Thermodynamic and Spectroscopic Characterization of Near Native States in Protein Folding

by

PRAJNA MISHRA

10CC15J26011

A thesis submitted to the

Academy of Scientific & Innovative Research

for the award of the degree of

DOCTOR OF PHILOSOPHY

in

SCIENCE

Under the supervision of
Dr. Santosh Kumar Jha

CSIR- National Chemical Laboratory, Pune

Academy of Scientific and Innovative Research

AcSIR Headquarters, CSIR-HRDC campus
Sector 19, Kamla Nehru Nagar,
Ghaziabad, U.P. – 201 002, India

March 2021

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, "*Thermodynamic and Spectroscopic Characterization of Near Native States in Protein Folding",* submitted by *Prajna Mishra* to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of *Doctor of Philosophy in Science*, embodies original research work carried-out by the student. We, further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material(s) obtained from other source(s) and used in this research work has/have been duly acknowledged in the thesis. Image(s), illustration(s), figure(s), table(s) etc., used in the thesis from other source(s), have also been duly cited and acknowledged.

Prajna Mishra

(Signature of Student)
Prajna Mishra
Date: 05/03/2021

(Signature of Supervisor)
Santosh Kumar Jha
Date: 05/03/2021

Statements of Academic Integrity

I, Prajna Mishra, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. 10CC15J26011 hereby undertake that, the thesis entitled "*Thermodynamic and Spectroscopic Characterization of Near Native States in Protein Folding"* has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)".*

Prajna Mishra

Signature of the Student

Date : 05/03/2021

Place : Pune

It is hereby certified that the work done by the student, under my/our supervision, is plagiarism-free in accordance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for "*Ethics in Research and in Governance (2020)".*

Signature of the Supervisor

Name: Santosh Kumar Jha

Date: 05/03/2021

Place: Pune

Acknowledgements

It is a great pleasure to present my research work in the form of this thesis. This work would not have been possible without the help and support of few people. I take this opportunity to express my sincerest gratitude to all those who helped me throughout this journey.

First and foremost, I am incredibly grateful to my guide Dr. Santosh Kumar Jha, for his invaluable advice, continuous support and patience throughout my Ph.D. journey. Not only he introduced me to this exciting field of protein folding, but he also provided me a platform where I can work with absolute freedom. His sincerity, hard work, optimistic outlook and confidence have always inspired me. He never stopped believing in me and always encouraged and motivated me at all the time of my research and writing this thesis. I have learnt a lot from him like how to think, formulate and have the right approach for solving a scientific problem, be self-critical while interpreting the experimental results, and explore new possibilities and define my path, to name a few. I believe there is still a lot to learn from him, and as my mentor, he has taught me more than I could ever encapsulate my words here. He has shown me, by his example, what a good scientist and person one should be. I believe working with him was a privilege and an excellent opportunity in my life.

I also express my sincere gratitude to my doctoral advisory committee members, Dr. Sayan Bagchi, Dr. J. Nithyanandhan and Dr. H. V. Thulasiram, for their constructive feedback, guidance and encouragement throughout my Ph.D. research.

Getting through my Ph.D. required more than academic support, and I have many people to thank for. A special thanks to my lab mates for providing me a lively work environment and for all the support and critical suggestions on various occasions. I did not realize when they transitioned from colleagues to great friends beyond the lab. This journey is incomplete without my lab members: Nirbhik, Meenakshi, Divya, Abhilasha, Sonal, Aman, Jivan, Minnu and Reshmi; and the extended lab members: Anjali, Anwesha and Avishkar.

I express my heartfelt thanks to all my friends in NCL for their love, care, cheerful company and support: Nirbhik, Ruchir, Deb, Singham, Parag, Priya, Monika and Kushal. They have made my stay on this campus memorable. I would surely treasure the days I spent with them in NCL.

I acknowledge UGC to provide the fellowship to pursue a Ph.D. in NCL and the Academy of Scientific and Innovative Research (AcSIR) for my enrolment in the Ph.D. programme. I gratefully

acknowledge the Director of CSIR-NCL, HoDs of the Physical and Materials Chemistry division, for allowing me to carry out my research and availing this institute's instrument facilities. I sincerely thank the staffs of the library, AcSIR office, Student academic office and all the administrative office for their unwavering support and cooperation.

I reserve for last my deepest gratitude to my family and all the friends beyond NCL for their unconditional love and support.

Prajna Mishra

Table of Contents

List of Figures

List of Tables

Synopsis of the Thesis

1) Introduction

The native structure of proteins is central to various functions performed in cells. An integral part of the structure-function paradigm of proteins is their inherent flexibility and dynamics. The dynamic interconversion between different conformational substates in the heterogeneous native state basin of the energy landscape enables a single protein molecule to perform multiple functions.[1-2] The dynamics among these near native states are assisted by the motion of different structural elements of a protein, out of which side-chains of amino acids hold a significant position due to their involvement in various functions such as molecular recognition and dynamic allostery.[3-4] Nevertheless, it has remained challenging to quantify the degree of side-chain conformational heterogeneity in the native state ensemble of proteins and the relative energetic contributions of the side-chain packing and the hydrophobic effect in protein stability. Recent studies envisage these native state populations as an ensemble of dry molten globular states (DMG).[5-7] DMG were conjectured to be expanded forms of the native state in which the core is dehydrated, the secondary structure is intact and side-chain packing is loose and flexible. However, a plethora of questions regarding thermodynamic contributions of physicochemical forces due to side-chain packing, the nature and extents of side-chain packing in the DMG-like native state ensemble remains to be answered. A detailed thermodynamic and structural understanding of these DMG-like near native states can open new horizons in the field of protein folding and engineering.

2) Statement of Problem

Flexibility and dynamics are the two strategies adapted by nature to enable proteins to perform complex cellular functions at reduced costs. The recently proposed conformational selection model postulates the pre-existence of multiple conformations of a protein and the binding of a ligand favors one conformation over others.[1] The arrangement of side-chains provides the groove or the interface where ligands can bind to the protein. Arrangement of side-chains in a protein can alter two important physicochemical forces, i.e., van der Waal's interaction and electrostatic interactions in

a protein that are crucial for its structure and function. In this context, understanding the nature, extent and thermodynamic contribution of side-chain packing in the native state ensemble becomes extremely essential.

3) Objectives

The following open questions were answered in the scope of this thesis work.

a) How mobile and different are the packing interactions in the core of the DMG as compared to the N state?
b) Are DMG a subset of native state ensemble?
c) Whether DMG-like properties can be localized only in the selective regions of the protein structure?
d) Does heterogeneity exists only on surfaces or is it extended to hydrophobic cores?
e) What is the temporal nature of fluctuations of protein side-chains in the hydrophobic core?
f) What structural adaptations stabilize the burial of a charged side-chain in the core of a protein?
g) How does the burial affect the thermodynamic stability of the protein?

4) Methodology

We used Human Serum Albumin (HSA) as the model protein to study the conformational heterogeneity due to side-chain packing. HSA is an all helical protein found abundantly in blood plasma and plays a distinct role in the transportation of metabolites, drugs and ions in the bloodstream and the maintenance of the pH of the plasma and extravascular fluids.[8] HSA maintains its native structure in the pH range of 5.0 – 9.0. We have site-specifically probed the energetics and structural composition of the native state ensemble of HSA using tools of thermodynamics and a battery of site-specific and high-resolution spectroscopic probes, including fluorescence resonance energy transfer (FRET), dynamic fluorescence quenching, red-edge excitation shift (REES), fluorescence anisotropy, and near- and far-UV circular dichroism (CD).

