoseconds (ps) and nanosecond (ns) timescales. Resolving the individual modes of motion allows one to describe the relaxation of a physical quantity with simpler function prototypes (Equation 1.1 constitutes an example) as well as shedding light into the dynamics of the folded state.

Most of the mathematical models of relaxations were launched with the intention of building more robust expressions that would capture the main features of dielectric polarization decays on the grounds that the latter have posed problems since the emergence of classical electrodynamics. Of the numerous attempts, the Cole-Davidson and Williams-Watt models have been two successful epitomes of generic functions to fit non-exponential decays comparison of which may be found elsewhere (Lindsey and Patterson, 1980). Reaction kinetics is another long studied field from which to borrow models and function types is lucrative. A model due to Frauenfelder is relevant for systems that bear rugged energy landscape (Frauenfelder, 1988). However, another example which has appeared more recently predicts a Lorentzian form for the Fourier-transformed time correlation of potential energy assuming a landscape with fractal properties for the case of plastocyanine (Carlini et al., 2002). The latter result is shown to comply with FBM.

This work aims to elaborate on Equation 1.1 and the physical implications of KWW type dependency, along with a decomposition in terms of single exponential functions. Biphasic nature of protein dynamics is explicitly shown through the β exponents which are obtained from a series of molecular dynamics simulations (MD) of three protein molecules at a wide temperature range (130K-340K). Spectral analysis of real space trajectories is carried out to reveal the underlying time-frequency windows and their interactions.

This work partly aligns itself with the “residue network” approach (Atilgan et al., 2004; Haliloglu, 1997; Yilmaz and Atilgan, 2000) where the whole protein is treated as a network of interacting amino acids. Time and temperature characteristics of the network structure and the fluctuation modes is juxtaposed with the spectral properties to propose a roadmap for carrying out the expansion given by Equation 1.1 robustly.

In what follows, Chapter 2 will give a layout of the theoretical background. Therein, section 2.1 gives a reasonably detailed exposition of MD simulations. In section 2.2 Brownian motion and fluctuations are discussed along with the residue network approach. In the following section (2.3) Hurst exponent and FBM are briefly introduced. Computational procedures for heat capacities from potential energy trajectories constitutes the section 2.4. Section 2.5 describes the Fast Fourier Transform (FFT) and recapitulates the basic mathematical properties of it. In the concluding section of Chapter 2 autocorrelation functions are described within the context of structural relaxations in proteins. Chapter 3 provides detailed information on the molecular systems modelled and the simulation details. Results and their interpretation are given in Chapter 4. It starts with the description of the “protein glass transition” as monitored through three physical parameters. In the section 4.2, fluctuations of dihedral angles pertaining to the surface residues are shown to exhibit a fractional Brownian

nature which follows from the deviations in the Hurst exponents. In the subsequent section 4.3, the origins of these deviations are investigated through the spectral analysis of the atomic displacements.

In the last section, three-mode model for describing the stretch exponential behaviour as a superposition of single exponentials is introduced and the physical implications are discussed within the context of protein structural relaxations. Chapter 5 is a recapitulation of the findings as well as the new directions that the results render necessary to follow. Protein glass transition and protein dynamics are put into a broader perspective with a novel way of looking at the weighted coupling of time-scales.

2. Theoretical Background

2.1 Molecular Dynamics

MD simulations uses Newton's laws of motion to simulate the dynamics of many-body systems under force-fields. Force fields provide approximations for the forces on each atom. Newton's second law is then explicitly solved for each atom i in the system for a given time step;

$$m_i \frac{d^2 r_i}{dt^2} = -F_i \quad (2.1.1)$$

r_i describes the displacement (or position) vector while F_i gives the total force acting upon the i^{th} particle in the system. Equations of motions are time reversible and the atomic trajectories created are deterministic, while the forces on each atom satisfy the minus gradient law of potentials;

$$F_i = [-\nabla U(r_1, r_2, \dots, r_n)]_i = -\frac{\partial U(r_1, r_2, \dots, r_n)}{\partial r_i} \quad (2.1.2)$$

The very first MD simulation that originated from the work of Alder and Wainwright using a hard sphere model (Frenkel, 2001; Leach, 1996) did not employ a continuous force field. The implementation of continuous force-fields had to wait until 1964, when Rahman simulated the motion of Argon atoms (Frenkel, 2001; Leach, 1996). MD methods currently have a myriad of applications and are routinely used for the simulation of biological molecules. However, the focus is mainly on the simulation of folded structures which assume stable tertiary structures rather than modelling of the folding phase, small molecular systems being exception (Duan and Kollman, 1998).

The integration of the equations of motions are carried out with a suitable finite difference scheme since there is no analytical solution in closed form for a many body problem. MD methods are deterministic; that is, by numerical integration one can go from acceleration to velocity and from velocity to positions and vice versa. All algorithms are based on the assumption that the dynamical variables are analytic functions which can be approximated by Taylor series expansions. The Verlet algorithm is worth mentioning on the grounds that its the adapted integration scheme for the MD simulations reported herein.

Verlet scheme is a finite difference method that exploits the series expansion of displacements. We expand displacements backward and forward shifted by h in time and put together the resulting expression to solve for $x(t+h)$ and $x(t-h)$.

$$\begin{aligned} x(t + \Delta t) &= x(t) + v(t) \cdot \Delta t + \frac{1}{2} \cdot a(t) \cdot \Delta t^2 + \frac{1}{6} \frac{d^3 x(t)}{dt^3} \Delta t^3 + O(h^4) \\ x(t - \Delta t) &= x(t) - v(t) \cdot \Delta t + \frac{1}{2} \cdot a(t) \cdot \Delta t^2 - \frac{1}{6} \frac{d^3 x(t)}{dt^3} \Delta t^3 + O(h^4) \end{aligned} \quad (2.1.3)$$

Velocities and third order terms cancel out each other leaving accelerations and 4th order error terms behind. This implies that the Verlet algorithm which is apparently 2nd order is more precisely a 4th order algorithm (Eberly, 2004). The ability to solve x in both directions in time brings out the feature that the Verlet algorithm is suitable for modelling conservative systems. Combining this with the Newton's second law, one ends up with;

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{1}{2} \cdot a(t) \cdot \Delta t^2 + O(h^4) \quad (2.1.4)$$

Velocities may still be estimated with various methods one of which produces an estimate at

half-step forward :

$$v(t + 1/2\Delta t) = \frac{x(t + \Delta t) - x(t)}{\Delta t} \quad (2.1.5)$$

On the other hand, a modified version, the so-called velocity-Verlet (Swope et al., 1982) algorithm, explicitly calculates the velocities as well as the positions and accelerations simultaneously. This is achieved through a three-step computation which involves;

$$\begin{aligned} x(t + \Delta t) &= r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \\ v(t + \Delta t) &= v(t) + \frac{1}{2}a(t)\Delta t + \frac{1}{2}a(t + \Delta t) \end{aligned} \quad (2.1.6)$$

2.2. Viscoelasticity, Brownian Motion and Fluctuations

Classical molecular theory of viscoelasticity describes the microscopic dynamics of a polymer network as a system of spherical Brownian particles connected by springs under incessant bombardment of solvent molecules (Doi and Edwards, 1986). The equilibrium dynamics of such a macromolecular system in solution might well be described by the Langevin equation,

$$\frac{d\mathbf{X}}{dt} = \mathbf{H}^{-1}(-\mathbf{\Gamma}\mathbf{X} + \Sigma\mathbf{F}) \quad (2.2.1)$$

where $\mathbf{\Gamma}$ is the Kirchoff matrix and \mathbf{H} stands for the inverse mobility matrix, whose diagonal accounts for the solvent friction coefficients due to Kirkwood (Kirkwood, 1949). The construction of $\mathbf{\Gamma}$ and \mathbf{H} is such that the k_{ij} element is the stiffness of the harmonic potential

between beads i and j , whereas the h_{ij} element of matrix \mathbf{H} gives the hydrodynamic interaction between these. Finally, the random forces are given by the \mathbf{F} matrix and the solvent bombardment is assumed to be a Gaussian process with white noise (Johnson noise) satisfying;

$$\langle f_i(t) f_j(t') \rangle = \delta_{ij} \delta(t - t') \quad (2.2.2)$$

The main distinguishing trait of a white noise process is the delta function nature of the time autocorrelations.

With elastic network models, one modifies the Γ matrix such that the Hookean springs are assumed between each residue as long as the two are in their close proximity below a certain cut-off distance. This is based on the assumption that the time evolution of the inter-atomic distances around their equilibrium position follows a Gaussian distribution (Bahar et al., 1997).

All residues are described by a generic bead and force constant in calculating the interactions irrespective of molecular weight and chemical composition. The elastic network, as established by the Γ matrix is totally isotropic, thus reducing the number of eigenmodes from $3N-6$ to $N-1$. This allows the construction of single harmonic potential based residue networks. Figure 2.2.1 is an illustration of the residue network generation by spatial proximity;

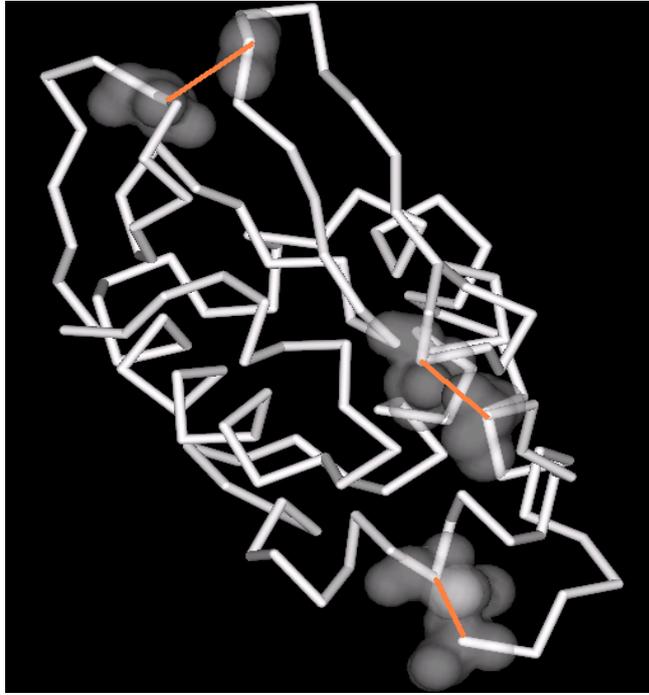


Figure 2.2.1. Nonbonded close contacts and the formation of residue network with generic potentials

2.3. Hurst Exponents

The trajectory of a particle in a liquid undergoing random collisions of the solvent molecules was termed as the Brownian motion in the previous section. The main characteristic of such a random walk is the 0.5 exponent which relates the elapsed time to the distance covered;

$$R = kT^{0.5} \quad (2.3.1)$$

Where R is the range and k is a multiplicative factor. Hurst exponent is a natural generalization of $H = 0.5$ exponent to cover the interval $[0,1]$, leaving 0.5 as a special case. However, it is validated by rescaling the range with the standard deviation S (Steeb, 2002).

$$\frac{R}{S} = kT^H \quad (2.3.2)$$

Moreover, it describes a generalized Brownian motion bearing a fractal distribution with inherent scaling and self-similarity. FBM is a Gaussian process, the prefix of which follows from the fractal dimensionality that stems from an Hurst exponent $H \neq 0.5$. Since 0.5 corresponds to the mid-point of the $[0,1]$ interval we may identify two more cases of Fractional Brownian Motion.

i. $0 < H < 0.5$ (Antipersistency): This H range is described as the pink-noise and the dominant characteristic of such a process is the mean-reverting tendency. Any time series assuming H within this interval is more volatile than a random walk and is ergodic.

ii. $H = 0.5$ (Independency): Classical Brownian motion with derivative giving rise to white-noise. This value of H defines the random walk and the pertinent statistical parameters of the time-series are time invariant.

iii. $0.5 < H < 1$ (Persistency): Such a process is indicative of long-memory effects such that the distance covered within time t is longer than it is for a pure Brownian motion. It immediately follows that the persistency favors trend formation; i.e. onset of a decrease is more likely to be followed by a decreasing trend, and vice versa. The noise for a persistent time series is named as black-noise. When H lies within this range, we expect to see positive correlations across time-scales.

What Hurst exponent provides is therefore a unitless measure to distinguish between random phenomena and deterministic processes exposed to random perturbations. Moreover, it is a means to assess the volatility of a data set. The apparent nonstationarity of FBM processes require a careful consideration for spectral decomposition. Although time-frequency analysis is still possible, unveiling the scaled stationarity lurking within time scales provides a more robust means to quantify FBM trajectories.

Relation of the Hurst exponent to the spectral methods and ergodicity is clearly deduced from the two identities ;

$$D_{fps} = \frac{1}{H} \quad D_{fts} = 2 - H \quad (2.3.3)$$

Here, D_{fps} gives the dimension of the probability space whereas D_{fts} is the fractal dimension of the time series. It is clear that the probability space is two dimensional for a random walk thus allowing the phase space to be fully spanned. However, for trajectories which assume $H > 0.5$ a portion of the phase space is not accessible at any time. Hurst exponents also give clues to the convergence and the power spectrum of the Fourier decomposition of the time series via the well established expression which gives the decay rate of the Fourier coefficients as;

$$\frac{1}{f^{(2H+1)}} \quad (2.3.4)$$

As the Hurst exponent converges to zero, the Fourier components tend to decay in a slower fashion, crowding the power spectrum.

