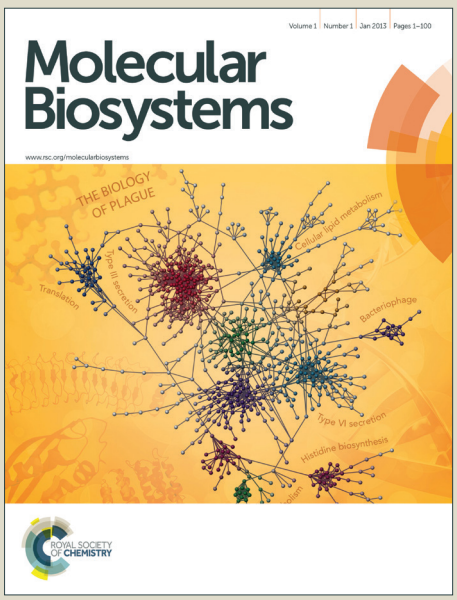
Molecular BioSystems

Accepted Manuscript

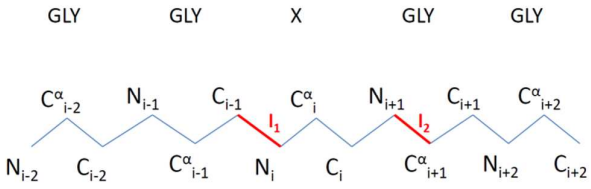


ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Table of Contents

We present a computational method to investigate conformational transitions of the twenty amino acids based on molecular dynamics (MD) simulations and the dynamic rotational isomeric state (DRIS) model. Local dynamics of twenty amino acids resulting from rotational transitions between isomeric states are analyzed.



Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Conformational Transitions in the Ramachandran Space of Amino Acids by the Dynamic Rotational Isomeric State (DRIS) Model

Cigdem Sevim Bayrak* and Burak Erman

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

The dynamic rotational isomeric state model is applied to predict the internal dynamics of the 20 amino acids. Transition rates between rotational isomeric states are calculated from molecular dynamics simulations of Gly-Gly-X-Gly-Gly peptides where X represents one of the 20 amino acids. Predicted relaxation times are in good agreement with fluorescence quenching rate measurements.

Introduction

Proteins are not rigid structures, they are dynamic molecules characterized by conformational ensembles at equilibrium. Proteins perform most of their function through rearrangements of their conformations, such as allosteric rearrangements, large scale fluctuations, tail motions etc. Therefore, the knowledge of only the average structure of a protein does not provide a complete description. Dynamic properties are essential for understanding the basic features of their functions. The knowledge of the popularity of the states, transition rates and the pathways are all necessary for a full understanding of a protein [1]. Proteins exhibit both internal motions including bond vibrations, bond stretching, side-chain motions, rotational transitions and global motions as rotational and translational motions [2, 3]. In this study, we consider dynamics at the single residue level and focus on the internal motions resulting from conformational transitions exhibited by the twenty amino acids. We adopt the dynamic rotational isomeric states (DRIS) formalism in explaining and interpreting the residue dynamics. The DRIS approach has been developed for predicting local internal dynamics of polymer chains by Bahar and Erman [4]. The model requires the knowledge of the exact locations and probabilities of all torsion states of each amino acid and the transition rates between torsion states. These parameters are determined in this work by an exhaustive sampling of amino acid Ramachandran plots by Molecular dynamics (MD) simulations. In a sense, the DRIS formalism is used to organize the MD simulations results in a compact and efficient way as explained in detail in the rest of this paper.

The DRIS model is an extension of the equilibrium theory of chain statistics by Flory [5] to the dynamic domain by Jernigan [6] and is conveniently used to calculate the internal time correlation functions of a chain and different dynamic properties associated with the transitions between the isomeric states [2, 4, 7-12]. Internal dynamics of a polyethylene chain was characterized by calculating the autocorrelation of a vector affixed to the middle of a sequence in the chain [4]. Orientational correlation functions, correlation times, and spectral densities [7] of poly(ethylene oxide) segments were determined. The method was also used to determine the activation energies of the local conformational transitions for a polyethylene chain [8]. DRIS is used to obtain the conformational autocorrelation functions, and first and second orientational autocorrelation functions of a polyethylene chain and it has been shown that the results agree satisfactorily with NMR experiments [9]. A mathematical scheme to examine the stochastic process of conformational transitions between rotational isomeric states in polymer chains was developed by Bahar [10]. According to the model, stochastic weights are assigned to the configurational transitions and stochastic weight matrices are defined. A matrix multiplication scheme is adopted to analyze the isomeric transitions by serial multiplication of stochastic weight matrices. Then, this approach is extended to efficiently calculate the first and second orientational autocorrelation functions [11]. The mechanism of local conformational transitions in poly(dialkylsiloxanes) which is a typical example that has a highly flexible backbone with bulky and highly articulated side groups is analyzed [12]. The results of DRIS method has been compared with molecular dynamics simulations and it has been shown that DRIS provides an efficient way of analyzing the mechanism of the local relaxation of a polymer chain rather than long simulations [2, 13].

A computational approach is given for measuring relaxation times of the amino acids. Here, the aim is to investigate the conformations of the twenty amino acids whose dynamics evolves only from transitions from one torsion minimum to another as it would obtain in the random coil state. The preferences of the amino acids extracted from the Protein Data Bank are biased by the presence of secondary and tertiary structure, and long range interactions such as contacts and hydrogen bonds of a given residue with the rest of the protein [14]. Hence, instead of analyzing the protein structures from the protein data bank (PDB) or running simulations of proteins we perform MD simulations of Gly-Gly-X-Gly-Gly pentapeptides. Here X represents the amino acid to be examined. The GGXGG peptides have been commonly used as models for the random coil state [14-19].

We use MD simulations to identify isomeric states of the amino

acids and to calculate the equilibrium probabilities of the corresponding states. Then, the rates of the transitions between the isomeric states are calculated using the DRIS approach based on the parameters obtained via MD simulations. The relaxation rates determined via DRIS model are compared with the fluorescence quenching rate constants obtained experimentally by Huang and Nau [20]. Results of our calculations agree with their experimental findings.