5) Summary

The summary of the working chapters are discussed as follows:

Chapter 2: In this chapter, we show, using multiple site-specific spectroscopic probes and tools of thermodynamics, that the native state ensemble of HSA contains an equilibrium intermediate (I) whose interdomain region resembles a DMG. In the I state, a tryptophan residue, present in a helical segment in the interdomain region is alternatively packed, but the global secondary structure and intradomain packing of the protein are native-like. The I state also has a more considerable interdomain distance, but the domain−domain interface is dry and molten. Our results indicate that hydrophobic desolvation and side-chain packing are decoupled during protein folding and that interdomain packing interactions have an important energetic contribution to protein stability. Dynamic interconversion between alternatively packed native-like states could be important for multiple allosteric and ligand binding functions of this protein.

Chapter 3: It has been extremely challenging to detect protein structures with a dynamic core, such as DMG, that remain in equilibrium with the tightly packed native state and that are important for a myriad of entropy-driven protein functions. Here, we detect the higher entropy conformations of HSA, using red-edge excitation shift experiments. We covalently introduced a fluorophore inside the protein core and observed that in a subset of the native population, the side-chains of the polar and buried residues have different spatial arrangements than the mean population and that they solvate the fluorophore on a timescale much slower than the nanosecond timescale of fluorescence. Our results provide direct evidence for the dense fluidity of protein core and show that alternate side-chain packing arrangements exist in the core that might be important for multiple binding functions of this protein.

Chapter 4: The contribution of electrostatic interactions in protein stability has not been fully understood. Burial of an ionizable amino acid inside the hydrophobic protein core can affect its ionization equilibrium and shift its pKa differentially in the native and the unfolded states of a protein and this coupling between the folding/unfolding cycle and the ionization equilibria of the ionizable residue can substantially influence the protein stability. Here, we studied the coupling of the folding/unfolding cycle with the ionization of a buried ionizable residue in HSA using fluorescence spectroscopy. A pH dependent change in the stability of HSA was observed in the near native pH range (pH 6.0 – 9.0). The protonation-deprotonation equilibrium of a single thiol residue that is buried in the protein structure was identified to give rise to the pH dependent protein

stability. We quantified the pKa of the thiol residue in the native and the unfolded states. The mean pKa of the thiol in the native state was upshifted by 0.5 units to 8.7, with respect to free cysteine in water, which was due to the burial of the thiol in the protein structure. Surprisingly, the mean pKa of the thiol in the unfolded state was observed to be downshifted by 1.3 units to 6.9 in comparison to free cysteine in water. These results indicate that some charged residues are spatially proximal to the thiol group in the unfolded state. Our results suggest that, in addition to the native state, electrostatic interactions in the unfolded state are important determinants of protein stability.

6) References

1. Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **2009,** *5* (11), 789-796.
2. Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem Rev* **2016,** *116* (11), 6516-6551.
3. Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* **2010,** *6* (5), 352-358.
4. Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J Am Chem Soc* **2017,** *139* (10), 3599-3602.
5. Neumaier, S.; Kiefhaber, T. Redefining the dry molten globule state of proteins. *J Mol Biol* **2014,** *426* (13), 2520-2528.
6. Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* **2010,** *78* (13), 2725-2737.
7. Baldwin, R. L.; Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* **2013,** *23* (1), 4-10.
8. Peters, T. U. *All about Albumin: Biochemistry, Genetics, and Medical Applications*. Academic Press: 1996.

Chapter 1.

The Native State Conformational Heterogeneity in the Energy Landscape of Protein Folding

1.1 Introduction

To function, a protein needs to fold into its distinct conformation(s) called the native (N) state. The process begins when a polypeptide chain starts self-assembling and ends with the formation of a characteristic three-dimensional N state. The spontaneity and the rapidity of the polypeptide chain to select the same unique conformation(s) over a myriad of available options has given rise to the commonly called "Protein folding problem". In one of the seminal works by Cyrus Levinthal, he pointed out that it would take an astronomical length of time for a polypeptide chain of 101 amino acid residues to find its native state over a vast number of available conformational space.[1-2] However, the real folding rates of proteins vary between microseconds to a few hundreds of seconds. This discrepancy between the estimated and the real folding times of proteins is known as the 'Levinthal Paradox' whose solution lies in proteins opting for defined folding pathways and not choosing random structures to fold to the native state.[3] In a folding pathway, a specific sequence of structural events is followed in which progressive stabilization of intermediate structures takes place.[4] However, it has remained extremely challenging to understand the nature, role and temporal order of the structural events associated with the folding pathways. Multiple theories and models have been put forward to understand the protein folding problem.[5-9] Experimental and theoretical studies have demonstrated that the structure of the protein develops hierarchically and the protein can follow several alternative trajectories to fold into its native state.[10-14] These alternate pathways arise due to the presence of structural heterogeneity at each step of the folding process. The heterogeneity encountered at each step (unfolded, intermediates and native) is an ensemble of several subpopulations existing in equilibrium with each other following the Boltzmann distribution of population. Of all, the nature and degree of conformational heterogeneity present in the native state ensemble are very poorly understood. The primary reason being these subpopulations are isostructural and isoenergetic and due to which most of the spectroscopic probes fail to detect them. The heterogeneity in the native state basin could arise due to the dynamics of multiple structural elements of a protein. The conformational ensemble sampled due to side-chain dynamics is a significant contributor to native state heterogeneity and are vital for numerous entropy-driven functions and dynamic allostery.[15-16] Few recent studies suggest that these conformational subpopulations could be like dry molten globules (DMG).[17-20] However, it has remained extremely challenging to detect and characterize these DMG-like states in the native state ensemble. Characterization and a detailed understanding of the

function of the native state ensemble can open new horizons in the field of protein folding and engineering.

This chapter accounts the current knowledge and understanding of the conformational heterogeneity experienced by globular proteins at various steps in the folding pathways and then narrows down and focuses on the heterogeneity existing in the near native states. The conformational heterogeneity sampled due to different structural elements of a protein in its native state basin is discussed in this review while emphasizing on the heterogeneity due to side-chain packing. Experimental methods used to detect these ensemble and their functional significance are also discussed briefly.

1.2 Pathways of protein folding and reduction of conformational heterogeneity

Multiple conceptual models have been put forward to understand the protein folding problem.[5] Nevertheless, in all the models, conformational heterogeneity is detected at all the stages a protein passes through. In the unfolded state, the heterogeneity could be due to native-like interactions,[21] non-native interactions,[22-23] prolyl isomerization[24] or non-prolyl peptide bond isomerization.[25] For various proteins like cytochrome C,[26] apomyoglobin,[27-28] ribonuclease A[29] and barstar,[30-31] structural heterogeneity has been detected in the early (milliseconds) collapsed intermediates and these intermediates are known to direct the folding process into different routes.[27] For all these proteins, coexisting populations of folding intermediates have also been observed.[27, 32-34] These heterogeneous ensembles of intermediates indicate not only towards multiple pathways of protein folding but also emphasize on the importance of folding conditions in the selection of a given folding trajectory.[34] Considering the unfolded state, which is a random coil-like structure, to be maximum structurally heterogeneous, and native state to be minimally heterogeneous due to its compact three-dimensional structure, heterogeneity is reduced significantly in the folding pathway.[35] Different models of protein folding and how these models envisage the reduction in heterogeneity in the folding pathway are discussed in this section (Figure 1.1):

i) **Framework Model:** This model proposes that the secondary structural elements are formed first due to local interactions in the early stages of protein folding.[6, 36-37] These local elements come together via arbitrary diffusion-collision until the

formation of a native-like tertiary structure.[12, 38-39] In this model, the heterogeneity is expected to reduce throughout the folding pathway.

ii) **Hydrophobic Collapse Model:** In this model, an initial clustering of hydrophobic residues occurs as the first step of the folding which significantly decreases the conformational search space.[7, 40-41] The formation of this entropically driven collapsed state then facilitates the development of the secondary and tertiary structures of the protein. Here a drastic reduction in heterogeneity occurs with the initial chain collapse. However, as the chain collapse is accompanied by the formation of many non-native interactions between different segments of the protein, some of the decrease in structural heterogeneity may occur at the later stage of the folding route.

iii) **Nucleation and Nucleation–Condensation Model:** In the nucleation model, a nucleation site is shaped initially in which secondary structures are formed by the key residues of the polypeptide chain.[8] The rest of the protein structure then develops around the nucleation site, which acts as a scaffold. The development of the folding nucleus is the rate-limiting step of the whole folding trajectory. The nucleation-condensation model is an extended version of the nucleation model which postulates that in the rate-limiting step, the formation of a folding nucleus and development of the tertiary structure of the protein happens in a concerted manner.[42-44] In these models, most of the conformational heterogeneity reduces in the first step, i.e., during the formation of the nucleation site.