Synthesis of FBM and the inverse problem of determining H from a given trajectory is now computationally well established, thanks to the exploitation of wavelet methods. Detailed coverage of such methods is beyond the scope of this work. For more details one can resort to Abry and Sellan's excellent paper (Abry and Sellan, 1996).

2.4 The Fast Fourier Transform (FFT)

Fourier transform, other than being a mathematical curiosity, provides the necessary link between the two facets of the same physical phenomenon that is; the time domain and the frequency domain. Moreover, Fourier transform is an invertible linear mapping that allows one to switch between the two domains back and forth. Using frequency representation, the continuous Fourier transform and its inverse are defined as;

$$F(t) = \int_{-\infty}^{\infty} G(f)e^{-2\pi ft} df \quad \text{and} \quad G(f) = \int_{-\infty}^{\infty} F(t)e^{2\pi ft} dt \quad (2.4.1)$$

Owing to the properties of the plane-wave kernel, there might be certain symmetries inherent in both domains. A list of the symmetry and domain dependent properties as adopted from Numerical Recipes (Press et al., 1992) is given below:

Time Domain	Frequency Domain
Real $F(t)$	$H(-f) = [H(f)]^*$
Complex $F(t)$	$H(-f) = -[H(f)]^*$
Even $F(t)$	$H(-f) = H(f)$ (even)
Odd $F(t)$	$H(-f) = -H(f)$ (odd)
Real and Even $F(t)$	$H(f)$ is real and even
Real and Odd $F(t)$	$H(f)$ is complex and odd
Complex and Even $F(t)$	$H(f)$ is complex and even
Complex and Odd $F(t)$	$H(f)$ is real and odd

Table 2.4.1 Properties of the Fourier Transform

However, very few scientific problems yield closed form analytical expressions as solutions rendering the use of discrete FT (DFT) inevitable. DFT, by definition takes N points and maps them into N new points in the transform domain. The number of independent points (which in turn corresponds to the eigenfrequencies) in the transformed domain is equal to the number of sampled points. Thus, for transforming complicated signals, larger trajectories are needed to extract the information content with the least amount of loss. The discrete transformation could be represented in closed form as;

$$F(n) = \sum_{k=0}^{N-1} \exp(-2\pi ink) G_k \quad G(k) = \sum_{n=0}^{N-1} \exp(2\pi ink) F_n \quad (2.4.2)$$

We can also explicitly write the above approximations in matrix form. The matrix obtained thereby is a special kind of Vandermonde matrix whose columns are orthogonal to each other ensuring the unitarity of the transformation (Meyer, 2004). From this point of vantage, the requirement for matrix multiplication entails DFT to be an $O(n^2)$ algorithm with multiplication operations carried over complex numbers. However, it was judiciously shown by Danielson and Lanczos that with two $N/2$ point transform it is possible to replicate the original N point transform (Press et al., 1992). Unless the subsequent divisions of N ends up in an odd number, one can go on dividing the function domain until all sub-domains are grouped as odd and even single points, which brings DFT back to the family of $O(n \log(n))$ algorithms. FFT is actually an alias for transform algorithms which operate with $O(n \log(n))$ floating point operations and could be implemented in both recursive and non-recursive fashion. The total number of points sampled therefore is chosen as integer powers of 2. Symbolic mathematics languages such as Matlab and Mathematica bears functions which automatically completes the number of points to the next power of 2 such that $N < 2n$. However this trick has to be incorporated into the subroutines in compiler level languages.

Fourier transform of MD trajectories constitute complex arrays, which does not leave much room for physical interpretation. However, it is well known that the modulus squared FTs are related to the power spectral density (PSD) or the power spectrum. One of the simpler ways of extracting the power spectrum from FT is the periodogram method which is defined as;

$$\begin{aligned}
 P(0) &= P(f_0) = \frac{1}{N^2} |C_0|^2 \\
 P(f_k) &= \frac{1}{N^2} \left[|C_k|^2 + |C_{n-k}|^2 \right] \quad k = 1, 2, \dots, ((N/2)-1) \\
 P(f_c) &= P(f_{N/2}) = \frac{1}{N^2} |C_{N/2}|^2
 \end{aligned} \tag{2.4.3}$$

However, one immediate problem is the assesment of the goodness of signal reconstruction with finite data points, which in turn imposes limitations on the frequency content. Shannon sampling frequency or the Nyquist frequency is a pertinent mathematical measure for the sampling rate to achieve an uncompromised recovery of the Fourier components. It is defined as;

$$f_c = \frac{1}{2\Delta} \tag{2.4.4}$$

For a bandwidth limited signal, sampling more than what the Nyquist frequency implies is a futile effort, since above that rate all components have zero power density in the power spectrum. However, sparser sampling is also troublesome on the grounds that it gives rise to information lost and the aliasing phenomenon. With sparser sampling, we mean any signal which is not bandwidth limited up to the Nyquist frequency. Aliasing is also termed as false translation since the higher frequency components fold back into the $[-f_c, f_c]$ closed interval, distorting the final waveform. From a computational point of view, the ideal case is to

transform a discretized sample whose highest frequency component is in the close proximity of the critical frequency.

There is, unfortunately, no panacea to remedy the disadvantages accompanying aliasing, leakage and related signal processing problems simultaneously, but one might still choose to apply low-pass filtering techniques before implementing a transformation. From the perspective of MD simulations, our current knowledge of the molecular modes whose frequency based classification is readily accessible by spectroscopy science provides important benchmarks to choose the optimum sampling rate for simulations.

Of course, not every trajectory is Fourier transformed and the computational intensity of equation of motion integrations usually establish themselves as the most decisive factor for choosing the density of sampling. However, from a spectroscopic point of vantage, the highest wavenumber which is physically important is about 4000 cm^{-1} .

For a protein, the residue-wise modal distribution is obtained via a discrete Fourier transform of the coordinate trajectories of each C_α atom. The finite time of the simulations imposes certain limitations on the Fourier transforms, such as the insertion of higher frequency artificial terms and peak broadening, both of which scale inversely with the total duration of simulation time.

2.5 Heat Capacities

Heat capacity is a fundamental thermodynamical entity which exhibits discontinuity at the transition temperature for a second order transition (for a first order transition it diverges to

infinity). Fortunately, statistical mechanics provides means to exploit the variance of energy throughout a MD trajectory so as to calculate C ;

$$C = \left[\frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2} \right] = \frac{\langle E - \langle E \rangle^2 \rangle}{k_B T^2} \quad (2.5.1)$$

Proof of this identity is quite straightforward and is given at Appendix 1. However, using the rightmost identity is a more robust way from a computational point due to the possibility of round-off errors in calculating $\langle E^2 \rangle$ and $\langle E \rangle^2$ first hand.

Heat capacity can also be calculated with pure thermodynamics by differentiating the energy function once we fit a differentiable function to the time evolution of the energy data with a reasonably small lack of fit. However, the length of the trajectories renders the possibility of identifying and best-fitting functions like polynomials of order n to the simulation data as prohibitively difficult.

2.6. Relaxation and Autocorrelation Functions

The autocorrelation function is a measure of how the memory is retained in time for a specific quantity. In general, an autocorrelation function is a monotonically decreasing function which starts from one at time $t = 0$ and converges to zero. The characteristic time for a system to lose the significant part of its memory is called the relaxation time. There is no universal procedure for the calculation of relaxation times and they are generally obtained by a substitution that cancels out the explicit time unit.

The resulting set of points in an autocorrelation function may also assume different shape characteristics, limiting the options for fitting functions. This work centers around the stretch exponential behavior, which is described by a Kohlrausch-Williams-Watt function (Kohlrausch, 1874; Williams and Watts, 1970);

$$C(t) = \exp\left[-\left[t/\tau\right]^\beta\right] \quad (2.6.1)$$

What is presented in this work is the correlation of a fluctuation vector, evaluation of which is quite tedious and requires conscientiousness. The structural relaxations are studied through the time- and temperature-dependent properties of the fluctuation vector $\Delta\mathbf{R}$ of C_α atoms of a given protein from its mean structure. For obtaining the mean structure we first remove the overall tumbling embedded in the MD trajectories; that is rigid body rotation and translation. This is achieved as follows: For any given t' ps sub-trajectory, one starts with a best-fit superposition of the recorded structures to the initial structure by minimizing the root mean

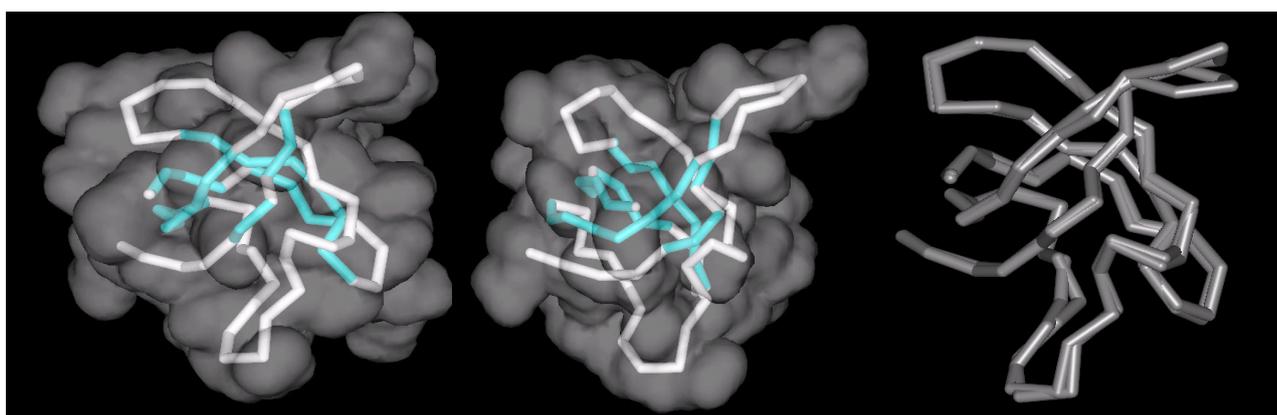


Figure 2.6.1 Two molecules at different times t_1 and t_2 along a trajectory is overlapped with the quaternonic molecule fit.

square deviations of the C_α atoms via quaternionic rotations and center of mass back-translation (Heisterberg, 1990). The use of quaternions renders the implementation of

composite rotations extremely simple and reliable. Rather than multiplying rotation matrices for each rotation, we multiply the quaternions whose arithmetic is very simple. After multiplying quaternions corresponding to each rotation in the quaternion space, we move back into the rotation matrix representation to modify the coordinates of each atom. One such fit of two structures is shown in fig. 2.4.1.

One then computes the average structure, $\langle \mathbf{R}(T) \rangle$, from the t' best-fitted structures. Here $\langle \cdot \rangle$ refers to time averaging. In the last step of the fit, quaternionic superposition of the recorded structures to this average structure is carried out. After subtracting rotational and translational degrees of freedom, the trajectory is reduced to a time series of internal motions for a given protein. Each structure of this final trajectory is denoted by $\mathbf{R}(t, T)$ and the coordinates of the i^{th} residue are denoted by $\mathbf{R}_i(t, T)$. The fluctuation vector for a given residue i at a given time t for a given protein at temperature T , $\Delta \mathbf{R}_i(t, T)$ is thus the difference between the position vectors for the i^{th} residue of the best-fitted and the average structures. The autocorrelation of the fluctuation vector is given as;

$$C(t) = \frac{\langle \overline{\Delta \mathbf{R}(0) \cdot \Delta \mathbf{R}(t)} \rangle}{\langle \overline{\Delta \mathbf{R}(0)^2} \rangle} \quad (2.6.2)$$

3. Molecular Systems

3.1 Biomolecular Systems

The initial structures for the three proteins were retrieved from Protein Data Bank (PDB) (Berman et al., 2000). Along with their PDB abbreviations these are namely; the Bovine Pancreatic Trypsin Inhibitor (PDB code: 5PTI) (Vlodawer, 1984), the Hen Egg White Lysozyme (PDB code: 193L) (Vaney, 1996), and the Cold Shock Protein from the Hyperthermophilic Bacterium *Thermotoga Maritima* (PDB code: 1G6P) (Kremer et al., 2001) Molecular visualizations of the three proteins are given below at the secondary structure level.