Methods and Materials

Here we apply the DRIS model to analyze the dynamics of the transitions of (φ,ψ) angles of the amino acids. These transitions are determined by the transitions between the specified regions of the Ramachandran map. Those regions are separated by energy barriers. We have defined $\nu = 8$ discrete regions for this purpose based on the molecular dynamics simulations we have performed. Let the torsion state of an amino acid be defined by

$$\{\Phi\}_i = \{\varphi_i, \psi_i\} \tag{1}$$

Based on the interdependence between the neighbor amino acids, the transitions may be defined as (i) first order (independent), According to this, the statistics of $\{\Phi\}_i$ is independent of the statistics of its neighbors, (ii) second order (pairwise dependent), where the statistics of $\{\Phi\}_i$ depends on the statistics of its preceding neighbor, and (iii) third order (triplewise dependent) where the statistics of $\{\Phi\}_i$ depends on the statistics of its preceding two neighbors. In this work, we adopt the Flory independence hypothesis and assume first order statistics [5]. For a first order transition of an amino acid, we define the master equation as

$$\frac{dP(t)}{dt} = AP(t) \tag{2}$$

where the vector $P(t)$ is the 8-dimensional vector of the probabilities of 8 states, and the transition rate matrix A is the 8×8 dimensional matrix. The state of a given amino acid is defined by one of the torsion states $\{\Phi\}_k$, $k = 1,\ldots,8$. Then, A_{ij}, the ijth element of the rate matrix, denotes the rate of the transition from state $\{\Phi\}_j$ to the state $\{\Phi\}_i$.

The formal solution of the master equation gives

$$P(t) = \exp\{At\}P(t=0) \tag{3}$$

$P(t=0)$ represents the vector of the equilibrium probabilities of the 8 states. The eigendecomposition of the matrix A gives $A = B\Lambda B^{-1}$. Here B is the matrix whose ith column is the ith eigenvector of A, Λ is the diagonal matrix whose diagonal elements are the eigenvalues λ_i, and B^{-1} is the inverse of B. Then,

$$P(t) = B\exp\{\Lambda t\}B^{-1}P(t=0) \tag{4}$$

The term $Be^{\Lambda t}B^{-1}$ defines the time-delayed conditional probability matrix $C(t)$. The element C_{ij} denotes the probability of being at state $\{\Phi\}_i$ at time t, given the initial state $\{\Phi\}_j$ at t=0. The elements of each column of C_{ij} comprise all possible transitions from a given initial state to one of the 8 possible states. Hence, the summation of each column is unity.

$$C(t) = B\exp\{\Lambda t\}B^{-1} \tag{5}$$

The total time-dependent joint probability matrix $P(t)$ is defined as

$$P(t) = C(t)\ \mathrm{diag}P(t=0) \tag{6}$$

The element P_{ij} represents the joint probability of occurrence of state $\{\Phi\}_i$ at time t and $\{\Phi\}_j$ at t=0. P_{ij} may be defined as

$$P_{ij} = \sum_k B_{ik}\exp\{\lambda_k t\}B_{kj}^{-1}P_j(0) \tag{7}$$

Knowledge of the joint probability matrix $P(t)$, or alternatively knowledge of conditional probability matrix $C(t)$ with the equilibrium probabilities $P(t=0)$ gives a complete description of the dynamics of a given amino acid.

For any dynamic property f_{ij} which is a function of the conformations $\{\Phi\}_i$ and $\{\Phi\}_j$, the average of the property $\langle f_{ij}\rangle$ over all possible transitions taking place from conformation j to i can be calculated by assigning a stochastic weight to each transition

$$\langle f_{ij}\rangle = \sum_i\sum_j P_{ij}(t)\, f_{ij} \tag{8}$$

Molecular Dynamics Simulations for Determination of the States and Equilibrium Probabilities

A capped Gly-Gly-X-Gly-Gly peptide model is used for the analysis of the amino acid X. Simulations are performed using Gromacs 4.5.5 package [21]. The –N and the –C termini of the pentapeptide models Gly-Gly-X-Gly-Gly are capped with ACE and NME respectively. The initial conformation is selected as extended.

OPLS/AA all atom force field is used and the system is solvated using four-atom water model TIP4P [22, 23]. Electrostatic interactions are calculated using PME-Switch method with Coulomb cut-off 1.0 nm. Van der Waals interactions are calculated using Switch method with VDW cut-off 1.0 nm.

An energy minimization using the conjugate gradient method with 5000 maximum number of steps and 0.001 ps time step is performed. To assure that the solvent configuration matches the peptide, the solvent and the hydrogen atoms are relaxed for 2500 steps while the peptide was restrained.

Before the production MD step, first a 20 ps temperature coupling MD simulation and then a 40 ps pressure coupling MD simulation is performed. During the 20 ps NVT ensemble simulation the Berendsen temperature coupling method is used [24]. The coupling constant is assigned as 0.1 ps. Then, during a 40 ps NPT ensemble simulation Berendsen temperature and pressure coupling methods are used. The temperature coupling constant is again assigned as 0.1 ps and the pressure coupling constant is taken as 0.5 ps.

A simulated annealing procedure is necessary to avoid a local minimum and to sample the energy surface properly [25]. We used 80 ps simulated annealing cycles during the production MD simulation. In each cycle, the temperature of the system is increased from 310 K to 1010 K in the first 20 ps. Then in the second 20 ps part the temperature of the system is decreased from