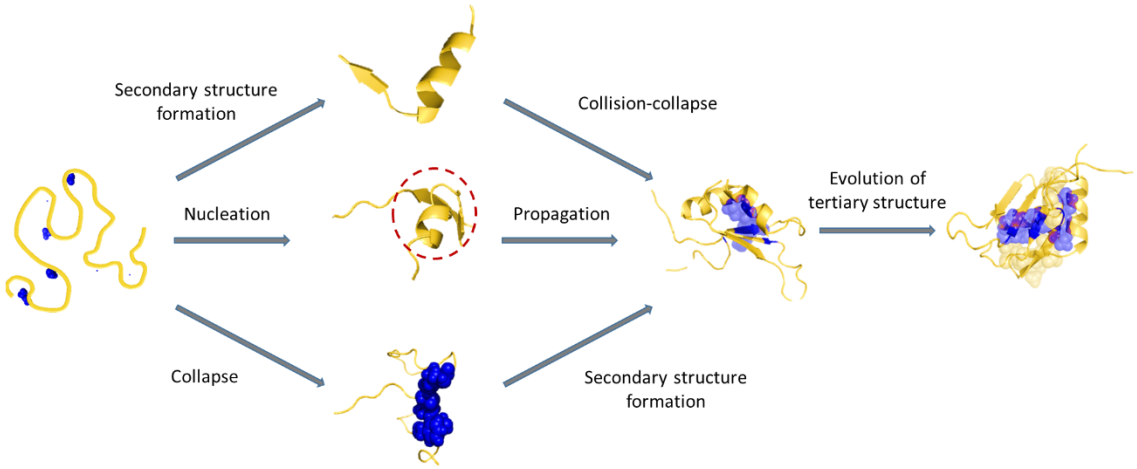


Figure 1.1. *Different models of protein folding. The framework model envisages the formation of the secondary structure before the development of the tertiary structure. In the nucleation*

model, a nucleation site of secondary structures is formed first, followed by the development of structure around the nucleus. The hydrophobic collapse model gives precedence to the clustering of hydrophobic residues before structure development. In the picture, yellow lines represent protein backbone and blue spheres represent hydrophobic amino acid residues. The red dashed circle represents the nucleation site in the nucleation model.

iv) **Energy Landscape Theory:** A relatively recent perspective on the protein folding mechanism, commonly called the Energy landscape model, envisages a funnel-shaped energy landscape which a protein follows while folding to its native state (Figure 1.2).[9, 45-47] In this model, an energy gradient prefers the native conformations and the folding is accompanied by a reduction in conformational entropy. This model also postulates multiple folding trajectories for a protein molecule and folding progresses via a series of intermediates. The intermediates are the ensemble of partially folded structures and the transition from one ensemble of structures to the other can happen in parallel ways.[48-49] The reduction in conformational heterogeneity in this model occurs gradually and the native state is supposed to be an ensemble of substates coexisting with each other.[50-51]

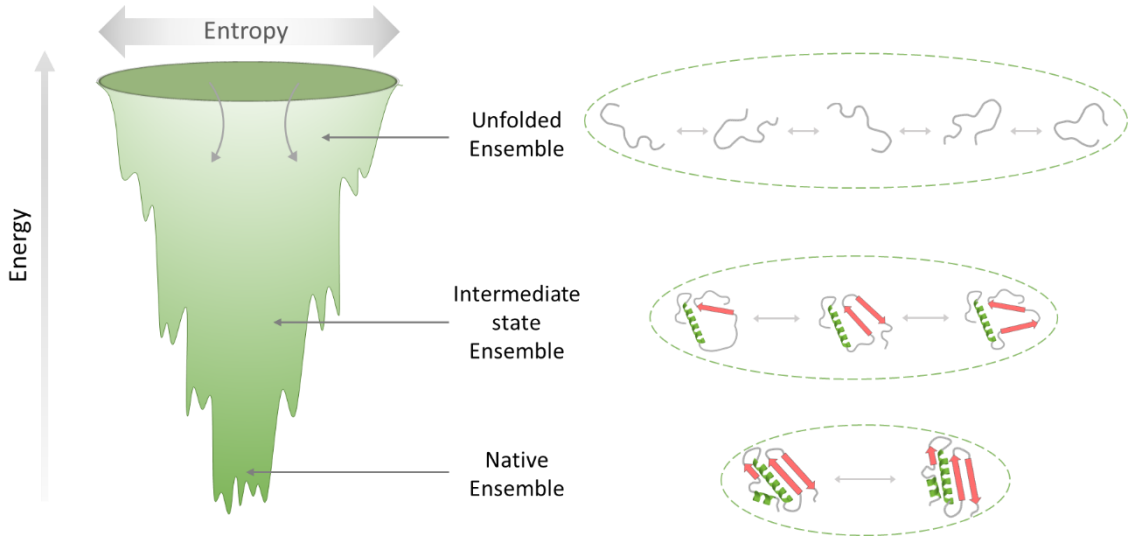


Figure 1.2. *The funneled view of protein folding in energy and entropy landscape. The unfolded protein can fold to its native structure via multiple intermediates by following different routes. This energy landscape projects an ensembles nature of various entities in the folding pathways.*

There is a reduction in conformational entropy and energy as the protein molecule travels downhill in the energy landscape.

1.3 Thermodynamic fluctuations are responsible for the conformational heterogeneity in the native state ensemble

The energy landscape model postulates that an ensemble of conformational subpopulations exists in equilibrium with the native states of the proteins.[50-52] These subpopulations follow a statistical thermodynamic distribution.[51] However, the physical principles that govern the heterogeneous behavior of the protein in its native state basin have not been fully understood. Protein molecules undergo considerable fluctuations around the mean value of their thermodynamic properties like internal energy, enthalpy, entropy and molecular volume.[53-54] The mean square fluctuation (Δ) around the internal energy (E) and total volume (V) which is a measure of the extent of distribution of these variables is given by:[53]

$$\overline{\Delta E^2} = k_B T^2 m C_V \qquad (eq\ 1.1)$$

$$\overline{\Delta V^2} = k_B T V \beta_T \qquad (eq\ 1.2)$$

where, k_B, T, m and β_T are the Boltzmann's constant, absolute temperature, mass and isothermal compressibility of the system, respectively. C_V is the heat capacity of the system at constant volume. For a representative protein molecule of 25 kDa (m = 4.2×10^{-20} g, V = 3.2×10^{-20} cm^3), the values of $\overline{\Delta E^2}$ and $\overline{\Delta V^2}$ will be 6.4×10^{-20} cal and 5×10^{-23} cm^3, respectively.[53-54] These intrinsic fluctuations in a typical protein molecule combined with the local fluctuations of water molecules[55] are sufficient to disrupt the folded structure to a vast number of conformational substates. Statistical information about all these conformational substates and their intrinsic fluctuations is provided by the energy landscape model.[50, 52] The native state basin of the energy landscape illustrates multiple energy valleys that are separated by energy barriers.[50-51] Each valley is a conformational substate and coexists with other substates in the native state basin. The temporal order of conformational exchange among these substates is determined by the height of the barrier between them.[56] The presence of these conformational substates gives a ruggedness to the bottom of the folding funnel which is measured by the thermodynamic fluctuations of the protein molecule. The ruggedness and the width of the global energy minima in the folding funnel regulate the conformational entropy of the native

state of the protein.[56] For example, if the native state basin is shallow, the thermodynamic fluctuations are expected to be larger and the number of conformational substates is more.