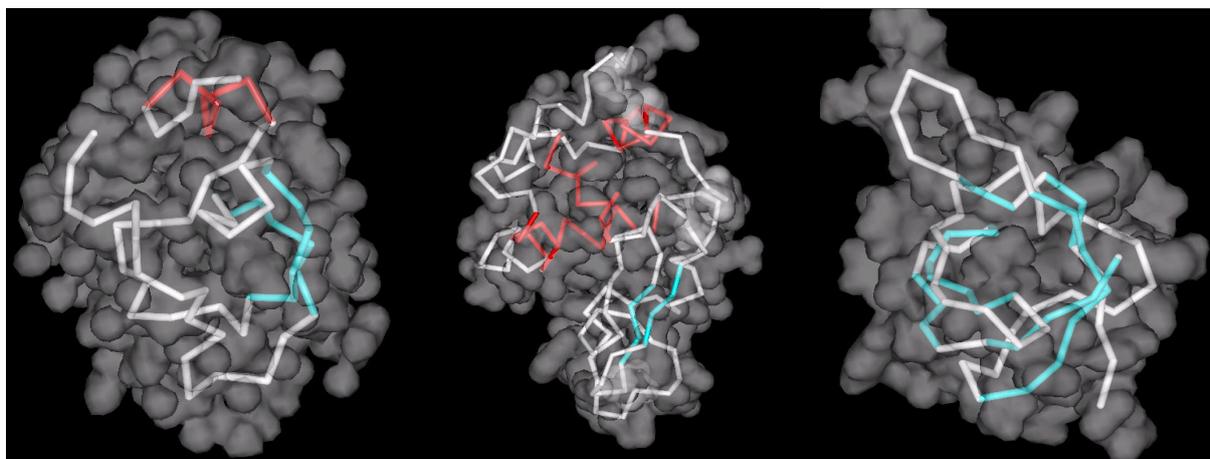


Figure 3.1.1. CA wire representation of a) 5PTI b) 193L c) 1G6P. Color coded in secondary structure and Van der Waals surface superimposed.

Molecular dynamics (MD) simulations were carried out with InsightII 2000 (Accelrys, 2001) commercial licence. The two proteins, 193L and 1G6P were soaked in a 7 Å thick water shell whereas 5PTI assumed a 6Å thick hydration layer in vacuo. Together with the structural water molecules reported in the PDB structure, this treatment leads to a total of 703 solvent

molecules for 5PTI, whereas for 193L and 1G6P the number of solvent molecules are 2490 and 1842, respectively. All atoms are treated explicitly during the simulations. Before the equilibration runs, the corresponding energy of each system is minimized with a multi-method minimization procedure due to the unphysical values of the energy derivative stemming from the close contacts formed by the packing algorithm and the uncertainties imposed by the X-ray diffraction, from which the atomic positions are obtained. For 5PTI, minimization is comprised of subsequent conjugate-gradients iterations until the energy-derivate reaches down to 7.5 kcal/mol. For the cases of 193L and 1G6P, minimization runs start with the steepest-descent method until the energy-derivative is reduced below 1000 kcal/mol and switches to conjugate-gradients until it reaches below 50 kcal/mol. Futher refinement is achieved through the Newton-Raphson scheme until the derivative is brought below 0.5 kcal/mol.

3.2. Equilibration and Simulations

All equilibration runs are of 200 ps in duration. Equilibration phases assume the simulation temperature of corresponding MD trajectories without resorting to thermal annealing. Adapted velocity control mechanism is chosen to be Andersen which assigns velocities to the molecules picked at random in accordance with the Maxwell-Boltzmann distribution.

For the cases of 193L and 1G6P all simulation are of 2 ns duration irrespective of the simulation temperature. The CVFF (Dauber-Osguthorpe et al., 1988) force field implemented within the Accelrys InsightII 2000 (San Diego, CA) software package is used throughout structure refinement and the subsequent MD runs of all three proteins. A modified version of the single point charge model is adopted for describing water molecules in the system

(Berendsen et al., 1987). Group-based cutoffs are set to 10\AA . A switching function is used, with the spline and buffer widths set to 1.0 and 0.5\AA , respectively. Integration algorithm is the velocity-Verlet method implemented with velocity and bond constraints from the RATTLE scheme (Andersen, 1983). Integration step is taken to be 1 fs for 5PTI and 2 fs for the remaining two. Data are recorded every 2 ps for all runs. Temperature control is achieved by the extended system method of Nose (Nose, 1984). The MD runs for 5PTI are of length 2.0–2.8 ns, depending on the temperature: 2.0 ns for $T < 230$ K, 2.4 ns for $230 < T < 290$ K, and 2.8 ns for $T > 290$ K. Also, second independent MD runs of duration 2.0 ns are carried out for $T > 290$ K. This yields trajectories of length 2.0–4.8 ns depending on the temperature. For 193L and 1G6P all runs are carried out for 2 ns with 20 K increments in the range 138 K–338 K and 158 K–358 K respectively. After all runs have been completed, the trajectories are parsed into portions with 200 ps temporal window size so as to obtain multiple data sets.

4. Results and Discussion

4.1. Phase Transitions

The phenomenon of dynamical phase transitions in proteins has first manifested itself in neutron scattering studies. Those studies are significant, since they clearly show the effect of water as solvent in dampening of the molecular flexibility and fluctuations below a temperature regime of 190-220 K (Zaccai, 2000; Tsai et al., 2001, Gabel et al., 2004). When the mean squared fluctuations are plotted against temperature one observes a change in the slope when the protein is hydrated in water. However, for completely dehydrated powders, such a transformation is absent (Figure 4.1.1 (a)). This is in accordance with our findings for each protein. Residue fluctuations versus temperature from our simulations are shown in Figure 4.1.1 (b). The transition temperature is extracted via the extrapolation of the two regimes marked by a change in the slope of residue fluctuations to find the intersection point. Transition temperatures calculated via this method is listed in Table 4.1.1

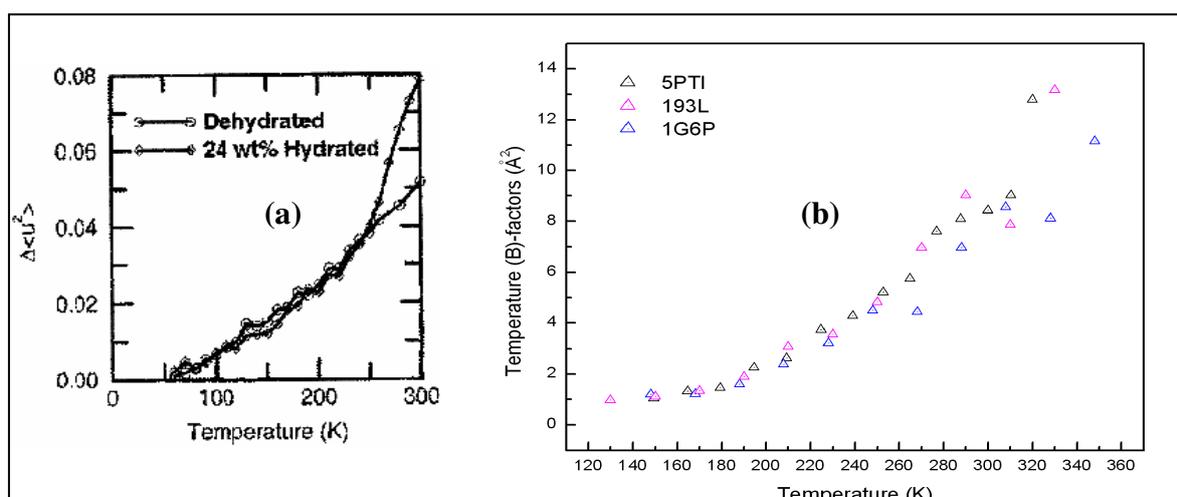


Figure 4.1.1 (a) Neutron scattering experiments on Rnase A powders (Tsai et al., 2001)
(b) Temperature (B)-factors($= 8\pi^2/3 \langle \Delta R_i \cdot \Delta R_j \rangle$) for the three proteins studied in this work

Structural relaxations, as defined by the autocorrelation functions are indicative of a universal feature of proteins. Our results show that the β exponents in stretch exponential decays vary between 0.2 and 0.4 irrespective of the sequence information. This is quite remarkable considering the algebraic simplicity of the model employed. Moreover, this unique property of relaxations imply that the concertion of molecular fluctuations with different time scales could well be decisive for putting the protein function in dynamical perspective.

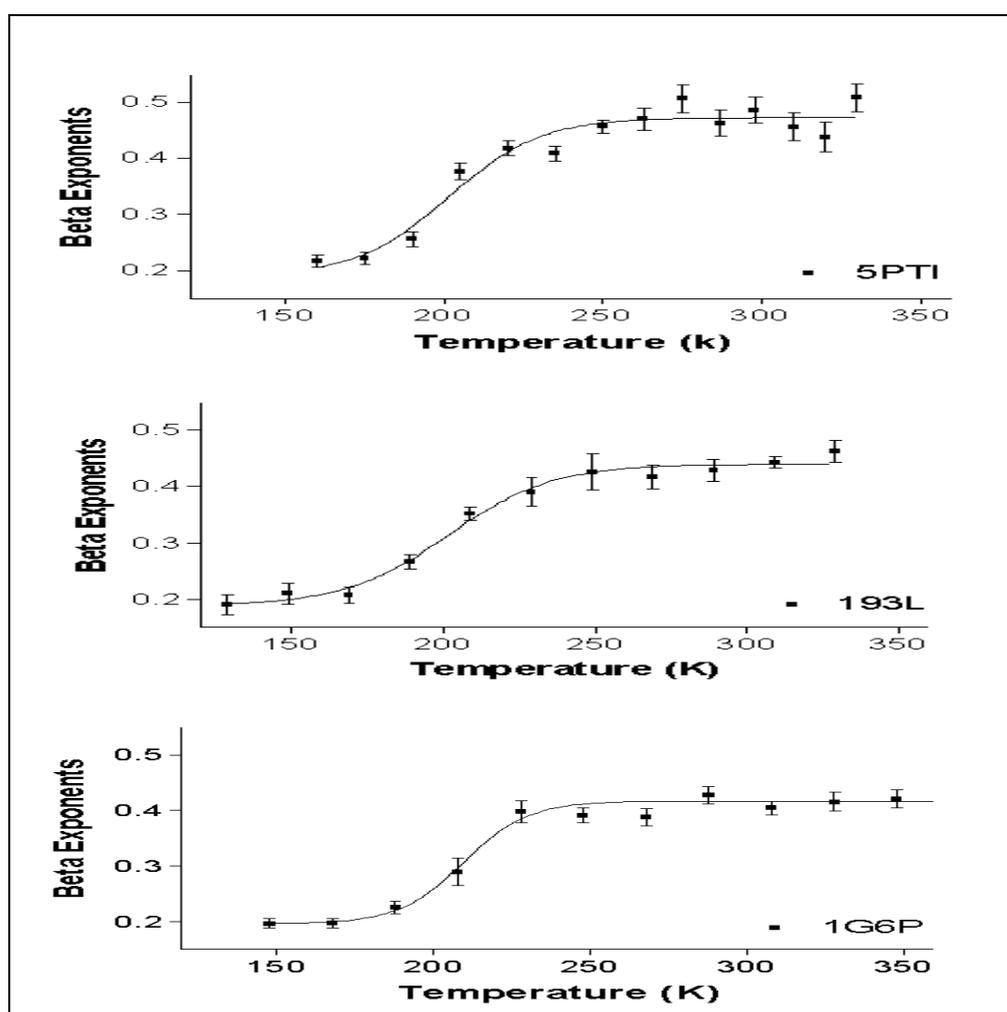


Figure 4.1.2. Temperature dependence of β exponents for three proteins

It is obvious that the sigmoidal curves obtained from the temperature dependency of β might be fitted with an analytic function to locate the inflection point at which the transition occurs.

This gives an estimate for the glass transition temperature. It has been shown that the Boltzmann-sigmoidal curve is a perfect candidate for fitting transitions of this shape (Baysal and Atilgan, 2005). For our case it is given by the formula;

$$\beta = \beta_0 + \frac{\beta_1 - \beta_0}{1 + \exp\left[\frac{(T_0 - T)}{\lambda}\right]} \quad (4.1.1)$$

Heat capacities as shown in Figure 4.1.2 also exhibit a similar trend with broader transition region. However the onset of transition is in accordance with the transitions in the β exponents. In fact, fits to these data are also made with a Boltzmann-sigmoidal curve using the formula;

$$C = C_0 + \frac{C_1 - C_0}{1 + \exp\left[\frac{(T_0 - T)}{\lambda}\right]} \quad 4.1.2$$

and the estimated transition temperatures are also listed in Table 4.1.1.