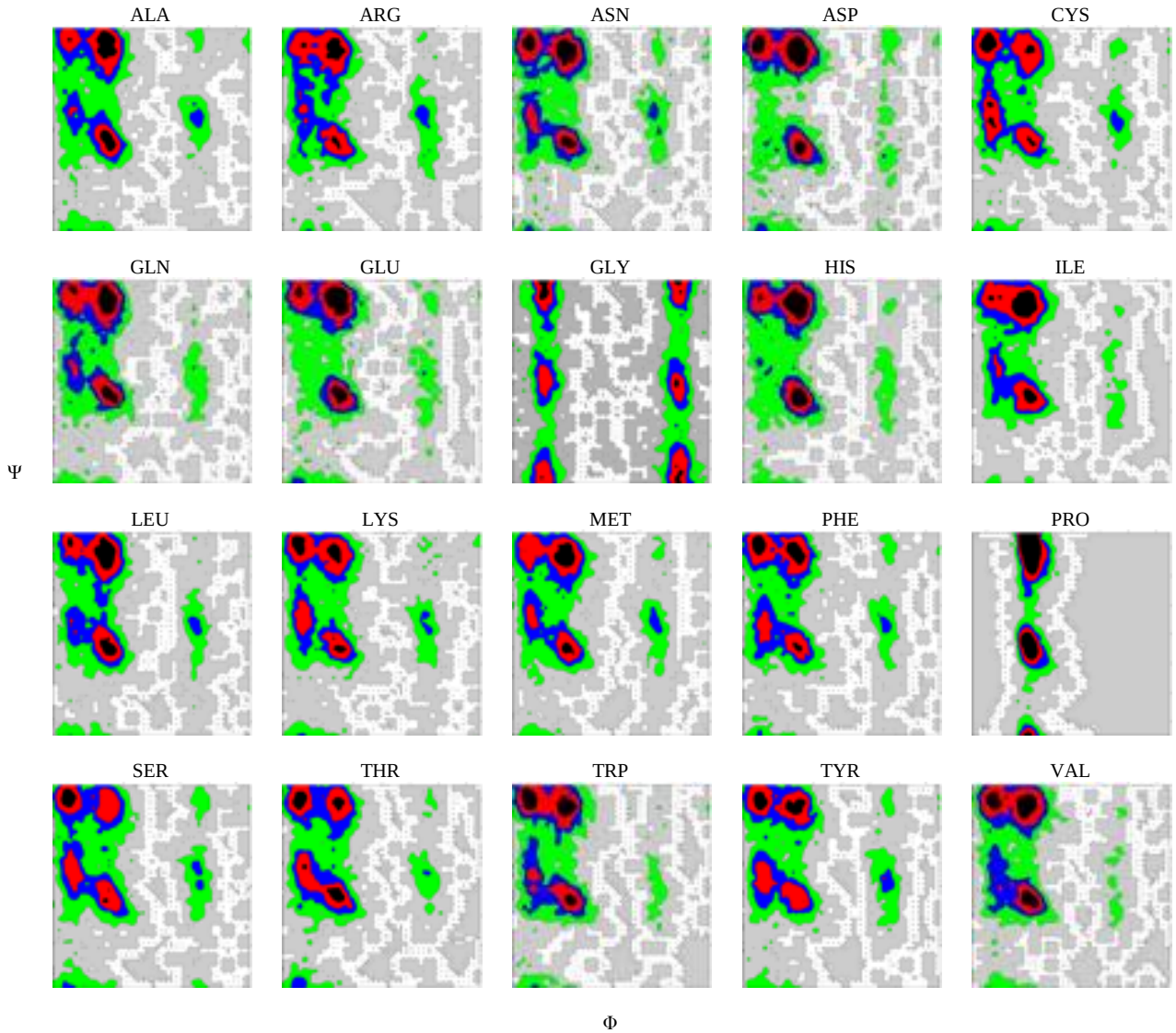


Fig. 1. Ramachandran plots for the twenty amino acids. The populations are obtained from MD simulations of GGXGG peptides.

1010 K to 310 K. Finally in the last 40 ps part the temperature of the system is kept constant at 310 K. This equilibration part is necessary since the collected data should be sampled from a Boltzmann distribution. Using the simulated annealing method, a different initial conformation is selected in each cycle. Since the system is equilibrated after heating and cooling procedures, we assure that the selected conformations are sampled from Boltzmann distribution.

Determination of the Rotational Isomeric States

The Ramachandran plots showing the distributions of the populated regions of 20 amino acids in the Gly-Gly-X-Gly-Gly peptide are constructed.. The Ramachandran plot of each amino acid is obtained directly from the simulation trajectories. The common populated regions are selected as the torsion states.

In **Fig. 1**, a contour plot for each amino acid plotted using $5^{\circ} \times 5^{\circ}$ grids where the x- y-axes represent φ and ψ angle distributions, respectively. The percentage of points in each grid are colored as: $[0,0.05)$: gray ; $[0.05,0.2)$: green ; $[0.2,0.4)$: blue ; $[0.4,0.8)$: red ; $[0.8,1]$: black that has been used by Beck et al. [14]. Based on the Ramachandran plots, 8 torsion states are determined. The selected regions are shown in **Fig. 2**. We have defined 7 states based on the popularity of the regions and identified 8^{th} state as the all other regions. Since Proline has a constrained phi angle, we have determined 4 torsion states for Proline. The identified states of Proline are shown in **Fig. 3**. Again, the 4^{th} state is defined as the all other allowed regions.

The regions are not so different from the regions introduced by Karplus [26]. Unal et al. also defined eleven states based on a knowledge based database which are similar to the regions we have defined by MD simulations [27-29]. In a knowledge based database, the preferences of the amino acids are biased by the presence of the protein. Furthermore, a coil library that is constructed by removing helical or beta structures may still include influences of conformations of proteins [14]. Therefore we perform MD simulations of pentapeptides which would present

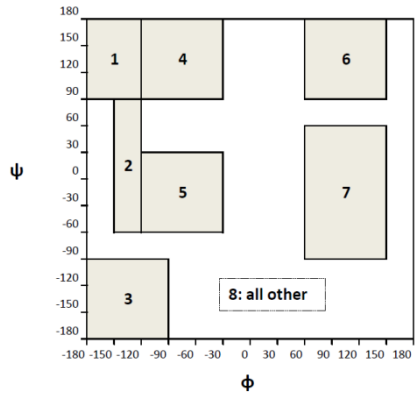


Fig. 2. The representation of 8 states on Ramachandran map.

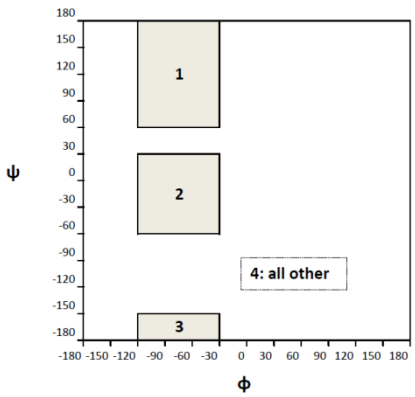


Fig. 3. The representation of 4 states of Proline.

better representatives for the random coil structures. The boundaries of the states are selected in such a way that the transitions between these states would be analyzed explicitly. We have defined rectangular shaped states centered on the most populated regions.

Calculating the Equilibrium Probabilities

The equilibrium probabilities are calculated from the equilibrium conformations data collected via MD simulations. For each amino acid, we simply count the number of states visited. The probability of observing an amino acid X in a state i is calculated as

$$P^X(s_i) = \frac{N^X(s_i)}{\sum_i N^X(s_i)} \qquad (9)$$

where $N^X(s_i)$ represents the number of occurrences of amino acid X in a state i during the simulation and the summation term indicates the total number of occurrences in all possible states by amino acid X.