For a protein to function, conformational exchange is necessary among these substates which are governed by different types of protein motions. One of the widely accepted models of protein motions derives its concepts from the physics of glasses and supercooled liquids, where a substate of a protein is similar to a polymorph of a glass.[57-58] This model postulates that there exists three types of motions in a protein molecule: α fluctuations, β_h fluctuations and vibrations.[59-60] α fluctuations principally regulate the conformational change in the protein molecule and the rates of α fluctuations are inversely proportional to the viscosity of the solvent.[61] For example, they control the opening and closing of the entrance of the heme pocket in myoglobin.[62] β_h fluctuations in a protein molecule are primarily due to the hydration shell around it and regulate some internal channels. These fluctuations are absent in a dehydrated molecule.[61, 63] Vibrational motions in a protein molecule are very large and important for conformational exchange among protein substates. All these motions combined together enable proteins to perform a multitude of functions.

1.4 Structural motions in the native state ensemble

Due to the thermodynamic fluctuations, protein molecules exist in a multitude of possible conformations in their native state. The major information about native protein structure are obtained from the crystal structures which often portray a static image of the protein molecules where packing density in the protein resembles with the crystalline amino acids and solids.[64-66] However, the crystal structures are more often affected by the conditions adopted during the crystallization process and, therefore, it is not necessary that they project the most populated or functionally active conformations of the protein. Nevertheless, there exists dynamic equilibrium among the conformational substates of the native proteins and the dynamics among these conformations are assisted by motions of various structural modules of the protein molecules which include loop dynamics, side-chain rotations, inter-domain linkers and hinge movements, collective motions of interconnected parts including backbone motions, etc.[67]

i) **Loops:** Loops are significant structural elements that impart flexibility and plasticity to the protein molecules. Loop motions can have significant functional roles such as at the binding sites where the movements of the loops facilitate ligand binding.[68-69] These motions can also have secondary roles like

chaperoning the passage of ligands from secondary to primary binding sites or having correlated dynamics in transmitting signals from one site to another.[68-69] Nevertheless, the presence of numerous loop structural arrangements in the native state ensemble has been detected in several theoretical and experimental studies. For example, in the binding pocket of a bacterial protein, tyrosine phosphatase, the opening and closing of a flexible 'WPD loop' plays a central role in its function.[70] The open and closed conformations of the loop have been found to alternate with each other with a rate constant of ~2.8×10^8 s^{-1} as observed from both theoretical and experimental studies.[70-71] In a protein, EphA4, the dynamics of the loops present in its ligand-binding domain modulate its conformational heterogeneity and specificity to bind several ephrin ligands and small molecules as revealed by MD simulations and NMR-HX experiments.[72] Similarly using NMR experiments, Wright and co-workers have observed that in the solution state, the active site Met20 loop in *E. coli* dihydrofolate reductase exists in either closed or occluded conformations which regulate different phases of the catalytic cycle of the enzyme.[73]

ii) **Linkers and hinges:** In the case of multi-domain proteins, different modules are linked by stretches of amino acids with disordered secondary characteristics called linkers. Besides connecting different domains, linkers associate in the cooperative movement of protein-protein and inter-domain interactions and provide communication among different functional parts of proteins.[74-76] Hinges, on the other hand, act as a screw axis around which parts of a protein can move as rigid bodies. Hinge movements are usually confined to loops and linker regions due to their flexibility and are highly significant for protein function.[68, 74] They help in major conformational changes in a protein without affecting the packing of individual domains and, thus, have a minimal impact on the thermodynamic stability of the protein. Owing to their flexibility, linkers and hinge motions can populate a large number of protein conformations and also act as binding sites for a multitude of ligands.[75, 77] For example, the active site of bacteriophage T4 lysozyme lies in the hinge region between its N- and C- terminal domains and the substrate access to the active site is gained by the equilibrium dynamics of the domain movements.[78-79] An array of intermediate states of this protein associated with the hinge movement have been observed

in fluorescence correlation analysis, small-angle X-ray scattering experiments and molecular dynamic simulations.[79-81] Lactoferrin is another example in which the N- and C-terminal lobes consist of two domains each.[82] In each lobe, the domains are attached by a flexible linker and the iron-binding sites are positioned at the domain interfaces.[82] Due to the high flexibility of the linker, an array of various open and partially closed conformational states are populated by the apo form of the lactoferrin and upon binding to the metal, the population shifts to a closed-form through hinge motions.[82-84] In another study, the conformational space of Lys63-linked diubiquitin was explored by Tang and co-workers using paramagnetic relaxation enhancement NMR spectroscopy, where they have observed that this protein populates several open and closed conformational states and a target protein can select and stabilize one of the pre-existing states.[85]

iii) **Backbone dynamics:** The synchronized vibrational motions of many atoms in a protein are known to provide preferential direction in many enzymatic catalysis processes.[86-88] The functionally relevant collective motions that occur at low frequencies (μs to ms timescale) are physically originated from the atomic fluctuations happening in the ps to the ns time domain.[89] For example, in adenylate kinase, the opening and closing of the nucleotide-binding lid (μs to ms timescale) are essential for its function. Using NMR spectroscopy and MD simulations, Kern and co-workers have observed an increase in the dynamics (in ps timescale) of a local backbone whose conformation must change for the closure of the nucleotide-binding lid.[89] In a different study, Kern and co-workers characterized the dynamics of nitrogen regulatory protein C in three different functional states using NMR and observed a strong correlation between the backbone dynamics (in μs timescale) and the phosphorylation driven activation of the protein.[90] Interestingly, the backbone movement occurred in the regions which underwent the maximum structural change upon phosphorylation.[90-91] In ubiquitin, correlation in backbone movement was observed between distant sites which is very essential in molecular recognition mechanisms and dynamic allostery.[92] In another study on ubiquitin, a collective global motion was observed that was explicitly related to a conformational switch away from the binding site.[93]

iv) **Secondary structure motion:** The dynamics of secondary structural elements can also sample conformational substates in the native state basin of a protein. For example, for a 62 residue protein gpW, a sparingly populated state was observed in equilibrium with the native state in which α helices were formed but the β hairpin was unfolded.[94]

v) **Side-chain dynamics:** The rotation of side-chains in a protein is also a substantial contributor to its conformational heterogeneity. The movement of side-chains regulates two important physicochemical forces present in proteins: van der Waal's forces of interactions and electrostatic interactions. The interplay of these two forces are crucial determinants of the thermodynamic stability and functions of proteins. Both experimental and computational studies have demonstrated the presence of conformational heterogeneity due to side-chain dynamics in multiple proteins such as ubiquitin, RNase H and β-lactamase.[95-97] In the next section (Section 1.5), the heterogeneity sampled in the native state ensemble due to side-chain packing is discussed in more detail.

The complete description of conformational heterogeneity present in the near native state in proteins requires the knowledge of timescales associated with the movement of different elements. The energy landscape not only defines the probabilities of the conformational substates, but also the time associated with the transition between them. The timescale of different dynamic processes can range in a multitude of orders of magnitude (fs to s) as depicted in Figure 1.3.[56] The slow timescale or Tier 0 dynamics corresponds to μs – s time range conformational fluctuations at physiological temperature and are usually collective or larger domain motions. Owing to the slow dynamics, these motions sample a relatively lesser number of states and can be detected in equilibrium and kinetic experiments. Tier 1 and tier 2 dynamics correspond to fast timescale fluctuations and populate a relatively larger number of substates. Structural changes such as loop motions and side-chain rotations happen in the ps – ns time domain whereas bond vibration occurs in fs – ps timescale.