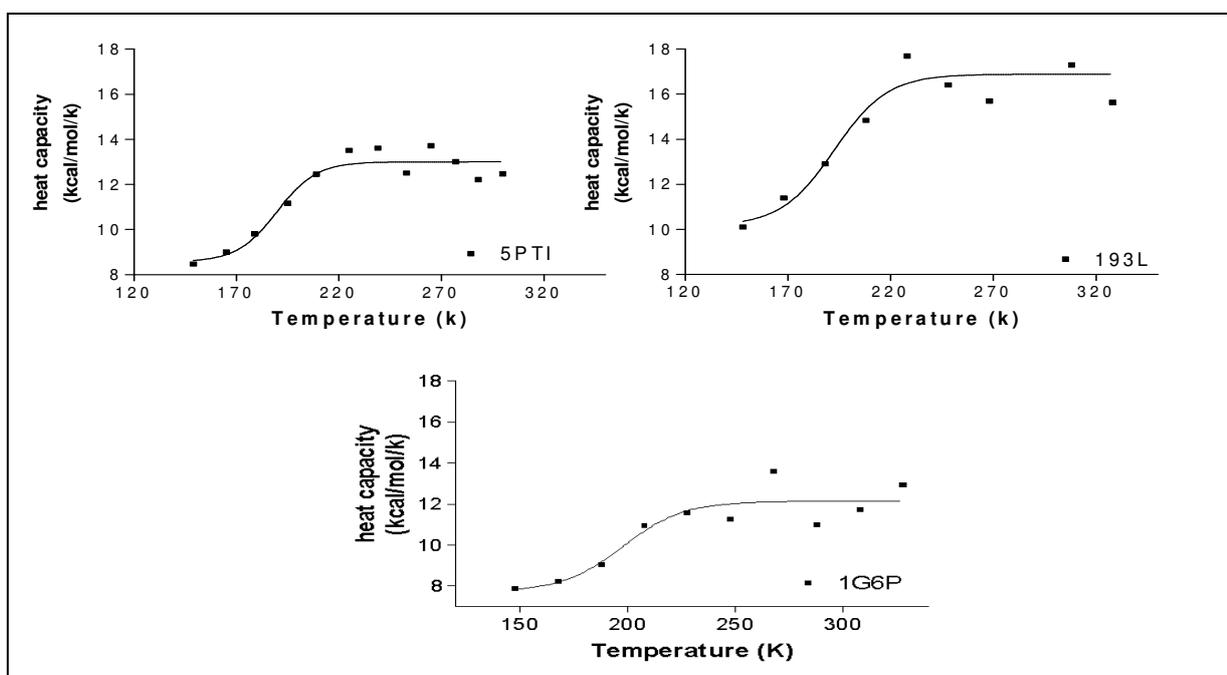


Figure 4.1.3. Temperature dependence of heat capacities

	5PTI	193L	1G6P
From Residue Fluctuations	190 K	193 K	198 K
From Stretch Exponentials	191 K \pm 5	202 K \pm 3	210 K \pm 6
From Heat Capacity	191 K \pm 4	187 K \pm 5	197 K \pm 4
Average	191 K \pm 9	194 K \pm 8	202 K \pm 10

Table 4.1.1 Transition temperatures obtained from three methods

A comparison of the results obtained from the three methodologies reveals that the transition temperatures predicted from three physical parameters with temperature, all point to the same value for a given protein. Of particular interest is the ca. 10 K higher transition of temperature obtained for 1G6P. This is a hyperthermophilic member of the cold shock protein family, which displays a remarkably high melting point of 360 K (Kremer et al., 2001), which is reflected to the protein glass transition temperature as well.

4.2 Deviating From Random Walk: Dihedral Fluctuations of Surface Residues

Characterizing the transitory dynamics of structural relaxations, we also monitor the fluctuations of side chain and backbone dihedral angles. Residues whose dihedral fluctuations are monitored are chosen from the solvent protein interface so that the solvent effects are fully accounted for. For the case of 193L, dihedral trajectories pertaining to four surface residues are chosen. The location of the 73th residue is illustrated below in Figure 4.2.1. Full trajectories for Arg-73 and pertinent Hurst exponents averaged over four residues are given in Figure 4.2.2 and Table 4.2.1 respectively.

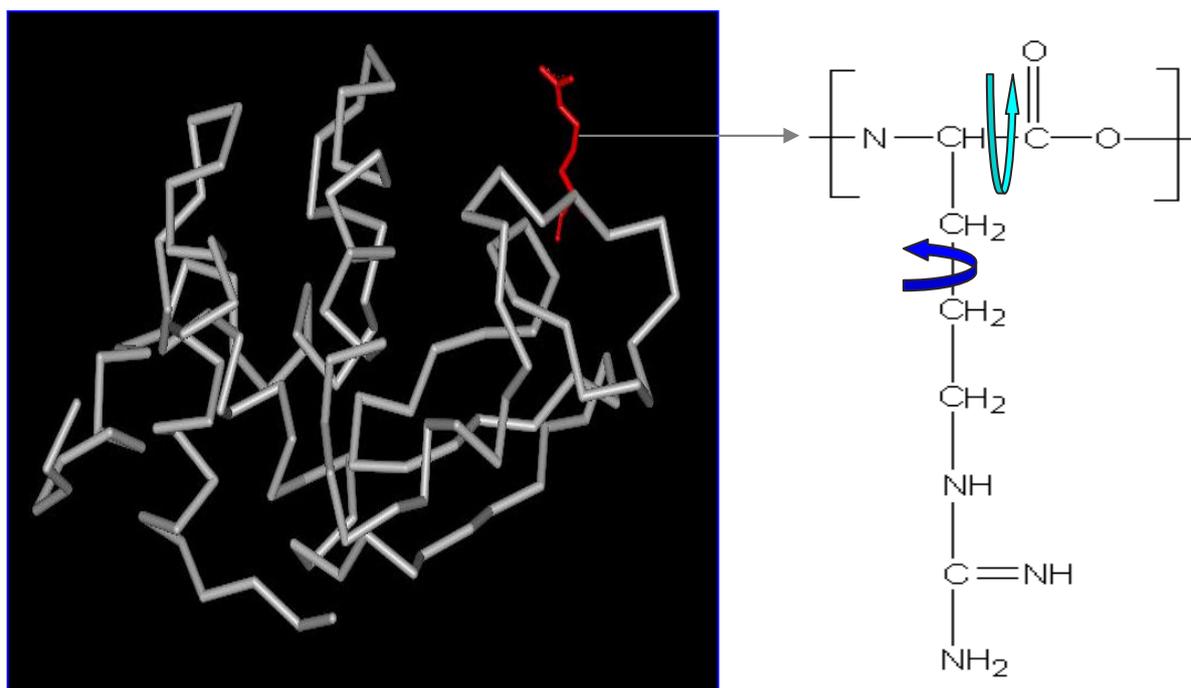


Figure 4.2.1 73th residue of 193L is shown along with the sidechain and backbone dihedrals that are plotted in Figure 4.2.2

The fluctuations of side chain and backbone dihedral angles, show an increasing trend in amplitude as the temperature increases. However, multiphase behavior is more pronounced for the dihedral fluctuations pertaining to side chains, which is indicative of a hopping behavior between local minima that becomes dominant, as the portion of phase space spanned gets larger. Hurst exponents of the dihedral trajectories show that both backbone and side-chain relaxations do form impersistent time-series with exponents $H < 0.5$. However as temperature increases there emerges an apparent trend that the exponent approaches to 0.5

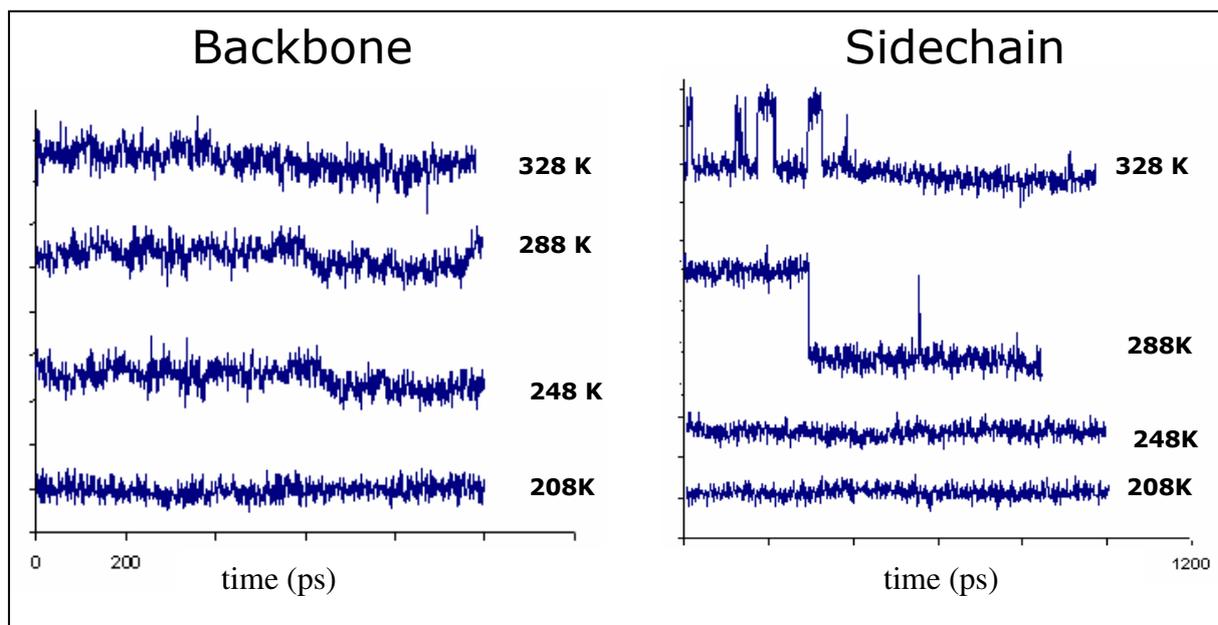


Figure 4.2.2 Dihedral fluctuations for Arg-73 residue of 193L 1) Backbone 2) Side Chain

Temperature	Side Chain	Backbone
208 K	0.03	0.01
248 K	0.08	0.07
288 K	0.14	0.12
328 K	0.22	0.15

Table 4.2.1 Hurst Exponents averaged over four surface residues of 193L at different temperatures

4.3 Spectral Analysis of Trajectories

The eigenvalue spectrum is provided in Figure 4.3.1 for the case of 193L protein at three different temperatures. Eigenvalue spectrum of the inverse of Kirchoff matrix reveals that the eigenmodes are quite stationary both in terms of magnitudes and distribution in time for a given temperature. This result also supports our findings with packing density and implies

that the network structure largely preserves its connectivity. At higher temperatures modal distribution of fluctuations. loosen its stationarity in time concurrent with the changes in packing density. However, one does not observe the emergence of new modes as the temperature is increased. This indicates that at the middle temperature range, the eigenmodes remain stationary, whereas new collective modes arise as the weights of each process start to show residue-wise changes. Thus the energy surface of a protein does not change with temperature. This picture is expected to change with the onset of unfolding. In fact, the spectrum at 330 K for 193L is clustered at higher values.

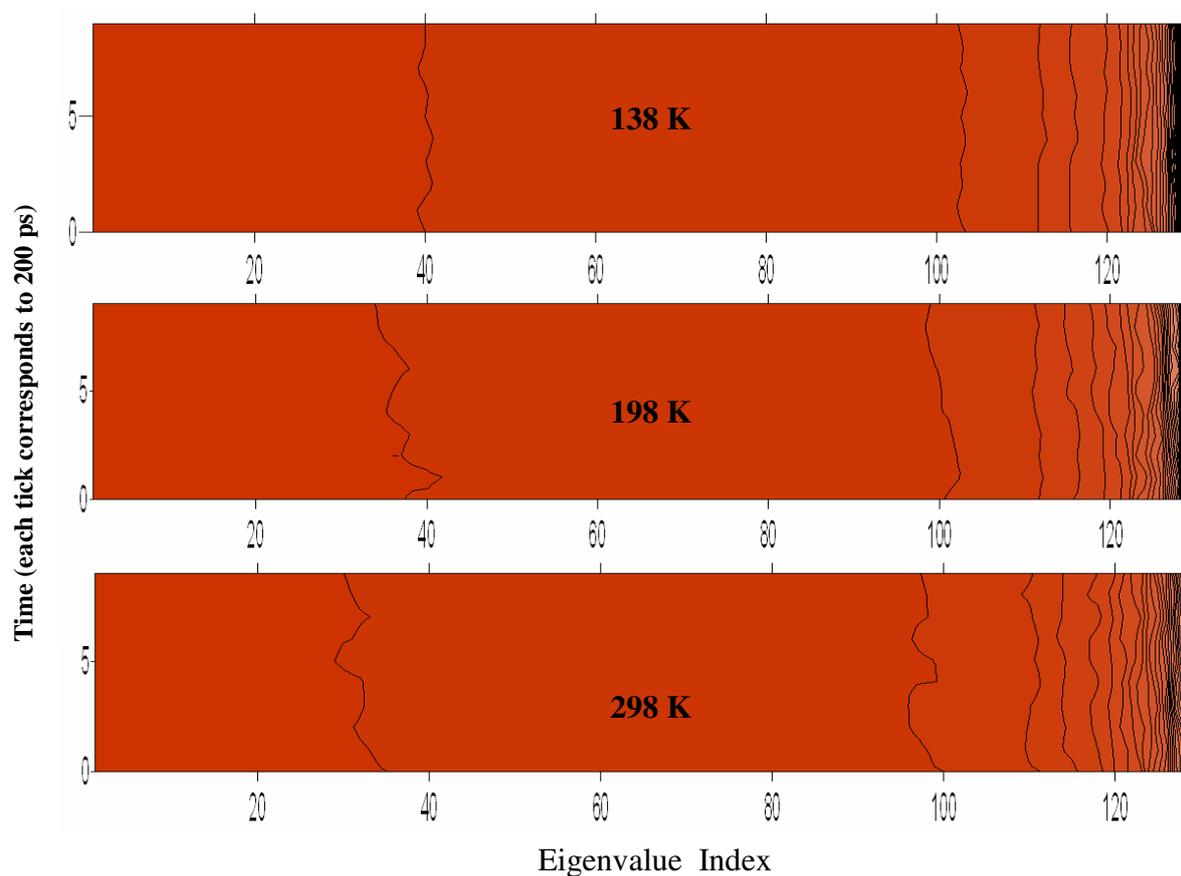


Figure 4.3.1. Evolution of the eigenvalue spectrum of 193L at different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.1 increment in contours.