The equilibrium probabilities of each amino acid obtained by MD simulations are presented in Table I.

Calculation of the Rates

In the previous applications of the DRIS, the rate constants related to transitions were calculated by using a kinetic scheme [4, 7]. The rates were determined by the activation energies E_a and a

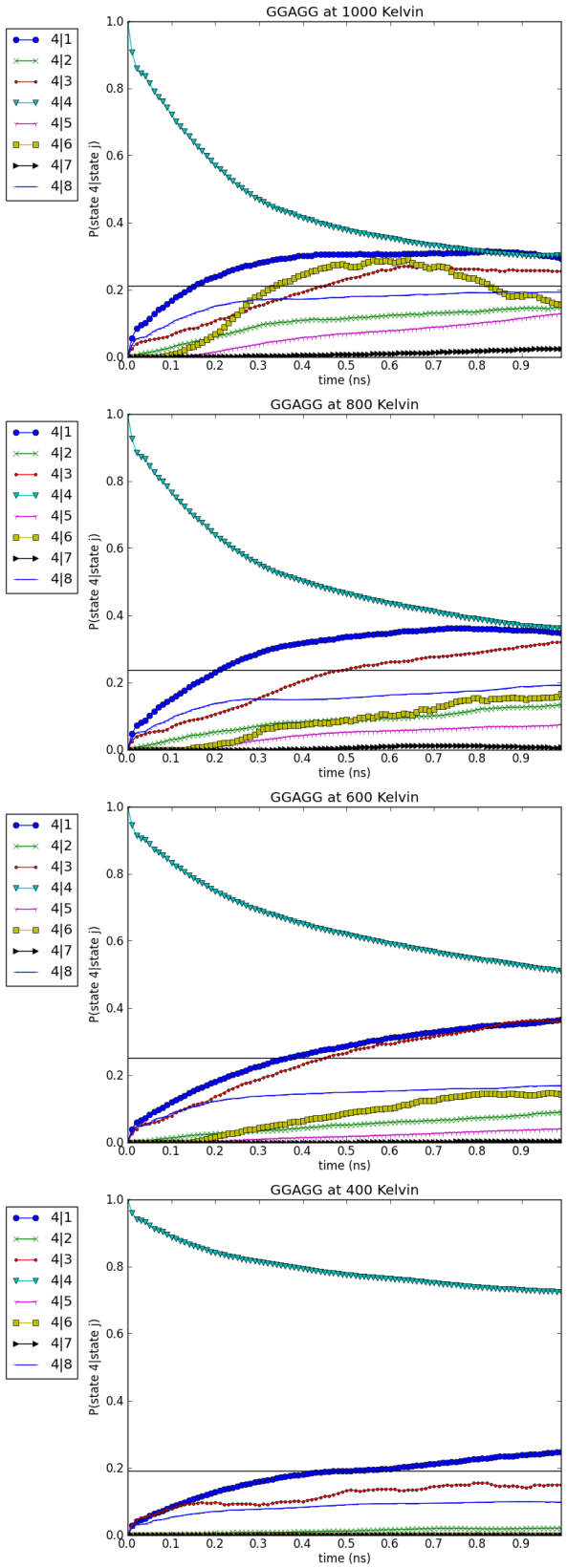


Fig. 4. Time-delayed conditional (transition) probabilities for Alanine at 1000 K, 800 K, 600 K, and 400 K. The probabilities are shown for all possible transitions from the eight states to the state 4. The black solid lines represent the equilibrium probability of the state 4 for Alanine.

front factor A_0.

$$r = A_0 \exp\{-E_a/RT\} \quad (10)$$

The activation energies are calculated from the heights of the saddle points in energy maps. The front factor A_0 is represented by Kramers' expression for high-friction Brownian motions [4]. It is the frequency of passing the energy barrier at a given temperature. Then the similarity transformation of the rate matrix A whose elements are the rates leads to the conditional probability matrix $C(t)$.

Alternatively, MD simulation trajectories of the chains may be used to determine the time-delayed conditional probability curves $C(t)$ [2]. We follow this approach and obtain the conditional probability curves for each amino acid. Then, the initial slopes of the curves give the transition rates following equation (5). Each curve converges to the equilibrium probabilities of the final state at infinite time.

Table I. The equilibrium probabilities of the amino acids based on 8 states

States:	1	2	3	4	5	6	7	8
ALA	0.14	0.08	0.03	0.33	0.21	0.01	0.05	0.13
ARG	0.17	0.11	0.02	0.30	0.19	0.01	0.07	0.13
ASN	0.20	0.11	0.03	0.31	0.18	0.01	0.05	0.11
ASP	0.21	0.05	0.04	0.37	0.19	0.02	0.03	0.08
CYS	0.22	0.16	0.03	0.23	0.16	0.01	0.05	0.14
GLN	0.17	0.10	0.02	0.33	0.21	0.01	0.04	0.12
GLU	0.17	0.04	0.04	0.40	0.21	0.01	0.04	0.09
GLY	0.10	0.07	0.14	0.09	0.07	0.13	0.16	0.25
HIS	0.16	0.06	0.02	0.37	0.23	0.01	0.04	0.11
ILE	0.16	0.10	0.01	0.39	0.22	0.01	0.03	0.08
LEU	0.14	0.08	0.03	0.34	0.23	0.01	0.05	0.11
LYS	0.20	0.13	0.03	0.27	0.18	0.01	0.05	0.12
MET	0.17	0.10	0.02	0.32	0.19	0.01	0.06	0.12
PHE	0.21	0.13	0.02	0.28	0.18	0.01	0.05	0.12
PRO	0.00	0.00	0.01	0.65	0.25	0.00	0.00	0.10
SER	0.21	0.13	0.04	0.22	0.19	0.02	0.06	0.15
THR	0.20	0.14	0.04	0.23	0.22	0.01	0.04	0.11
TRP	0.22	0.11	0.03	0.30	0.20	0.01	0.03	0.10
TYR	0.20	0.13	0.03	0.26	0.18	0.02	0.07	0.12
VAL	0.19	0.10	0.02	0.35	0.23	0.01	0.02	0.08

Calculation of the Time-Delayed Conditional Probabilities

Given state s_i at time $t = 0$, the probability of observing a state s_j at time t is given as

$$C(s_j, t; s_i, 0) = \frac{P(s_i, 0; s_j, t)}{P(s_i, 0)} \quad (11)$$

where $P(s_i, 0; s_j, t)$ is the joint probability that the system is in state i at $t = 0$ and in state j at time t.