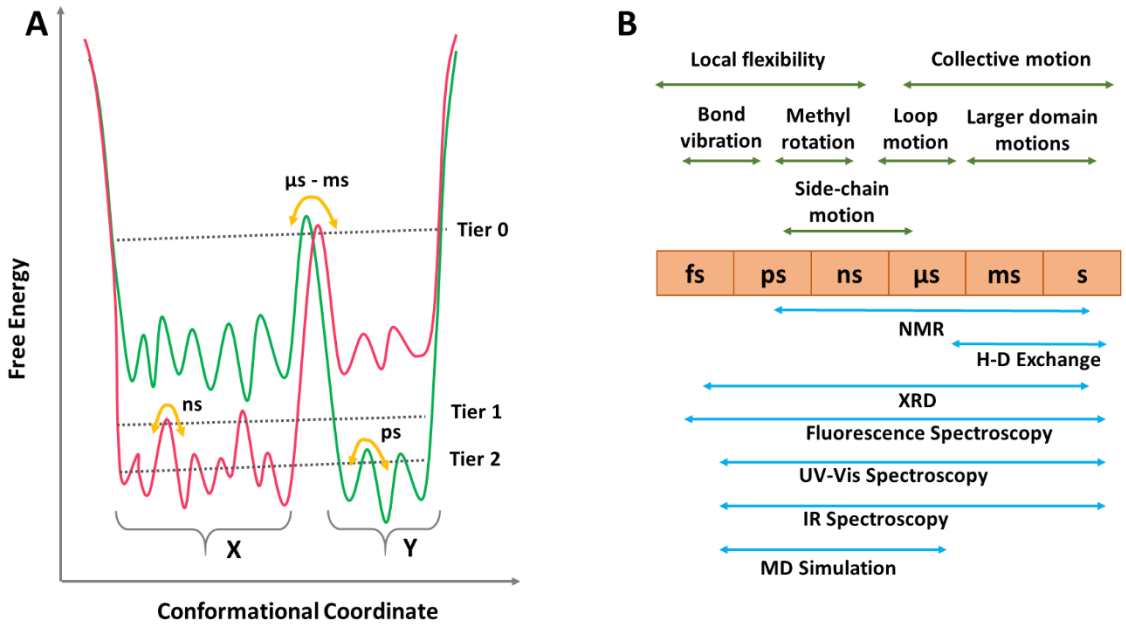


Figure 1.3: *The relative probabilities and timescale of conversion between conformational substates of a protein are defined by its energy landscape. (A) The energy landscape of a protein presenting showing the order of conformational dynamics and the energy barriers between substates. Tiers are classified as suggested by Frauenfelder and co-workers.[98] Tier 0 depicts slow timescale and tiers 1 and 2 represent fast timescale fluctuations. X and Y are two tier 0 states and a change in the system can change the landscape as shown by red and green lines. (B) Timescale associated with the structurally dynamic processes and the experimental methods that can detect these processes.*

1.5 Conformational heterogeneity in the native state ensemble due to side-chain packing

The conformational diversities in a protein due to the dynamics of different structural modules and their hierarchical stability can lead to a complex free energy landscape. Side-chains are one of the significant structural elements of a protein whose dynamics contribute substantially to the number of substates sampled at the native state basin. The interplay of physicochemical forces associated with these motions is significant for the thermodynamic stability of the native protein. In addition, these motions play an important role in many entropy-driven functions of a protein such as molecular recognition and dynamic allostery.[15-16] However, the extent of conformational heterogeneity due to side-chain dynamics in the native state basin of a protein and the nature of the structural substates sampled due to side-chain movement are very poorly understood.

Crystal structures show that the side chains are packed tightly in the stable native state of a protein and experience limited degrees of freedom. This has led to the traditional belief that proteins are like aperiodic crystals[99] and their densities resemble solid hydrocarbons.[100] This has been challenged recently by multiple studies. For example, using molecular dynamic simulations, Karplus and co-workers have observed proteins to be surface molten solids.[101] The liquid-like surface of a protein helps in function while the solid-like interior maintains the stability. Subsequent studies revealed that side-chains of a protein can have considerable dynamics with respect to a more rigid backbone.[95, 102-103] For example, Vendruscolo and co-workers have observed liquid-like mobility of the side-chain atoms in human ubiquitin.[95] The mobility of the side-chains, even in the core, was observed to be substantial in the context of backbone fluctuations. In another study by Geissler and co-workers on 12 small proteins, the authors observed fluctuations among the side-chains, even with a fixed backbone constraint.[103] Interestingly, the same group used Markov state models on three proteins (ubiquitin, RNase H and β-lactamase) and observed that every residue visits at least two rotamers in hundreds of μs time frame.[96] The rotameric transition rate could range from ns to tens of μs. Nevertheless, the observation that the core of the proteins, despite their high packing fractions, had some residual entropy and were like dense fluids challenged the pre-existing notion that protein cores are similar to organic solids.[96] These results, altogether, confirmed that heterogeneity due to side-chain packing in proteins exists to different extents and the transitions can happen on different timescales.

Understanding the nature of these conformationally heterogeneous ensemble due to side-chain dynamics has remained challenging. Recent studies advocate these substates could be like dry molten globules (DMG).[17-20] In DMG-like states of a protein, the core remains dehydrated but the side-chain packing is loose and molten while the protein possesses an intact secondary structure.[18] Using phase diagrams, it was initially hypothesized that proteins encounter DMG-like transition states during unfolding.[104-105] Experimentally, DMG have been detected in kinetic studies as transient intermediates during the unfolding of few proteins.[106-110] In equilibrium studies, DMG-like states have also been observed for multiple proteins.[111-115] The growing evidence for DMG-like states during the unfolding of multiple proteins indicates that they possibly lie on the native side of the energy barrier in the folding funnel. The DMG-like intermediates can be populated by perturbing the system using pH,[111-113] pressure[114, 116] and denaturants.[108-110] Interestingly, few equilibrium unfolding studies also indicated the presence of DMG-like intermediates in equilibrium with the native protein. For

example, in the case of human serum albumin (HSA), DMG-like intermediates in equilibrium with the native conformation were observed during the pH-induced unfolding.[112-113] However, the structure of the DMG-like intermediate detected during the acid-induced unfolding was considerably different than the base-induced unfolding intermediate. HSA is an all helical, three-domain protein. In the DMG-like intermediate, during the acid-induced unfolding, the side-chain packing of the two domains was broken while the secondary structure was intact and native-like.[113] Whereas in the case of base-induced unfolding intermediate, only the side-chain packings of the inter-domain region were loosened while the intra-domain packings were intact as shown in Figure 1.4.[112] Nevertheless, these results indicated that the tertiary structure loosening can happen differently in different parts of a protein (Figure 1.4). In other words, conformational substates may not have uniform DMG-like character and side-chain loosening can be localized to selective regions of a protein. In the case of ubiquitin, using pressure perturbation studies, DMG-like members were observed in the native state ensemble.[114] Apart from HSA and ubiquitin, a DMG-like state has been observed in equilibrium with the native state of a small protein, HP35.[115] The DMG-like unlocked the native state of HP35 showed increased conformational flexibility than the native state. Further studies on the DMG-like state of HP35 revealed that this state was alternatively packed and compact in comparison to the native state whereas the transition state for the native ⇌ DMG-like state had expanded volume.[17] These results suggest that the native state ensemble can contain several compact and expanded folded states.