A close inspection of Figure 4.3.1 shows that below the transition temperature, the eigenvalues are clustered at higher values so there is a thicker band at the corresponding color coded region. At intermediate temperatures however, eigenvalues are seemingly more uniformly distributed. This is more clearly monitored in the evolution of packing density of each residue, exemplified by 193L at different points in time as well as at different temperatures (Figure 4.3.2).

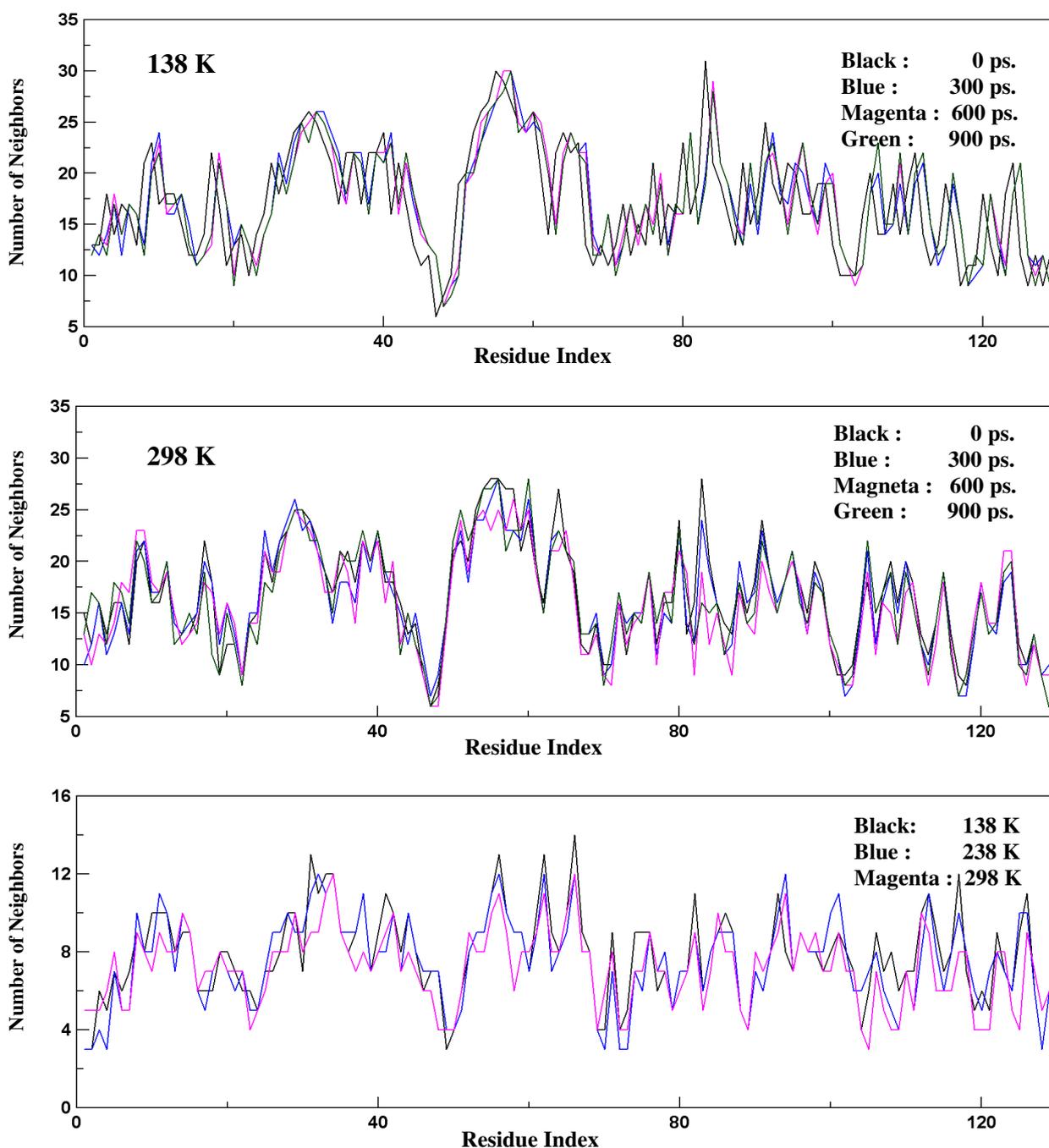


Figure 4.3.2 : Evolution of packing density in time and temperature domains for 193L

A network with a generic interaction term shall only allow new eigenmodes to emerge if and only if new connections are established. Therefore, residue-wise power spectrums, shown in Figure 4.4.3 as obtained from the displacements of C_α atoms are rich in information within the context of resolving time-scales. One immediate finding is that at lower temperatures all residues assume same Fourier components throughout their spectrum.

As the temperature increases juxtaposed power spectrums gain residue-wise asymmetry along with a more even distribution of intensities in the power spectra. This result is quite intuitive since it complies with the well established fact that the protein function is domain specific. At even higher temperatures some domains in the protein start to exhibit strong localization in the power density which favors the dominance of certain low frequency modes. Starting from the transition temperature the Fourier components become residue specific.

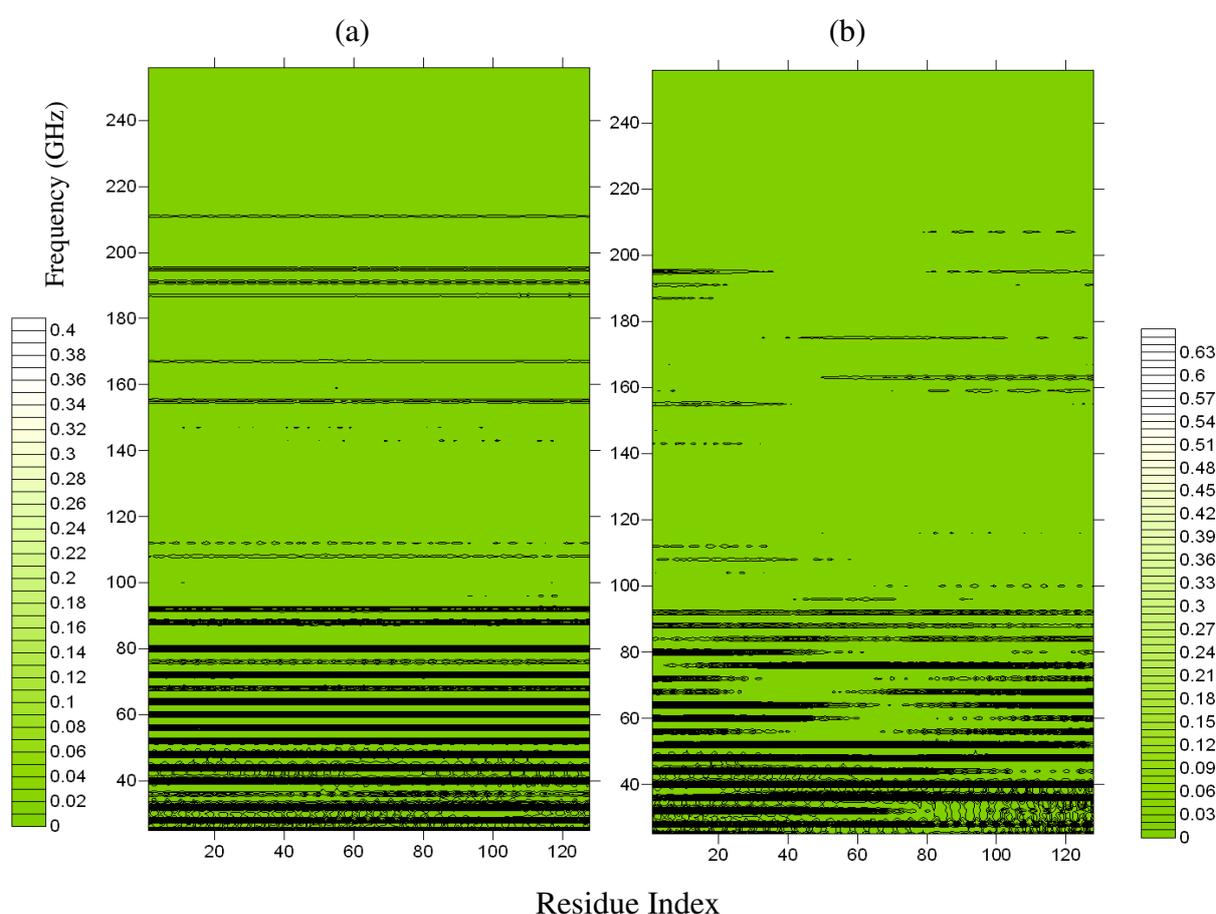


Figure 4.3.3 Power spectrum of the displacements of C_α atoms pertaining to Lysozyme at 158 K (a) and 298 K (b)

4.4 A Three-Mode Model for Stretch Exponential Decays

Our findings of the previous section point that the distribution of the characteristic frequencies are maintained throughout the temperature range studied, irrespective of the protein's size, type or function. The intensities at these frequencies, however, change with temperature. In fact we find that they also become residue specific above the protein transition temperature. We would thus like to test the idea that a strong change in the stretch exponent may be obtained (as implied by the stretch exponential fits of figure 4.1.2) with shifts in the intensities of the modes, while maintaining the frequencies of the modes.

Based on the assumption that each mode is governed by a simple exponential decay in the time window studied (2 ps – 2ns), we make a simple three-mode model to see the effect of the weight with which each mode contributes to the overall decay behavior of a correlation function. The generalized model is given by;

$$C(t) = \sum_{i=1}^3 A_i \exp\left[-\frac{t}{\tau_i}\right] = \exp\left[\left[-\frac{t}{\tau}\right]^\beta\right] \quad (4.4.1)$$

In this model A_i are treated as thermodynamical entities which do not change with temperature whereas the τ_i are treated as kinetic parameters, both the number and values of which are modified with temperature. We assume that the three modes have well separated time scales, e.g. each of them may be one order of magnitude apart. We also assume that the weights A_i of the modes change. We thus make three models whose parameters are listed in Table 4.4.1.

	Mode weights ($\tau_1 = 1$ ps, $\tau_2 = 10$ ps, $\tau_3 = 100$ ps)			KWW best-fit parameters	
	A_1	A_2	A_3	τ (ps)	β
Model 1	1/3	1/3	1/3	10.9	0.44
Model 2	0.4	0.3	0.3	14.1	0.40
Model 3	0.1	0.2	0.7	31.4	0.48

Table 4.4.1. Three models according to the first equality in equation 4.4.1, and the best-fitted KWW parameters to these models due to the second equality in equation 4.4.1.

In these models, pre-assigned τ_i s are taken to be 1,10 and 100 ps, whereas the weights, A_i , are assigned such that they add up to 1. In model 1, all the weights are taken to be equal. In models 2 and 3, larger weights are added to the high and low-frequency processes respectively. With only three parameters under control, it is not possible to sample all combinations of β and τ pairs, however it still provides a means to build a simple model which validates the claim that a range of stretch exponentials can be approximated by the superposition of single exponentials with varying weights. Stretch exponential decays as obtained from three different combinations of A_i are illustrated below in Figure 4.4.1;

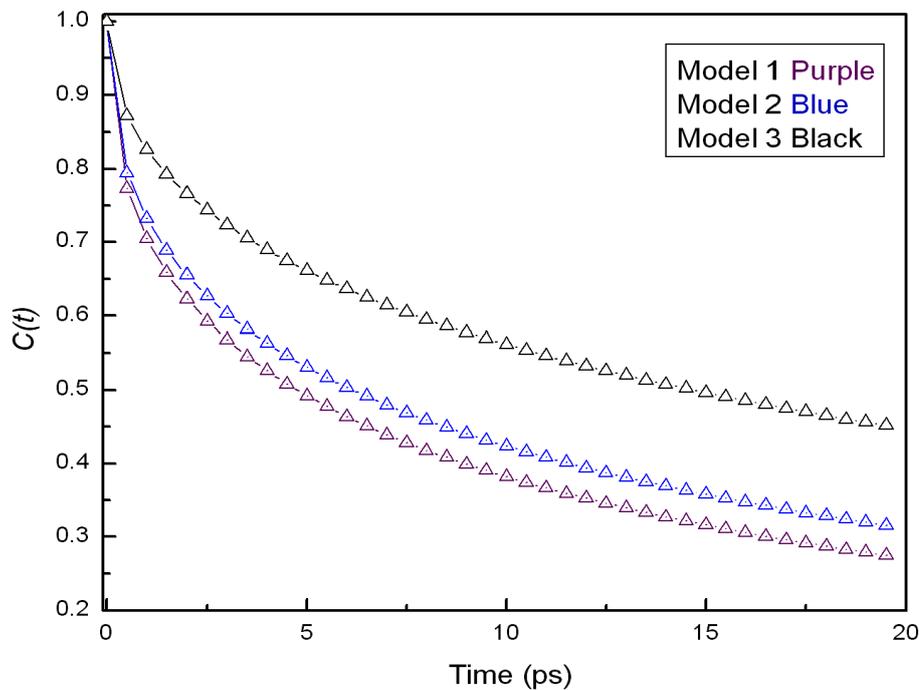


Figure 4.4.1 Stretch exponential functions as described in Table 4.4.1 are drawn together

Recently, a similar model was proposed (Baysal and Atilgan 2005) whereby, the transition in β exponents is ascribed to the insertion of an intermediate time scale, or mathematically, adding new terms to the decomposition given by Equation 4.4.1.