For a chosen time step $t = \tau$, we record the joint observations of s_i and s_j in time interval τ throughout the simulation. Then, the joint probability that the system is in state s_i at $t = 0$ and in state s_j at $t = \tau$ is

$$P(s_i, 0; s_j, \tau) = \frac{N_{ij}}{N_{total}} \quad (12)$$

Here N_{ij} is the number of times when the system is in s_i at time t and in s_j at time $t + \tau$, and N_{total} is the total number of observations.

After repeating the counting scheme for different chosen $t = \tau, 2\tau, 3\tau, \ldots$ we obtain the joint probabilities. Then the joint probabilities lead determination of the time-delayed conditional probabilities.

The time-delayed conditional probability curves C_τ are obtained for each amino acid using this approach. The transition probabilities for all possible transition between eight states are represented. All of the curves are expected to approach to the equilibrium probability at infinite time. However, this condition is not satisfied in reasonable time for MD simulations at lower temperatures. Hence, the simulations are performed for higher temperatures as 400 K, 600 K, 800 K, and 1000 K. Then, the aim is to estimate the rates for 310 K based on the values obtained for higher temperatures.

In **Fig. 4**, the conditional probability curves are given for Alanine evaluated by MD simulations at 400 K, 600 K, 800 K, and 1000 K. It is evident that the curves converge to the equilibrium probability faster for the higher temperatures. Based on this fact, the calculations are performed for higher temperatures. Then, the dynamical behavior at 310 K is estimated for each amino acid.

Determination of the Rate Matrices

The initial slopes of the time-delayed conditional probability curves give the transition rate matrix elements. Since $C_\tau = \exp\{A\tau\}$,

$$\left.\frac{dC_\tau}{d\tau}\right|_{\tau=0} = A\exp\{A\tau\}\big|_{\tau=0} = A \quad (13)$$

One may derive the derivative at $\tau = 0$ using the finite differences. The forward difference is defined by

$$\Delta f(x) = f(x+h) - f(x) \quad (14)$$

Furthermore, the higher order differences are derived by

$$\Delta^k f(x) = \Delta^{k-1} f(x+h) - \Delta^{k-1} f(x) \quad (15)$$

Therefore, $f'(x=0)$ may be estimated by the sum of forward differences as follows

$$\left.\frac{df}{dx}\right|_{x=0} \cong \frac{1}{h}\left(\Delta f(0) - \frac{1}{2}\Delta^2 f(0) + \frac{1}{3}\Delta^3 f(0) + \cdots\right) \quad (16)$$

The rates are estimated using finite differences up to order 5, using only the first six data points of the conditional probability curves. After construction of the rate matrices, the matrices of eigenvectors and eigenvalues are determined using the eigenvalue decomposition (2)-(4). Then, any dynamic property may be obtained by going through equations (7) and (8).

Molecular Dynamics Simulations for Determination of the Rates

The same peptide model Gly-Gly-X-Gly-Gly and the same parameters that have been used in MD simulations for determinations of the states are used. However, this time we don't use a simulated annealing procedure since the aim is to collect data for the transitions happened during the simulations. 1 ns

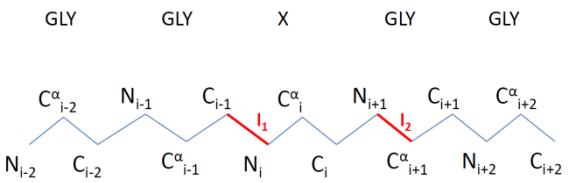


Fig. 5. l_1 and l_2 vectors of the GGXGG peptide

simulations are performed for each type of amino acids and the data collected in 1 ps periods.

Internal Flexibility

Any dynamic property associated with conformational transitions, $\langle f(\tau)\rangle$, may be computed by using the rate matrices. Similarity transformation of the rate matrix ($A = B\Lambda B^{-1}$) leads to the determination of a mean transient property as follows

$$\langle f(\tau)\rangle = \sum_{\gamma} k_{\gamma} \exp(\lambda_{\gamma}\tau) \qquad (17)$$

where λ_{γ} is the γ^{th} eigenvalue of Λ and k_{γ} denotes the amplitude factor given by

$$k_{\gamma} = \sum_{\alpha}\sum_{\beta} B_{\alpha\gamma}\left[B^{-1}\right]_{\gamma\beta} P_0(\Phi_{\beta}) f(\Phi_{\alpha};\Phi_{\beta}) \qquad (18)$$

We are interested in calculation of the local relaxation properties of the amino acids. The internal dynamics of a given residue X may be analyzed by a function that is associated with conformational transitions taking place between the torsion states. Hence, we consider two vectors; (i) l_1 between the atoms C_{i-1} and N_i, and (ii) l_2 between the atoms N_{i+1} and C^{α}_{i+1} of the backbone of $Gly_{i-2}-Gly_{i-1}-X_i-Gly_{i+1}-Gly_{i+2}$ for this purpose. The conformations of the vectors will change with time related to the values of (φ,ψ).

For this purpose we define two functions;

i. $\langle f_1(\tau)\rangle$ that gives the average cosine of the angle between l_1 vectors at $t=0$ and $t=\tau$ when the vector l_2 is kept fixed,

ii. $\langle f_2(\tau)\rangle$ that gives the average cosine of the angle between the cross product of the two vectors, $l_3 = l_1 \times l_2$, that gives information on conformational change in the normal.

Here, $\langle\ \rangle$ means averaging over all time steps. Then, $\langle f_1(\tau)\rangle$ is evaluated by

$$\langle f_1(\tau)\rangle = \left\langle \frac{l_1(0)\bullet l_1(\tau)}{\|l_1\|\|l_1\|} \right\rangle = \sum_{\gamma} k_{\gamma} \exp\{\lambda_{\gamma}\tau\} \qquad (19)$$

$$k_{\gamma} = \sum_{i}\sum_{j} B_{i\gamma}\left[B^{-1}\right]_{\gamma j} P_0(s_j) \frac{l_1\bullet l_1(s_i;s_j)}{\|l_1\|\|l_1\|} \qquad (20)$$

Here $l_1\bullet l_1(s_i;s_j)$ is the dot product of the l_1 vector with itself when the transition takes place from state s_j to state s_i.