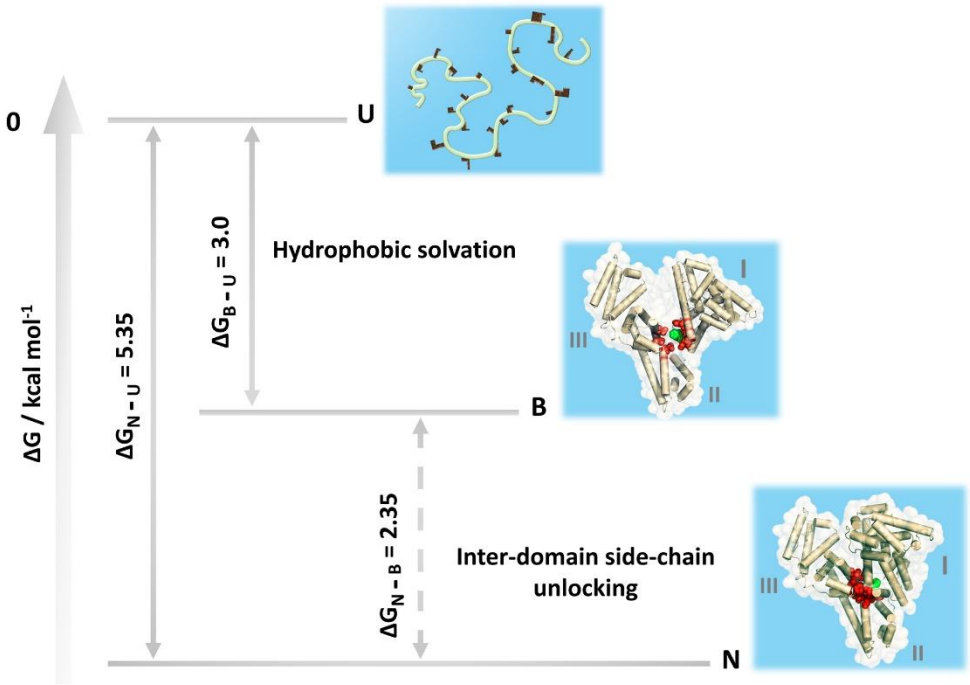


Figure 1.4. *Model energy diagram of HSA showing the relative energies of the native form (N), the base induced form (B) and the unfolded form (U).*[112] *In the structural models of the N form and the B form, the cylinders represent a-helical secondary structures, red spheres represent the side-chains of the residues involved in the packing of the inter-domain interface, and the white surface around the protein shows the solvent accessible surface area of the protein. The model of the U form shows only the peptide backbone (white thread) and the representative side-chains (black color) of the unfolded protein. In the B form, the inter-domain packing is disrupted without hydrophobic solvation.*

The role of side-chain conformational entropy in protein folding and thermodynamic stability of the native state has not been clearly understood.[117-120] The side-chains of amino acids are packed tightly in proteins in order to maximize the van der Waal's force of interactions. However, the contribution of van der Waal's interactions in protein stability is not yet clear.[121-124] It is commonly believed that, of all the physicochemical forces contributing to the thermodynamic stability of the native state, the hydrophobic effect is the maximum contributor.[125-128] Hydrophobic interactions refer to the exclusion of hydrophobic amino acid residues from water and their burial in the core of the protein. The packing interactions among the side-chains of all the residues are the contributors to the van der Waal's force of interactions in the protein. In the case of DMG-like states, the side-chain packing is broken but the hydrophobicity of the core is intact. Upon populating and characterizing the DMG-like states in equilibrium conditions, the thermodynamic contribution of van der Waal's interactions in comparison to the hydrophobic effect can be dissected (Figure 1.4).[18, 112-113]

Other than van der Waal's interactions, side-chain rotations can influence electrostatic forces of interactions in a protein such as salt bridge formation.[129-130] The side-chains of polar and charged amino acid residues are the major contributors of electrostatic interactions in a protein. Rotation of side-chains can modulate the interactions among different ion pairs in a protein, thereby affecting the overall protein structure. For example, Nussinov and co-workers analyzed the stabilizing and destabilizing effects of ion pairs in a study comprising of 11 proteins.[129] Each protein in the study was known to exist in at least 40 different conformations. It was observed that the electrostatic contributions of ion pairs to the protein stability varied substantially among the different conformers of each protein. The observed result was due to variabilities in the location and orientation of the charged residues in each protein

conformation. The study also revealed that, due to different side-chain movements, in some conformations of a protein salt bridges may exist and in others, they may not.[129]

1.6 Experimental and theoretical methods to detect conformational ensemble due to side-chain heterogeneity

The spectrum of conformations sampled due to side-chain dynamics in the native state basin of a protein is highly heterogeneous. However, it has remained extremely challenging to collect the structural and dynamic information of these heterogeneous ensembles. Multiple theoretical, experimental and spectroscopic methods have been developed to map the conformation and dynamics of proteins due to side-chain packing. These methods include but are not limited to NMR spectroscopy, thiol-disulfide exchange experiments, fluorescence based methods such as FRET and molecular dynamic simulations. Some of these methods and their application in understanding the side-chain movements of proteins are discussed in this section.

NMR has been an elegant and widely used method of choice for residue level dynamics characterization of proteins.[131-133] The temporal changes in the side-chains span from ps to μs timescales and are stochastic in nature. This influences the correlation function of the magnetic dipoles. In other words, how the dipoles relax after they are excited is dictated by side-chain fluctuations. Spin relaxation experiments that provide information about the amplitude and temporal order of the movement of side-chains have been developed and implemented extensively to several types of nucleus (^{1}H, ^{2}H, ^{13}C and ^{15}N).[102, 131-132, 134] The rates of relaxation are affected by various factors such as quadrupolar interactions, dipole-dipole interactions and chemical shift anisotropy. The effective local magnetic field, which is susceptible to the dynamics of side-chains of a protein, gets modulated by these factors. This establishes a direct relationship between spin relaxation and internal motions in a protein. It has been perceived from several NMR studies that the side-chain dynamics of proteins are heterogeneous and tunable. For example, using NMR spin relaxation measurements of methyl side-chains, Wand and co-workers observed significant variations in the internal conformational dynamics of the calcium-binding protein, calmodulin upon binding to various target domains.[135] Interestingly, the dynamics of the methyl bearing side-chains of residues varied significantly as a function of the nature of the target domain while the backbone dynamics of the protein remained invariant. The conformational entropy arising due to the movement of side-chains in a protein was revealed to be a substantial contributor to its molecular recognition.[15, 135] In a separate study,