The previous model is not contradictory to the current one, in that the former is a specialized case where some of the mode weights approach zero. It is, however, important to realize that significant modifications in the stretch exponent are made possible (eg. from 0.4 to 0.48 in Table 4.4.1) by changes in the pre-exponential factors.

In the application of these ideas to the data originating from the MD simulations, the τ_i are to be assigned from the inverse of the frequencies, and the pre-exponential factors are to be best-fitted to give the original relaxation data. Note, however, that the number of modes in the actual data are much larger than three, rendering the best-fit rather complicated. The relationship between the computed intensity values and the pre-exponential factors is also of great scientific interest.

5. Conclusions and Future Prospects

We investigate the temperature dependent fluctuation characteristics of three proteins in their folded state using extensive MD simulations. The studied temperature range is ca. 130 – 330 K, where the lower limit corresponds to temperatures that are well below the so-called “protein glass transition” occurring at ca. 190 – 220 K, and the upper limit spans temperatures where the protein is still functional. The main variable that we monitor is $\Delta\mathbf{R}$ that measures the fluctuations of the C_α atoms of the proteins around an averaged structure, where the average structure is calculated after the overall translational modes and tumbling of the proteins are eliminated.

We find that, although the relative values of the residue-by-residue fluctuations are the same throughout the temperature range studied (see figure 1 of Baysal and Atilgan, 2005), the average size of the fluctuations show a linear increase in two different regimes, depending on temperature (figure 4.1.1 b). This finding is in agreement with the experimental findings on hydrated protein samples. Such a two-regime behavior is not expected of proteins that are not in water (e.g. compare the neutron scattering experiment results on partially hydrated and dehydrated samples in figure 4.1.1 a). The region where the change in slope between the two regimes occurs is attributed to the protein glass transition. It is also the temperature region above which the protein becomes functional.

All three proteins are shown to exhibit this universal character of structural relaxation which is also marked by a phase transition manifested in the change in heat capacities (figure 4.1.2) and the stretch exponential relaxation model (equation 1.1 and figure 4.1.3). Of particular interest is the cold shock protein 1G6P as an extreme case of functional proteins, which

exhibits the universal behaviour described above with an apparent temperature up-shift (Table 4.1.1). This protein is known to have a melting point ca. 20 K above many proteins that operate under mild conditions. Yet, it is interesting to see that this up-shift in the first order phase transition at the high temperatures is reflected to the higher order transition that determines the functionality of the protein, and calls for deeper investigation in future work dedicated to different cold shock proteins.

Meanwhile, the change in the flexibility of the protein is expected to find its roots in the way the landscape is sampled. Such a change may be due to two scenarios: (i) the landscape itself might be changing upon the transition, or (ii) the landscape remains the same, but there might be shifts in the way it is being sampled. Scenario (i) is unlikely, due to the fact that the residue-by-residue fluctuations show the same trend throughout the temperature window studied, suggesting that the landscape of the folded protein remains the same with increasing temperature. To completely rule out scenario (i), we further study the eigenvalue spectra of the C_{α} fluctuations (figures 4.3.1 and 4.3.3). The spectral analysis of the atomic coordinates reveals that, around their folded state, proteins fluctuate with a finite number of eigenmodes. Those modes and their location remain intact against changes in temperature. At lower temperatures, certain modes dominate the spectra, but all the residues exhibit very similar frequency distribution in their power spectra. The mode intensities, however, change with temperature, supporting scenario (ii). Moreover, at higher temperatures, the residue-wise symmetry of the power spectra is broken, such that some modes show different intensities for different residues. Note that all these observations are made while the three dimensional network structure is largely preserved in time and across a temperature range spanning the interval 140 K – 310 K (Figure 4.4.2).

We thus attribute the change in the flexibility of folded proteins to the onset of coupled

dynamics between solvent molecules and the protein surface. In fact, above the transition temperature, conformational jumps between the minima of the dihedral angles are observed for the side-chain torsional angles of surface residues, whereas their backbone angles do not visit different minima during the 2 ns trajectories. This change in the dynamics of the side chains is expected to be directly reflected to the relaxation behavior of the C_α atoms that they are attached to. The fact that the best-fitted stretch exponent of these relaxations is much less than 1 suggests a fractional Brownian dynamics character for the chain. In fact, inspection of the Hurst exponents of the dihedral angles belonging to the side-chain and the backbone of the solvent exposed surface residues implies a large deviation from classical Brownian motion, and at high temperatures the exponents are larger for the side chains than they are for the backbone. Note that, as the temperature is increased, dihedral fluctuations start to lose their mean reverting tendency (Table 4.2.1), although they are still very far from the random walk character that would be expected from a random coil.

Finally, we make a simple model to show that the changes in the relative weight of the modes can, by themselves, modify the stretch exponent considerably, while their frequencies remain the same. Our findings show that, starting with the same distribution of relaxation times, one can update β via changing the weights of the low frequency and higher frequency components (Table 4.4.1).

Future work shall involve carrying out the decomposition proposed in (1.1) via a robust way. Rather than carrying out the optimization for the best-fit of the relaxation functions $C(t)$ computed from the trajectories, for both τ_i 's and front factors A_i , one can use the power spectra data to estimate where the series shall be truncated and the values of τ_i 's in that range. This immensely reduces the dimensionality of the space of objective function parameters, thus allowing a nonlinear constrained optimization to be performed. The relationship between the

best-fitted pre-exponential factors and the intensities computed directly from the trajectories may, along with interesting physics, also reveal the modes that become more pronounced with temperature, letting us pinpoint the actual source of the transition.

Finally, the effect of solvent is definitely worthwhile to ponder on, especially considering the peculiar properties of the solvent used: water. Carrying out simulations with different solvents will test the validity of the universality of the protein glass transition. Owing to its molecular structure that allows hydrogen bonding and crystal-like phases, water may dampen the structural relaxations considerably. Therefore, it is much required to trace out the water dependency of protein dynamics so that the true relationship between stretch exponential decay and the protein function is revealed.

Appendix A: Relationship Between Heat Capacity at Constant Temperature and the Variance of Energy

(Adapted from *Molecular Modelling* by A. R. Leach (Leach, 1996))

Given the statistical mechanical relationship;

$$U = \langle E \rangle = \frac{k_B T^2}{Z} \frac{\partial Z}{\partial T} \quad (\text{A.1})$$

and using the well known expression for C_v , we obtain;

$$C_v = \left(\frac{\partial U}{\partial T} \right)_v = \frac{\partial}{\partial T} \left(\frac{k_B T^2}{Z} \frac{\partial Z}{\partial T} \right)_v = \frac{k_B T^2}{Z} \frac{\partial^2 Z}{\partial T^2} + \frac{2k_B T}{Z} \frac{\partial Z}{\partial T} - \frac{k_B T^2}{Z^2} \left(\frac{\partial Z}{\partial T} \right)^2 \quad (\text{A.2})$$

After differentiation, the first term on the left hand side becomes the only one that require more effort to write explicitly in $\langle E \rangle$, so we make use of another identity ;

$$\frac{\partial}{\partial T} \left(\frac{\langle E \rangle}{k_B T^2} \right) = \frac{\partial}{\partial T} \left\{ \frac{1}{Z} \left(\frac{\partial Z}{\partial T} \right) \right\} \longrightarrow - \left(\frac{\langle E \rangle}{k_B T^3} \right) = \frac{1}{Z} \frac{\partial^2 Z}{\partial T^2} + \frac{\partial Z}{\partial T} \frac{\partial}{\partial T} \left(\frac{1}{Z} \right) \quad (\text{A.3})$$

$$\frac{\partial Z}{\partial T} \frac{\partial}{\partial T} \left(\frac{1}{Z} \right) = \frac{\partial Z}{\partial T} \frac{\partial Z}{\partial T} \frac{\partial}{\partial T} \left(\frac{1}{Z} \right) = - \left(\frac{\partial Z}{\partial T} \right)^2 \left(\frac{1}{Z} \right)^2 \longrightarrow \frac{k_B T^2}{Z} \frac{\partial^2 Z}{\partial T^2} = \frac{-2 \langle E \rangle}{k_B T^3} + \frac{\langle E^2 \rangle}{k_B^2 T^4} \quad (\text{A.4})$$

By collecting all terms together we relate the energetic fluctuations to heat capacity as;

$$C_v = k_B T^2 \left[\frac{-2 \langle E \rangle}{k_B T^3} + \frac{\langle E^2 \rangle}{k_B^2 T^4} \right] + 2 \frac{\langle E \rangle}{T} - \frac{\langle E \rangle^2}{k_B T^2} = \frac{(\langle E^2 \rangle - \langle E \rangle^2)}{k_B T^2} \quad (\text{A.5})$$

Appendix B : Fourier and Eigenvalue Spectrums of 5PTI, 193L and 1G6P over Different Temperatures of Interest

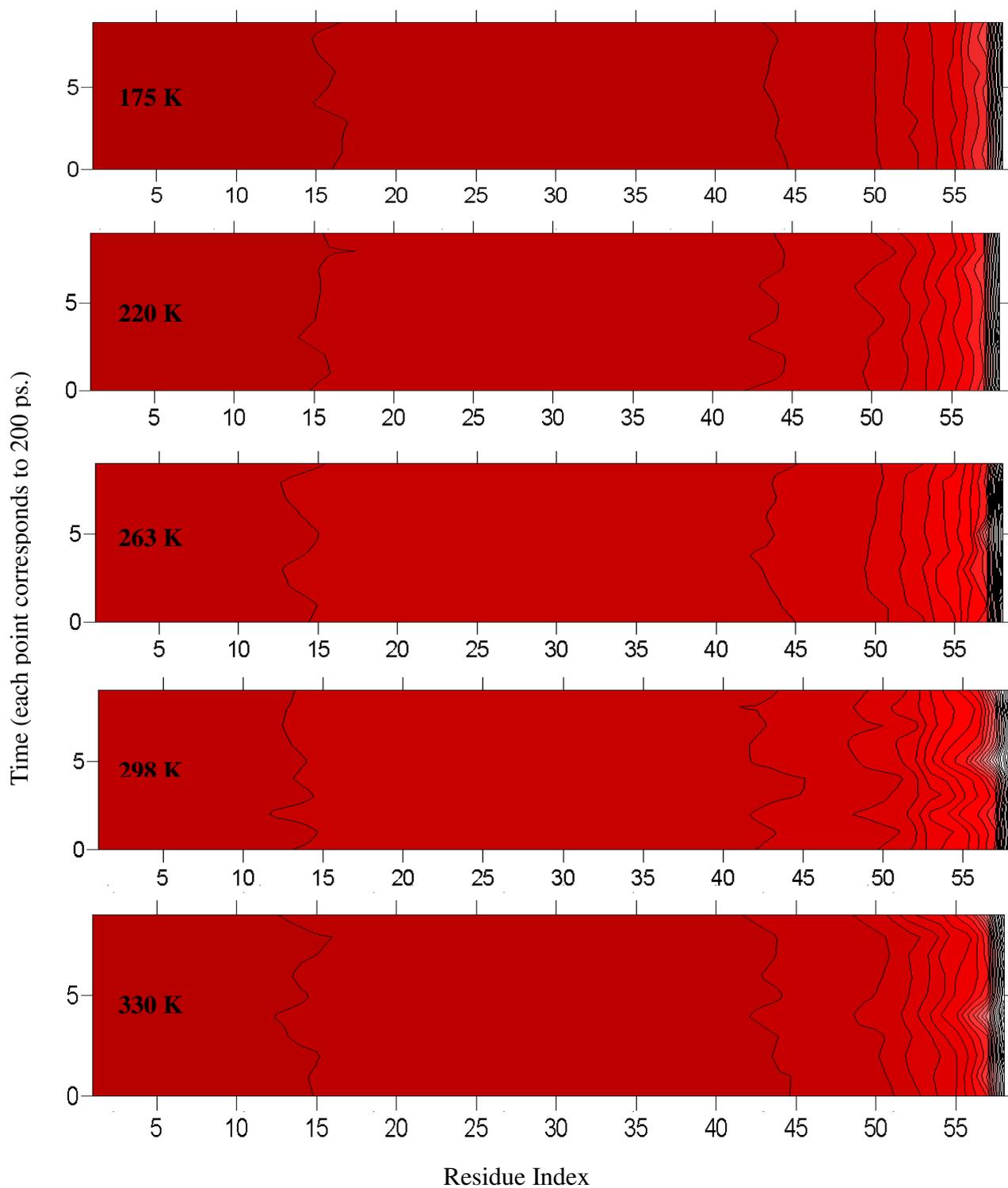


Figure B.1.1 Eigenvalue Spectrum for 5PTI at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger with 0.2 increment in contour levels).