Similarly, the cross product of the two vectors, $l_3 = l_1 \times l_2$, is analyzed by calculating $\langle f_2(\tau)\rangle$.

$$\langle f_2(\tau)\rangle = \left\langle \frac{l_3(0)\bullet l_3(\tau)}{\|l_3\|\|l_3\|} \right\rangle = \sum_{\gamma} k_{\gamma} \exp\{\lambda_{\gamma}\tau\} \qquad (21)$$

$$k_{\gamma} = \sum_{i}\sum_{j} B_{i\gamma}\left[B^{-1}\right]_{\gamma j} P_0(s_j) \frac{l_3\bullet l_3(s_i;s_j)}{\|l_3\|\|l_3\|} \qquad (22)$$

Results and Discussion

The two correlation functions $\langle f_1(t)\rangle$ and $\langle f_2(t)\rangle$ over time for some amino acids (Asn, Cys, Gly, Trp) at T=400, 600, 800, 1000 Kelvin are given in Fig. S1. The decay curves presented in the figures show that these functions exhibit exponential decay behavior. Hence, it is possible to obtain the relaxation times by fitting the curves with formula $A+B\exp(-t/C)$. Then, the parameter C that gives the relaxation time τ_R is determined for all amino acids at different temperatures.

Table II. Estimated relaxation times for twenty amino acids at 310 Kelvin.

Amino Acid	Relaxation time1 (τ_1)	Relaxation time2 (τ_2)
ALA	0.33	0.40
ARG	0.32	0.38
ASN	0.36	0.48
ASP	0.29	0.33
CYS	0.28	0.28
GLN	0.54	0.81
GLU	0.39	0.56
GLY	0.19	0.22
HIS	0.45	0.61
ILE	0.55	1.12
LEU	0.27	0.33
LYS	0.49	0.70
MET	0.40	0.58
PHE	0.41	0.54
PRO	1.51	1.51
SER	0.23	0.21
THR	0.72	1.05
TRP	0.50	0.78
TYR	0.39	0.53
VAL	0.37	0.52

Arrhenius' equation is a mathematical expression that gives the relation between rate constants and temperature. The equation was proposed by Svante Arrhenius in 1889 [30] by means of experiments. That is

$$k = A\exp(-E_a / k_B T) \qquad (23)$$

where k is the rate constant of a chemical reaction. A is the pre-exponential factor or the frequency factor of the reaction. E_a is the activation energy, k_B is the Boltzmann constant and T is the absolute temperature. Taking the natural logarithm of the equation yields

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

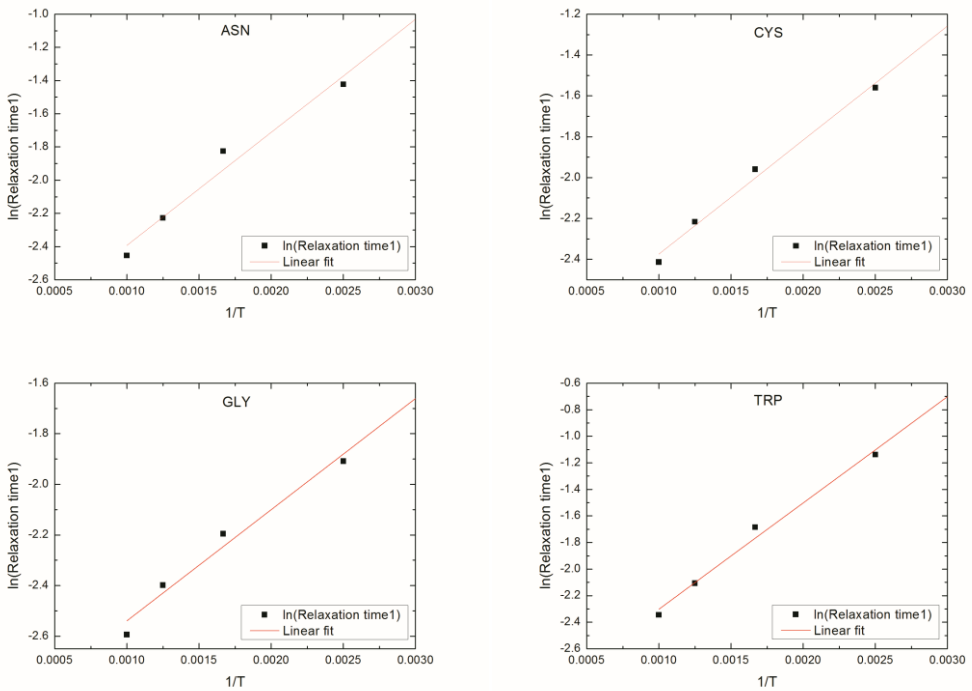


Fig. 6. Relaxation times for amino acids Asn, Cys, Gly, and Trp calculated via <f1(t)>. Red line represents the linear curve fitting.

$$\ln(k) = \ln(A) - \frac{E_a}{k_B T} \qquad (24)$$

When $\ln(k)$ is plotted as a function of $1/T$, this relation, called the Arrhenius equation, gives a straight line since it is in the form of $y = mx + b$.

Hence, our aim is to estimate the straight lines by linear least squares fitting technique after obtaining the relaxation times. These lines determined for each amino acid at different temperatures lead to relaxation times at 310 K by extending the lines until $1/T = 1/310 \approx 0.003$.

The relaxation times determined via fitting $\langle f_1(t) \rangle$ to exponential decays are shown in **Fig. 6** for some amino acids. Curve fitting the natural logarithm of relaxation times to linear lines yield the red lines shown in the graphs. The relaxation times at T=310 K are estimated by this lines since the dynamic functions obey Arrhenius' equation. Similarly, the relaxation times obtained via fitting $\langle f_2(t) \rangle$ are presented in **Fig. 7**.

In Table II, the estimated relaxation times τ_1 and τ_2 for T=310 K are presented. Glycine and Serine that are known as small amino acids relax faster than the other amino acids. Glycine is the least restricted amino acid while Proline has restrictions in Ramachandran map and has the largest relaxation time.

Isoleucine and Threonine relax more slowly than the other amino acids. These amino acids are restricted in the conformations since they contain two non-hydrogen substituent attached to their C-beta carbon. They are known as C-beta branched amino acids. Hence, it is not surprising that these amino acids relax slower than the other amino acids.