using high field relaxation and high-resolution relaxometry, the authors determined the amplitude and timescales of methyl side-chain dynamics of ubiquitin.[136] The observed dynamics in the ns time range could be attributed to the transition between different rotamers of the side-chains. The application of NMR to study the dynamics of side-chains has been majorly limited to small soluble proteins. However, using NMR, Kay and colleagues, in their studies on malate synthase G, a 723 residue enzyme, have been able to study the dynamics of methyl group bearing side chains successfully.[137-139] In an interesting study on dihydrofolate reductase, where a combination of NMR and X-ray crystallography was used, it was observed that alternate conformations of the proteins were sampled in the ps – ns time scale due to the backbone and side-chain motions.[140] This study also revealed that these dynamics are quenched at cryogenic temperatures in crystal structures of proteins. The aromatic side-chain dynamics can be monitored by using ^{13}C and ^{19}F NMR.[141-143] For example, using a specific biosynthetic method, ^{13}C was specifically incorporated at the δ-position of the aromatic rings of phenylalanine, tyrosine and histidine residues of E140Q mutant of the C-terminal domain of calmodulin and a μs - ms dynamics of these side-chains was observed in the relaxation dispersions NMR experiments.[142] ^{1}H, ^{13}C and ^{15}N spin relaxation studies have been implemented to study the dynamics of side-chains of polar / charged amino acids such as lysine and arginine.[144] One of the major advantages of NMR is that it needs isotopic labeling strategies, which do not disrupt structure/dynamics, unlike other measurement methods such as FRET and thiol-disulphide exchange. Although it is expensive, NMR can provide information over a large range of timescales and for many residues simultaneously. The temporal resolution of NMR is poor, and, therefore, it is almost impossible to carry out NMR experiments to study side-chain dynamics in a real time manner. Therefore, NMR measurements are typically carried out under equilibrium conditions. One of the major limitations of solution NMR is the simultaneous determination of timescales and amplitudes due to the tumbling of the protein.[145] However, in solid state NMR experiments, this limitation is avoided due to the absence of overall tumbling. In a study on the SH3 domain of α-spectrin, Skrynnikov and colleagues compared the side-chain dynamics of the protein from the solution and solid state NMR experiments. They observed that the methyl side-chain dynamics in the solid state and solution experiments were similar and rotameric jumps of amino acids in the protein were happening in a ns - μs timescale.[146] In a separate NMR study on the crystalline globular protein GB1 protein, Emsley and co-workers measured protein backbone, side-chain and solvent motions over a range of temperature and timescales. They observed that as

temperature increases, different dynamic modes get activated due to the coupling between these different motions.[147]

Fluorescence based methods such as Foster resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCS) can help in mapping conformational heterogeneity due to side-chain dynamics in a protein. If the experiments are carried out in a time-resolved manner, the structural change in the protein as a function of time can be calculated quantitatively. Single molecule FRET experiments, in which dynamics of a single protein molecule can be observed in real-time, can be used to detect conformational heterogeneity, sequence of events during a specific function, transient or metastable populations sampled, etc.[148] The timescale of energy transfer in FRET experiments is usually in ps which is much faster than the fluorescence decay rate (ns) of the donor. Nevertheless, timescales of side-chain dynamics (ns - μs) that occur slower than the donor decay rate can be detected in FRET experiments. FRET experiments have few limitations such as: a) determination of the labeling site is not trivial. b) labeling can substantially alter the structure and dynamics of the protein that cannot predict a-priori. Another major drawback of this method is that only a single distance change can be mapped in a protein at a specific time. However, by having multiple FRET pairs and generating a FRET distance map, this limitation can be avoided. For example, in a single molecule FRET study on the N-terminal domain of P53 and full length P53, the existence of multiple conformations of the protein was observed.[149] These conformations were often remained undetected in the ensemble averaged techniques. In time-resolved FRET experiments, the lifetimes of a fluorophore are measured from which the distance changes are calculated. In FCS, the correlation between the fluorescence lifetimes is measured. Using two-dimensional fluorescence lifetime correlation spectroscopy (FLCS), Tahara and co-workers have detected the presence of conformational ensembles and measured the μs timescale dynamics among the ensembles in cytochrome C in an acidic condition.[150] In a different study using FLCS on the B domain of protein A, the authors observed a highly heterogeneous nature of the native and the unfolded state of the protein. The structural inter-conversion in each ensemble was happening at a timescale of < 10 μs.[151]

Thiol-disulphide exchange experiments have also been employed to measure side-chain dynamics in multiple proteins. In this method, the intrinsic or engineered cysteine groups present in a protein are site-specifically labeled by thiol labeling reagents and the extent of thiol-disulfide exchange between cysteine side-chains and labeling dyes are determined by different spectroscopic methods such as absorbance and mass spectrometry. By monitoring the

rate of exchange reaction, this method helps in studying the structural changes and dynamics at individual side-chain level. For example, for proteins such as barstar,[152] and monellin,[153] thiol-disulfide exchange experiments have been performed to obtain site-specific dynamic information of cysteine side-chains. During the refolding study of apomyoglobin, Loh and co-workers have obtained the temporal information of cysteine side-chain movements and formation of side-chain packing during the structure gain of the protein.[154] In the case of RNAse H, Marqusee and co-workers detected some rarely populated states on the native side of the free energy barrier. Using thiol-disulfide exchange coupled to mass spectrometry, the authors identified and measured the dynamics of the side-chains that are modified upon crossing the energy barrier.[155]

Computation based methods have been widely employed to study the dynamics of proteins.[156-157] These methods individually and along with a combination of other spectroscopic techniques such as NMR have paved for a better understanding of temporal and structural order of side-chain dynamics in proteins.[95, 97, 136] Using Markov state models and molecular dynamic simulations, many groups have been able to explore the timescale of side-chain fluctuations,[158-159] orders of fluctuations at different structural sites in a protein,[96-97] and the correlations among side-chain motions[160] and their role in allosteric regulations[161] and protein-ligand interactions.[103]

1.7 Role of side-chain dynamics in protein functions

Molecular recognition is central to all protein functions. Several hypotheses have been proposed to understand mechanisms of molecular recognition in proteins which include lock and key mechanism,[162] Koshland's induced fit hypothesis[163] and recently proposed conformational selection model.[164] However, the inherently dynamic nature of proteins and growing experimental evidence[165-166] support the conformational selection model which postulates the pre-existence of multiple conformations of a protein and the binding of a ligand favors one conformation over others.[167] Following binding, the population redistribution of conformational substates happens. In comparison to the lock and key mechanism and induced fit hypothesis, which advocate that protein exists as a single stable conformation at a given condition, the conformational selection model derives its base from the energy landscape theory which postulates the presence of both low-energy and high-energy conformational substates at the native state basin.[50] The native substates vary from each other in the arrangement of different structural elements (Figure 1.5). In this context, the arrangement of

side-chains at the binding site provides the groove or the interface where the ligands can bind to the protein. For example, Allain and co-workers have observed distinct substates with different aromatic side-chains orientations at the RNA binding interface of the CUG binding protein 2. A conformational switch of these side-chains tunes the binding affinity of this protein.[168] An important question that arises here is that do proteins with similar functions have similar side-chain dynamics as in the case of homologous proteins. In an NMR study on the three PDZ domains with 30% sequence identity, Lee and co-workers have observed an organized and conserved side-chain dynamics in all the three homologous proteins which were in the ps-ns timescale.[169] These conserved dynamics of side-chains directly supported their significance in protein function.

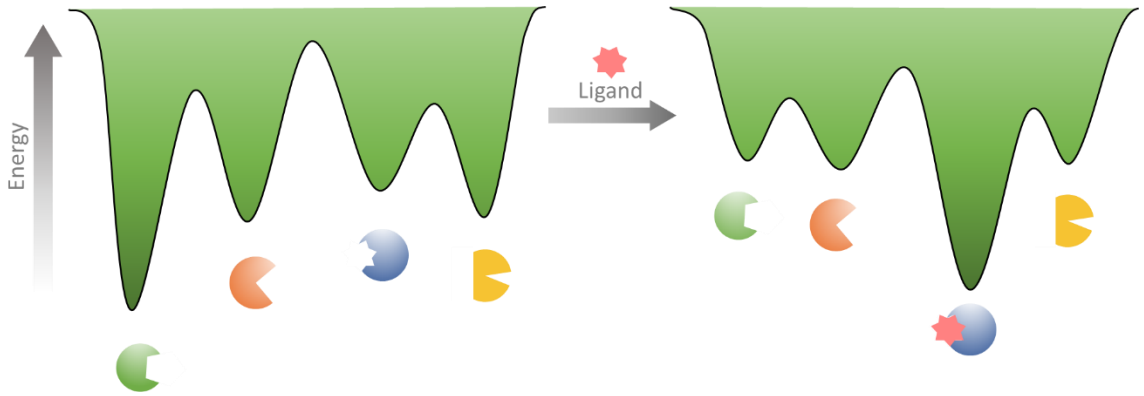


Figure 1.5. *The conformational selection model. Before the addition of ligand, multiple conformational substates pre-exist in equilibrium with each other in the native state basin of the energy landscape. Based on the structural arrangement, the ligand prefers one conformation over the others. After ligand binding, the population shift and conformational redistributions occur among the substates.*