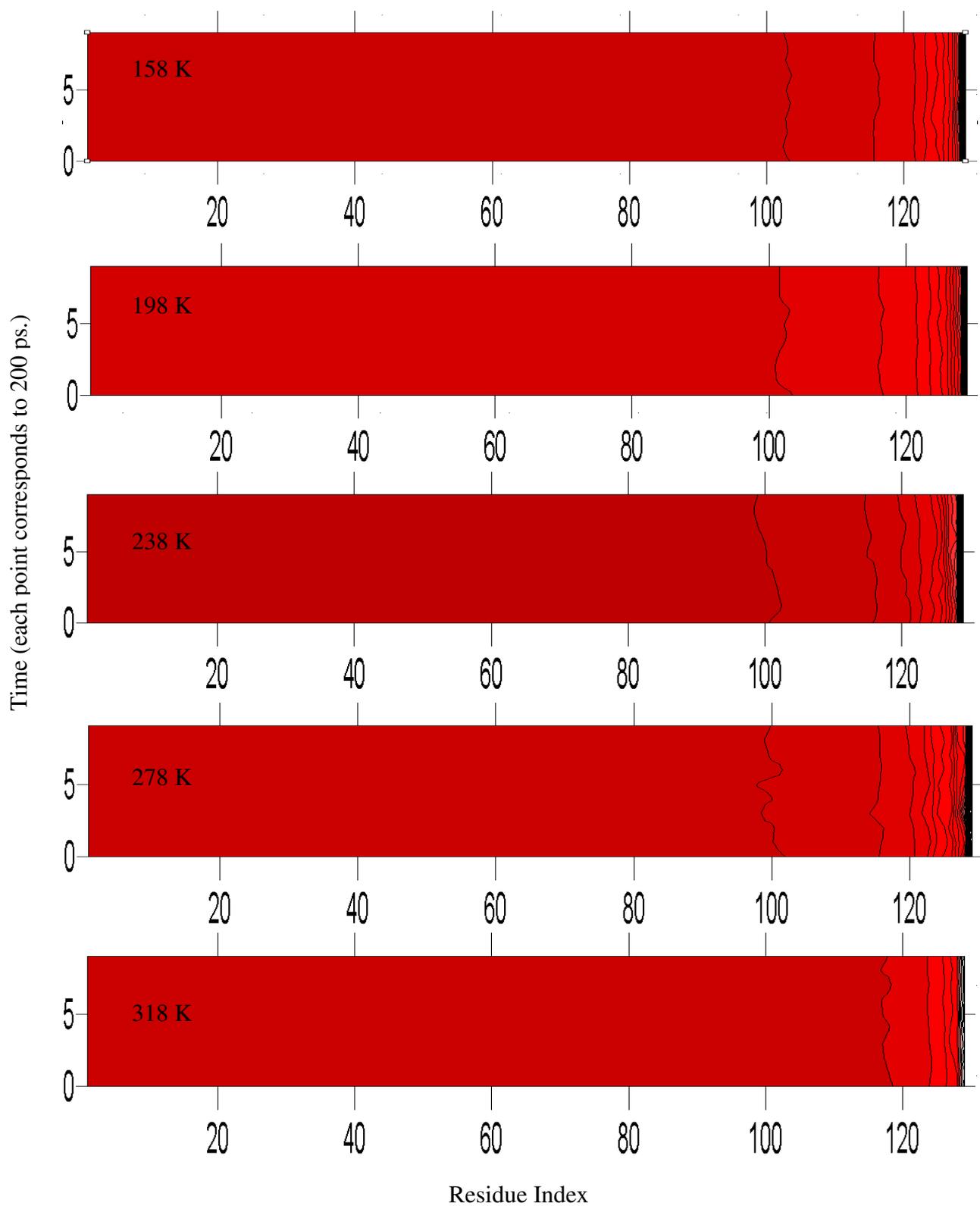


Figure B.2 Eigenvalue Spectrum for 193L at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.2 increment in contour levels.

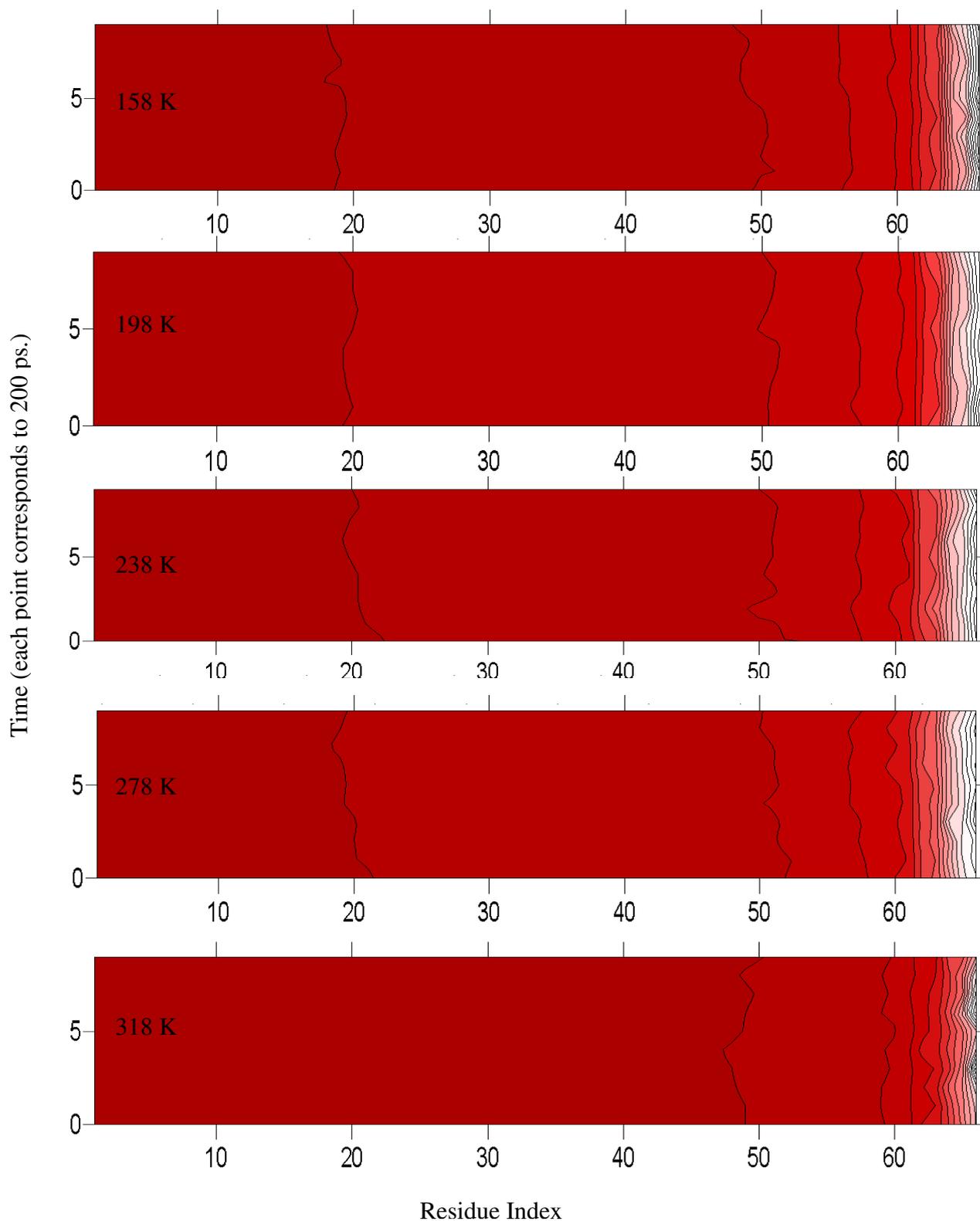


Figure B.3 Eigenvalue Spectrum for 1G6P at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.2 increment in contour levels.

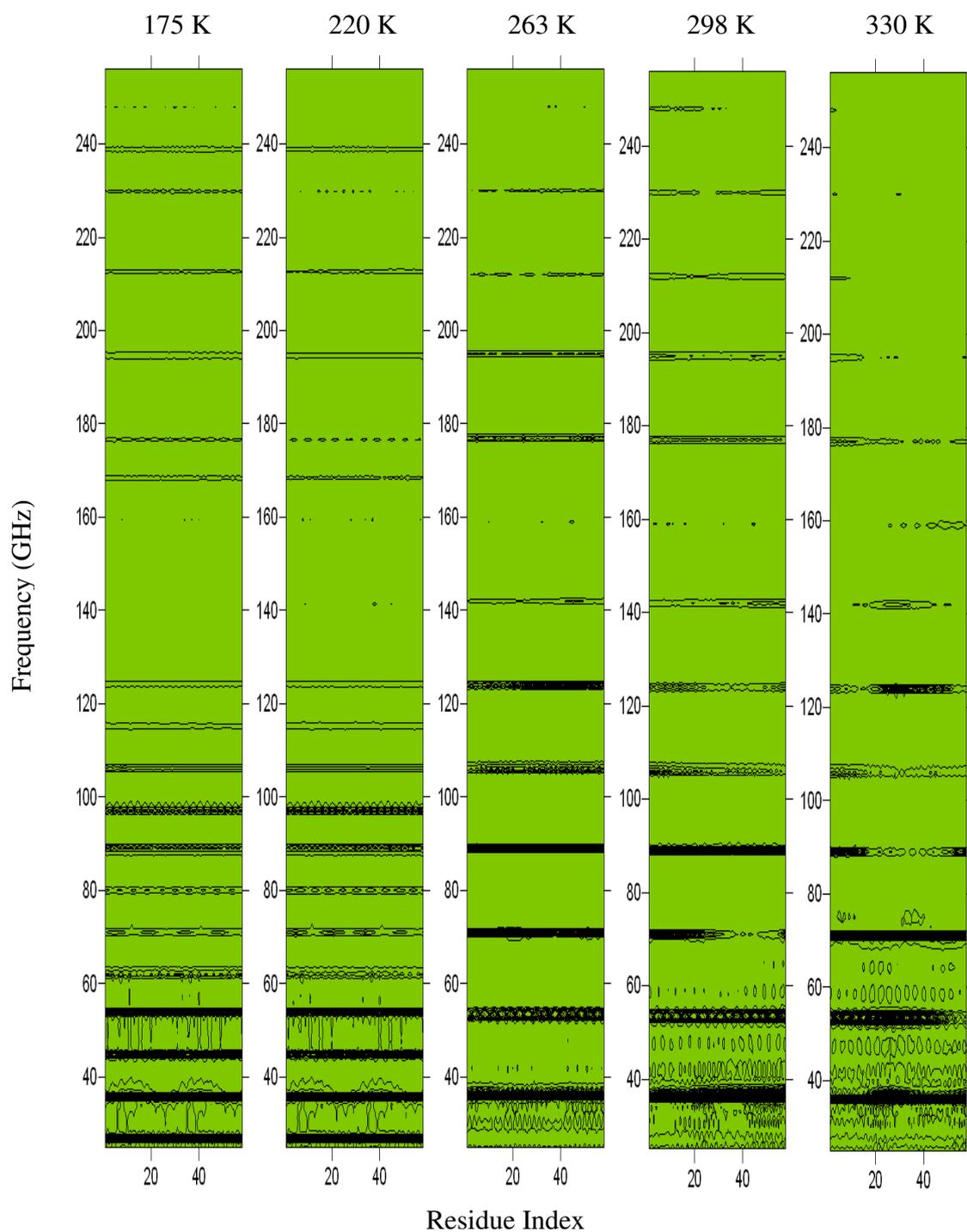


Figure B.4 Eigenvalue Spectrum for 5PTI at five different temperatures. Color coded for real eigenvalues (Green (Smaller) to White (Bigger)) with 0.03 increment in contour levels.

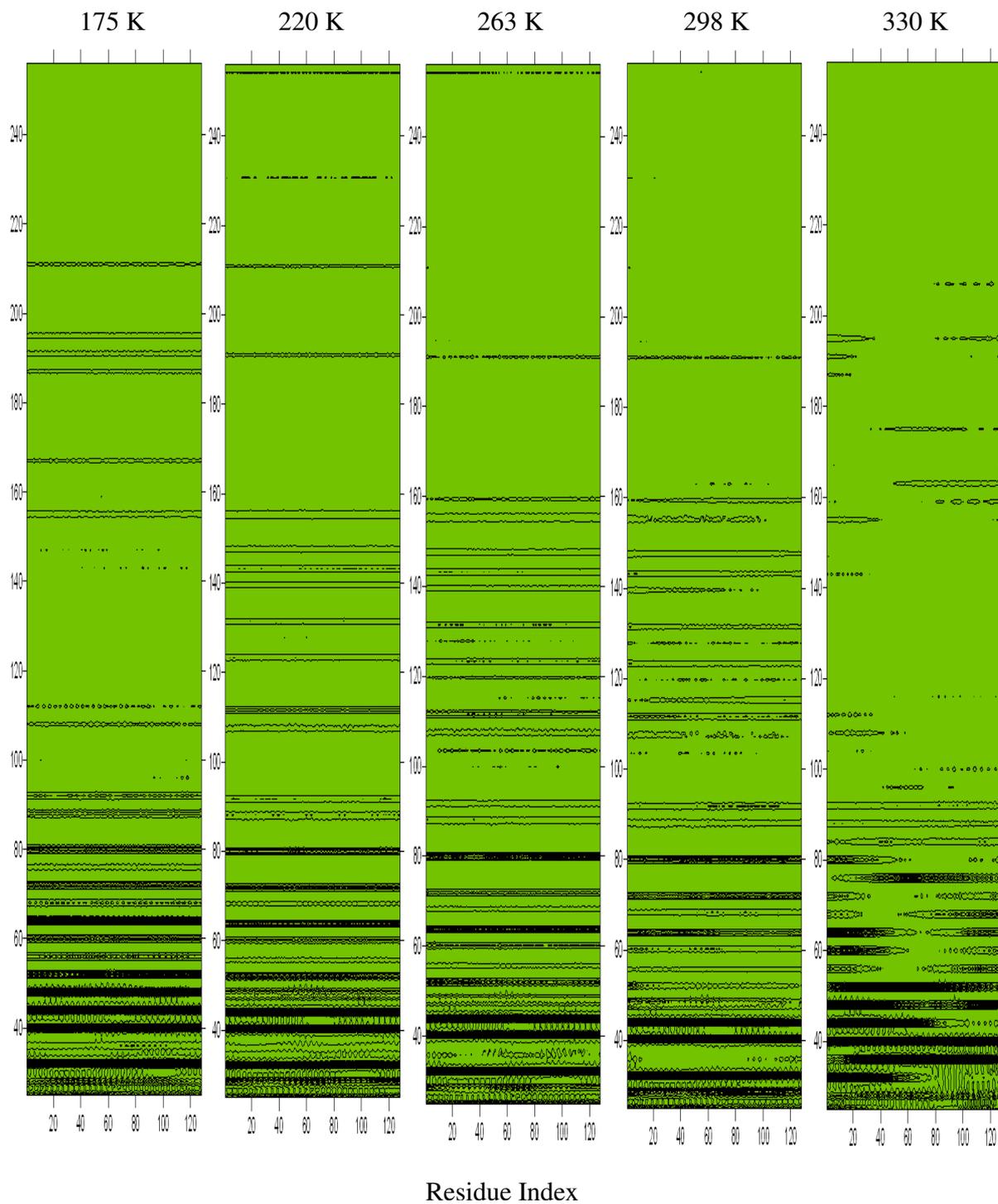


Figure B.5 Eigenvalue Spectrum for 193L at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.03 increment in contour levels.