Huang and Nau determined the fluorescence lifetimes of unstructured peptides as a function of the amino acid type [20]. It has been reported that the quenching rate constant is a scale for the flexibility of amino acids. The rate constants of sixteen amino acids are given except for Trp, Tyr, Cys, and Met since these four amino acids themselves are known as efficient quenchers [31].

In **Fig. 8**, we compare our results with the fluorescence results of Huang and Nau [20]. In general, the results are in good agreement. The quenching rates are determined as a measure for flexibility and the results are related with the residue size. In **Fig. 9**, we provide a comparison of quenching time with molecular weight. We removed Proline in the figure since it is an outlier with a relaxation time value of more than 10 microseconds. The quenching times of Ile and Val are higher than the others. These amino acids are β-branched amino acids that have restricted main chain conformations. On the other hand, β-branched Thr seems to be more flexible than Ile and Val. Huang and Nau proposed that since Thr has a secondary hydroxyl group, it limits the flexibility of Thr but much less than Val and Ile. Furthermore, if we compare Leu and Ile; Leucine seems to be more flexible than Isoleucine. Huang and Nau stated that this is because Ile is a β-branched amino acid.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

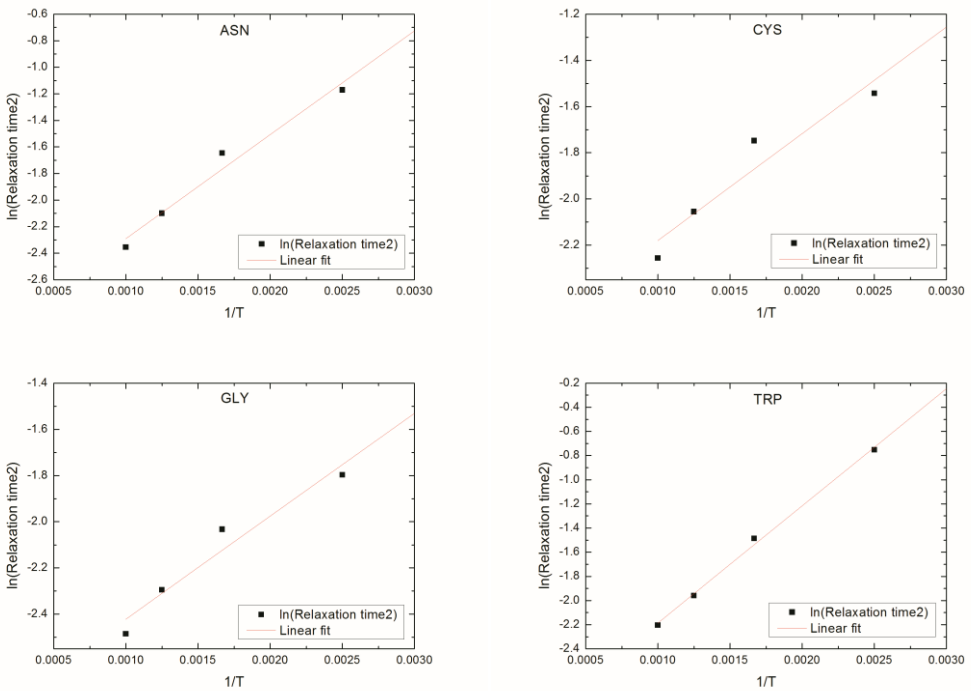


Fig. 7. Relaxation times for amino acids Asn, Cys, Gly, and Trp calculated via <f2(t)>. Red line represents the linear curve fitting.

The scale between relaxation time determined by DRIS and molecular weight is given in **Fig. 10**. The relation is determined after excluding Thr since it is an outlier. Here two C-beta branched amino acids, Thr and Ile, relax more slowly than the other amino acids while the other C-beta branched amino acid Val relaxes faster than these and our explanation parallels that of Huang and Nau.

The order of relaxation time τ_1 is Gly< Ser< Leu< Cys< Asp< Arg< Ala< Asn< Val< Tyr,Glu< Met< Phe< His< Lys< Trp< Gln< Ile< Thr< Pro. Similarly, the order of relaxation time τ_2 is Ser< Gly< Cys< Asp, Leu< Arg< Ala< Asn< Val< Tyr< Phe< Glu< Met< His< Lys< Trp< Gln< Thr< Ile< Pro.

Although the results fit together with the data given by Huang and Nau [20], there are some differences. This may be a consequence of different macrocyclization probabilities of different amino acid sequences, which plays a role in the experiments of Huang and Nau but is not consequential in our single amino acid calculations. The proportion of macrocyclic constituents at equilibrium is related to the macrocyclizatiion probability that is the probability of coincidence of the two ends of a sequence [32-34]. Hence, it effects the statistical configurations of the chain. Therefore, it is related to chain flexibility and the choice of chain length may affect measurements of the quenching.

Conclusion

The rotational transitions between the torsion states of the bonds of an amino acid have an important role in local dynamics. DRIS is an efficient computational method that yields determination of a dynamic property of the chain as a function of local transitions based on the frequencies of each state.

In this study, the rates of first order transitions between the specified states are determined by the conditional probability curves that obtained from MD simulation trajectories of amino acids. Since the probability curves converge to the equilibrium probabilities faster at higher temperatures, the calculations are performed for 400, 600, 800, and 1000 K. The initial slopes of the time-delayed conditional probability curves give the transition rates that are implemented in the rate matrices. The identified dynamic properties $\langle f_1(t)\rangle$ and $\langle f_2(t)\rangle$ are determined based on the DRIS method. Since these functions decay exponentially, respective single exponent relaxation times at different temperatures are obtained.

Our calculations showed that $\langle f_1(t)\rangle$ and $\langle f_2(t)\rangle$ have Arrhenius behavior. Based on this observation, we extrapolated the rates for higher temperatures and obtained the relaxation times at 310 K for all amino acids.

Relaxation times of the amino acids are related to the flexibility of the molecules, smaller relaxation times indicating higher flexibility while larger relaxation times indicating higher rigidity.

We applied the DRIS method to analyze the internal dynamics of the amino acids. The DRIS method is a general method that can

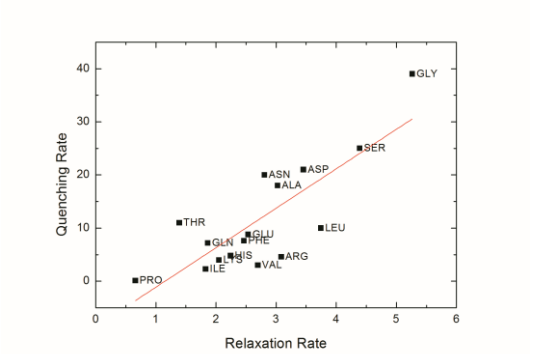


Fig. 8. Comparison of the relaxation rates determined via DRIS with the quenching rates obtained experimentally by [20].

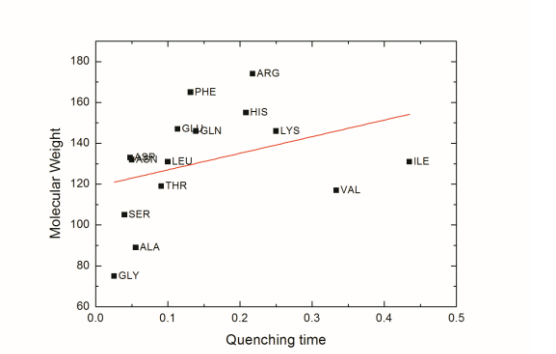


Fig. 9. The relation between quenching time [20] and molecular weight of the fifteen amino acids. Pro is excluded.

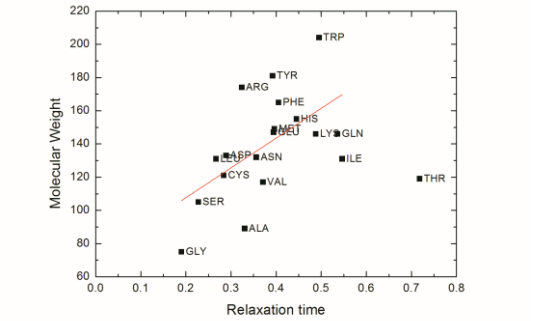


Fig. 10. The relation between relaxation time determined by DRIS and molecular weight of all amino acids except Pro and Thr. The relation is determined by excluding Thr since it is an outlier.

be applied to determine the internal dynamics of the peptides. One may adopt the same method to analyze dynamics of different length peptides in the random coil state [13].

Notes and references

Computational Science and Engineering Program, Koc University, 34450, Sariyer, Istanbul, Turkey. Tel: +90 212 338 1739; Fax: +90 212 338 1548; E-mail:csevim@ku.edu.tr

† Electronic Supplementary Information (ESI) available: Supplementary Figure S1. See DOI: 10.1039/b000000x/

1. I. R. Kleckner and M. P. Foster, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2011, **1814**, 942-968.
2. T. Haliloglu, I. Bahar, B. Erman, E.-G. Kim and W. L. Mattice, *The Journal of Chemical Physics*, 1996, **104**, 4828-4834.
3. M. B. Hamaneh, L. Zhang and M. Buck, *Biophysical journal*, 2011, **101**, 196-204.
4. I. Bahar and B. Erman, *Macromolecules*, 1987, **20**, 1368-1376.
5. P. J. Flory, *Statistical mechanics of chain molecules*, Interscience Publishers, New York,, 1969.
6. R. L. Jernigan, in *Dielectric Properties of Polymers*, ed. F. Karasz, Springer US, 1972, pp. 99-128.
7. I. Bahar, B. Erman and L. Monnerie, *Macromolecules*, 1989, **22**, 2396-2403.
8. I. Bahar and B. Erman, *Macromolecules*, 1987, **20**, 2310-2311.
9. I. Bahar, B. Erman and L. Monnerie, *Macromolecules*, 1989, **22**, 431-437.
10. I. Bahar, *The Journal of Chemical Physics*, 1989, **91**, 6525-6531.
11. I. Bahar and W. L. Mattice, *Macromolecules*, 1990, **23**, 2719-2723.
12. I. Bahar, N. Neuburger and W. L. Mattice, *Macromolecules*, 1992, **25**, 4619-4625.
13. A. M. Ovacik, Master's thesis, Koc University, 2005.
14. D. A. C. Beck, D. O. V. Alonso, D. Inoyama and V. Daggett, *Proceedings of the National Academy of Sciences*, 2008, **105**, 12259-12264.
15. K. Plaxco, C. Morton, S. Grimshaw, J. Jones, M. Pitkeathly, I. Campbell and C. Dobson, *J Biomol NMR*, 1997, **10**, 221-230.
16. S. Schwarzinger, G. A. Kroon, T. Foss, P. Wright and H. J. Dyson, *J Biomol NMR*, 2000, **18**, 43-48.
17. L. Ding, K. Chen, P. A. Santini, Z. Shi and N. R. Kallenbach, *Journal of the American Chemical Society*, 2003, **125**, 8092-8093.
18. Z. Shi, K. Chen, Z. Liu and N. R. Kallenbach, *Chem Rev*, 2006, **106**, 1877-1897.
19. L. Wang and J. Markley, *J Biomol NMR*, 2009, **44**, 95-99.
20. F. Huang and W. M. Nau, *Angewandte Chemie International Edition*, 2003, **42**, 2269-2272.
21. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *Journal of Computational Chemistry*, 2005, **26**, 1701-1718.
22. W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey and M. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.
23. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *Journal of the American Chemical Society*, 1996, **118**, 11225-11236.
24. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684-3690.
25. S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science*, 1983, **220**, 671-680.
26. P. A. Karplus, *Protein Sci*, 1996, **5**, 1406-1420.
27. E. B. Unal, A. Gursoy and B. Erman, *Phys Biol*, 2009, **6**, 036014.
28. B. Unal, A. Gursoy and B. Erman, *PloS one*, 2010, **5**, e10926.
29. C. S. Bayrak and B. Erman, *Molecular BioSystems*, 2012, **8**, 3010-3016.
30. K. J. Laidler, *Journal of Chemical Education*, 1984, **61**, 494.
31. R. R. Hudgins, F. Huang, G. Gramlich and W. M. Nau, *Journal of the American Chemical Society*, 2002, **124**, 556-564.
32. P. J. Flory and J. A. Semlyen, *Journal of the American Chemical Society*, 1966, **88**, 3209-3212.

33. U. W. Suter, M. Mutter and P. J. Flory, *Journal of the American Chemical Society*, 1976, **98**, 5740-5745.

34. M. Mutter, U. W. Suter and P. J. Flory, *Journal of the American Chemical Society*, 1976, **98**, 5745-5748.