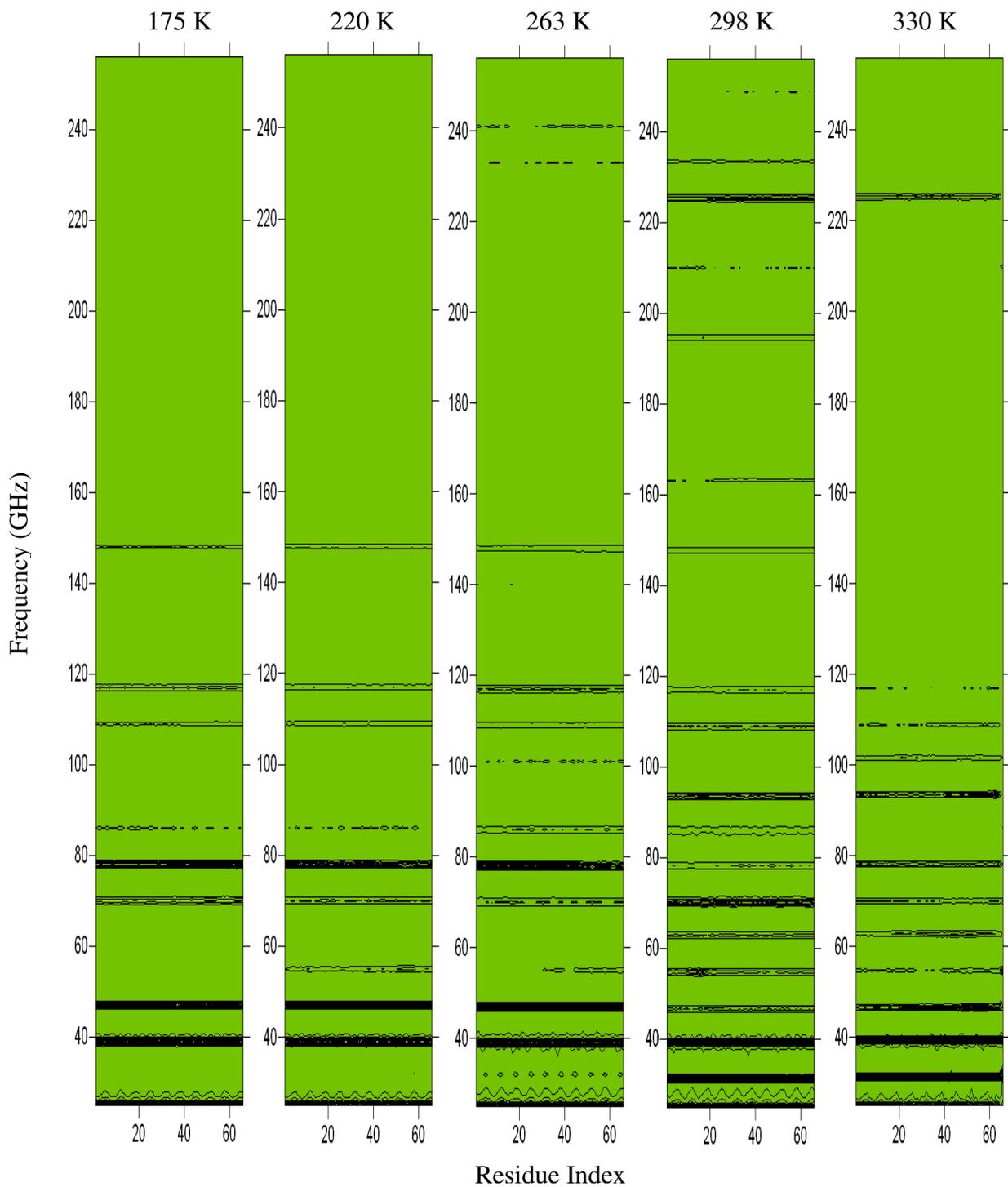


Figure B.6 Eigenvalue Spectrum for 1G6P at five different temperatures. Color coded for real eigenvalues (Red (Smaller) to White (Bigger)) with 0.03 increment in contour levels.

References

Abry, P., F. Sellan, 1996, "The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation." *Appl. And Comp. Harmonic Anal.*, 3(4):377-383

Accelrys Inc, San Diego, CA, U.S.A. (Insight II 4.0.0 P+)

Allen, M. P., D. J. Tildesley, 1991, *Computer Simulation of Liquids*, Clarendon Press, Oxford

Andersen H.C.. 1980. "Molecular Dynamics Simulations at Constant Pressure and/or Temperature." *J. of Chem. Phys.* 72(4): 2384-2393

Andersen H.C.. 1983. "Rattle: A 'Velocity' Version of the Shake Algorithm for Molecular Dynamics Calculations." *J. of Comp. Phys.*, 54:24-34

Atilgan, A. R., P. Akan, C. Baysal. 2004. "Small-world communication of residues and significance for protein dynamics." *Biophys. J.* 86:85-91

Bahar I., A. R. Atilgan, B. Erman. 1997. "Direct evaluation of thermal fluctuations in protein using a single parameter potential." *Folding and Design* 2. (3):173-181

Baysal, C., A. R. Atilgan. 2002. "Relaxation kinetics and the glassiness of proteins: the case of bovine pancreatic trypsin inhibitor." *Biophys. J.* 83:699–705

Baysal, C., A. R. Atilgan. 2005. "Relaxation Kinetics and the Glassiness of Native Proteins: Coupling of Timescales." *Biophys. J.* 88:1570–1576

Berendsen, H. J. C., J. R. Grigera, T. P. Straatsma. 1987. "The missing term in effective pair potentials", *J. Phys. Chem.* 91:6269-6271

Berman H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissing, I. N. Shindyalov, P. E. Bourne. 2000. "The Protein Data Bank." *Nuc. Acid. Res.* 28:235-242

Dastidar, S. G., C. Mukhopadhyay. 2003. "Structure, dynamics, and energetics of water at the surface of a small globular protein: a molecular dynamics simulation". *Phys. Rev. E.* 68 (021921):8044–8052

Daniel, R. M., J. C. Smith, M. Ferrand, S. Hery, R. Dunn, and J. L. Finney. 1998. "Enzyme activity below the dynamical transition at 220 K." *Biophys. J.* 75:2504–2507

Dauber-Osguthorpe, P., V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler. 1988. "Structure and energetics of ligand binding to proteins: E. coli dihydrofolate reductase-trimethoprim, a drugreceptor system." *Proteins.* 4:31–47

Doi, M., and S. F. Edwards. 1986. *The Theory of Polymer Dynamics*. Oxford Science, Oxford

Duan Y., P. A. Kollman. 1998. "Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution." *Science* 282:740-743

Dvorsky, R., J. Sevcik, L. S. D. Caves, R. E. Hubbard, and C. S. Verma. 2000. "Temperature effects on protein motions: a molecular dynamics study of RNase-Sa." *J. Phys. Chem.* 104:10387–10397

Eberly D. H., 2004. *Game Physics*. Morgan Kaufmann

Fenimore, P. W., H. Frauenfelder, B. H. McMahon, and F. G. Parak. 2002. "Slaving: solvent fluctuations dominate protein dynamics and functions." *Proc. Natl. Acad. Sci.* 99:16047–16051

Frauenfelder, H., F. Parak, and R. Young. 1988. "Conformational substates in proteins." *Annu. Rev. Biophys. Chem.* 17:451-479

Frenkel, D., Smit B., 2001. *Understanding Molecular Simulation: from Algorithms to Applications*. Academic Press

Gabel, G., M. Weik, B. P. Doctor, A. Saxena, D. Fournier, L. Brochier, F. Renault, P. Masson, I. Silman, and G. Zaccai. 2004. "The Influence of Solvent Composition on Global Dynamics of Human Butyrylcholinesterase Powders: A Neutron-Scattering Study." *Biophys. J.* 86: 3152–3165

Garcia, A. E., R. Blumenfeld, G. Hummer, and J. A. Krumhansl. 1997. "Multi-basin dynamics of a protein in a crystal environment." *Physica D*. 107:225–239

Haliloglu T., I. Bahar, B. Erman. 1997. "Gaussian dynamics of folded proteins." *Phys. Rev. Lett.* 79:3090-3093

Heisterberg, D. J. 1990. "QUATFIT algorithm for superimposing two similar rigid molecules." The Ohio Supercomputer Center, Ohio State University, Columbus, OH.

Kohlrausch, R. 1874. "Theorie des elektrischen Ruckstandes in der Leidener Flasche". *Ann. Phys. Chem. (Leipzig)* 91:179–214

Kremer, W., B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke, H. R. Kalbitzer. 2001, "Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritime*." *Eur. J. Biochem.* 268:2527-2539

Kubo, R. 1986. "Brownian motion and non-equilibrium statistical mechanics." *Science* 233:330–334

Kubo, R., M. Toda, N. Hashitsume. 1978. *Statistical Physics II: Nonequilibrium Statistical Mechanics*. Springer-Verlag, New York

Leach, A. R.. 1996. *Molecular Modelling, Principles and Applications*. Longman. London

Lindsey C. and G. Patterson. 1980. "Detailed comparison of the Williams-Watts and Cole-Davidson functions." *J. Chem. Phys.* 73:3348 -1980

Nose, S. 1984. "A molecular dynamics method for simulations in the canonical ensemble." *Mol. Phys.* 50:255–268

Pal, S. K., J. Peon, and A. H. Zewail. 2002. "Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution". *Proc. Natl. Acad. Sci.* 99:1763–1768

Press W. H., S. A. Teukolsky, W. T. Vetterling, B. Falnnery. 1992. *Numerical Recipes in Fortran 77*. Cambridge University Press. Cambridge

Steeb W., 2002. *The Nonlinear Handbook*. World-Scientific. Singapore

Swope W. C., H. C. Anderson, P. H. Berens, K. R. Wilson, 1982, "A computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters," *J. Chem. Phys.* 93:9042-9048

Tarek, M., and D. J. Tobias. 2002. "Role of protein-water hydrogen bond dynamics in the protein dynamical transition." *Phys. Rev. Lett.* 88:138101–138104

Tournier, A. L., J. Xu, and J. C. Smith. 2003. Translational hydration water dynamics drives the protein glass transition.” *Biophys. J.* 85:1871–1875

Tsai, A. M., T. J. Udovic, and D. A. Neumann. 2001. “The Inverse relationship between protein dynamics and thermal stability.” *Biophys. J.* 81:2339–2343

Vaney, M. C., S. Maignan, M. RiesKautt, A. Ducruix. 1996. “High resolution structure (1.33 angstrom of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission.” *Acta Crystallogr. D. Biol. Crystallogr.* 52:505-517

Vitkup, D., D. Ringe, G. A. Petsko, and M. Karplus. 2000. “Solvent mobility and the protein ‘glass’ transition.” *Nat. Struct. Biol.* 7:34–38

Williams, G., and D. C. Watts. 1970. “Non-symmetrical dielectric relaxation behavior arising from a simple empirical decay function.” *Trans. Faraday Soc.* 66:80–85

Wlodawer, A., J. Walter, L. Huber, and L. Sjolin. 1984. “Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and x-ray refinement of crystal form II.” *J. Mol. Biol.* 180:301–329

Yilmaz, L. S., A. R. Atilgan. 2000. “Identifying the adaptive mechanism in globular proteins: fluctuations in densely packed regions manipulate flexible parts.” *J. Chem. Phys.* 113:4454-4464

Zaccai, G. 2000. "How soft is a protein? A protein dynamics force constant measured by neutron scattering." *Science* 288:1604–1607