Conformational flexibility in proteins is vital for cellular functions. The thermodynamic manifestation of flexibility is covered in the entropic effects which protein experiences.[19] Of all different structural elements of a protein, the entropic contribution due to side-chain flexibility and fluctuations is a major contributor to total conformational entropy. This conformational entropy observed due to side-chain packing plays a significant role in the free energy of ligand binding and molecular association.[15, 135] In proteins such as calmodulin[135] and catabolite activator protein,[170-171] the protein function was observed to be regulated by conformational entropy. Apart from molecular recognition, correlated movements of side-chains have a role to play in the allosteric propagations within a protein.[16, 172-174] For example,

in a study on catabolite activator protein, Kalodimos and co-workers observed that the binding of cAMP to one subunit of the proteins does not affect the other subunit.[172] However, upon sequential ligand binding, it was observed that the first subunit enhanced while the second subunit completely quenched protein dynamics. This study revealed that the conformational entropic penalty was incurred due to second ligand binding which resulted in negative cooperativity.

1.8 Conclusion

The mechanism by which a single protein performs multiple functions has not been well understood. Recent studies imply that all proteins sample several conformational substates that collectively constitute the native state ensemble. Though their distribution is determined by statistical thermodynamics, they are mostly tweaked to perform specific biological functions. Flexibility and dynamics are the two strategies adapted by nature to enable proteins to perform complex functions at reduced costs. However, the complete understanding of the mechanism of protein function requires accurate knowledge of the structure and dynamics of protein molecules before, during and after they perform a function. Of all structural modules, the side-chain packing in a protein is a substantial contributor to its thermodynamic stability and their dynamics greatly contribute to the conformational entropy of the protein. The native conformational substates sampled due to side-chain packing are recently proposed to be like DMG. Understanding the nature and thermodynamics of these DMG-like states can reveal information on the contribution of van der Waal's forces to protein stability and the role of conformational entropy in protein function. Within the broad framework of this chapter, we discussed the origin of conformational heterogeneity in the native state ensemble, the structural modules that assist the fluctuations and equilibrium between the native substates while focussing on side-chain packing, the experimental and spectroscopic techniques to characterize side-chain dynamics and their significance in cellular functions such as molecular recognition and dynamic allostery. However, many open questions still remain unexplored such as: (i) If in the native state ensemble, heterogeneous states due to side-chain packing are like DMG, then how mobile and different are the packing interactions in the core of the DMG as compared to the native state? (ii) Are DMG a subset of native state ensemble? (iii) Whether DMG-like properties can be localized only in the selective regions of the protein structure? (iv) Does heterogeneity due to side-chain packing exist only on surfaces or is it extended to hydrophobic cores? (v) What is the temporal nature of fluctuations of protein side-chains in the hydrophobic

core? We believe that in the future, more sophisticated and developed computational and experimental techniques will stimulate a detailed atomic-level understanding of side-chain dynamics of proteins and their significance in protein structure, stability and function.

1.9 References

1. Levinthal, C. Are there pathways for protein folding? *J Chim Phys* **1968,** *65*, 44-45.
2. Levinthal, C. How to fold graciously. *Mossbauer spectroscopy in biological systems* **1969,** *67*, 22-24.
3. Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's paradox. *Proc Natl Acad Sci U S A* **1992,** *89* (1), 20-22.
4. Kim, P. S.; Baldwin, R. L. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* **1990,** *59*, 631-660.
5. Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. Macmillan: 1999.
6. Ptitsyn, O. B. Stages in the mechanism of self-organization of protein molecules. *Dokl Akad Nauk SSSR* **1973,** *210* (5), 1213-1215.
7. Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985,** *24* (6), 1501-1509.
8. Wetlaufer, D. B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* **1973,** *70* (3), 697-701.
9. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **1997,** *48*, 545-600.
10. Baldwin, R. L. The nature of protein folding pathways: the classical versus the new view. *J Biomol NMR* **1995,** *5* (2), 103-109.
11. Baldwin, R. L.; Rose, G. D. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* **1999,** *24* (1), 26-33.
12. Baldwin, R. L.; Rose, G. D. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci* **1999,** *24* (2), 77-83.
13. Englander, S. W. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* **2000,** *29*, 213-238.

14. Bhatia, S.; Krishnamoorthy, G.; Dhar, D.; Udgaonkar, J. B. Observation of Continuous Contraction and a Metastable Misfolded State during the Collapse and Folding of a Small Protein. *J Mol Biol* **2019,** *431* (19), 3814-3826.

15. Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* **2010,** *6* (5), 352-358.

16. Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J Am Chem Soc* **2017,** *139* (10), 3599-3602.

17. Neumaier, S.; Kiefhaber, T. Redefining the dry molten globule state of proteins. *J Mol Biol* **2014,** *426* (13), 2520-2528.

18. Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* **2010,** *78* (13), 2725-2737.

19. Baldwin, R. L.; Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* **2013,** *23* (1), 4-10.

20. Prajapati, R. S.; Indu, S.; Varadarajan, R. Identification and thermodynamic characterization of molten globule states of periplasmic binding proteins. *Biochemistry* **2007,** *46* (36), 10339-10352.

21. Ackerman, M. S.; Shortle, D. Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry* **2002,** *41* (46), 13791-13797.

22. Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. Long-range interactions within a nonnative protein. *Science* **2002,** *295* (5560), 1719-1722.

23. Mok, K. H.; Kuhn, L. T.; Goez, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* **2007,** *447* (7140), 106-109.

24. Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. Proline cis-trans isomerization and protein folding. *Biochemistry* **2002,** *41* (50), 14637-14644.

25. Odefey, C.; Mayr, L. M.; Schmid, F. X. Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J Mol Biol* **1995,** *245* (1), 69-78.

26. Akiyama, S.; Takahashi, S.; Ishimori, K.; Morishima, I. Stepwise formation of alpha-helices during cytochrome c folding. *Nat Struct Biol* **2000,** *7* (6), 514-520.

27. Nishimura, C.; Dyson, H. J.; Wright, P. E. The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J Mol Biol* **2002,** *322* (3), 483-489.

28. Nishimura, C.; Dyson, H. J.; Wright, P. E. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol* **2006,** *355* (1), 139-156.

29. Houry, W. A.; Scheraga, H. A. Structure of a hydrophobically collapsed intermediate on the conformational folding pathway of ribonuclease A probed by hydrogen-deuterium exchange. *Biochemistry* **1996,** *35* (36), 11734-11746.

30. Pradeep, L.; Udgaonkar, J. B. Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. *J Mol Biol* **2002,** *324* (2), 331-347.

31. Pradeep, L.; Udgaonkar, J. B. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J Biol Chem* **2004,** *279* (39), 40303-40313.

32. Yang, H.; Smith, D. L. Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. *Biochemistry* **1997,** *36* (48), 14992-14999.

33. Udgaonkar, J. B.; Baldwin, R. L. Early folding intermediate of ribonuclease A. *Proc Natl Acad Sci U S A* **1990,** *87* (21), 8197-8201.

34. Sridevi, K.; Lakshmikanth, G. S.; Krishnamoorthy, G.; Udgaonkar, J. B. Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J Mol Biol* **2004,** *337* (3), 699-711.

35. Udgaonkar, J. B. Multiple routes and structural heterogeneity in protein folding. *Annu Rev Biophys* **2008,** *37*, 489-510.

36. Kim, P. S.; Baldwin, R. L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **1982,** *51*, 459-489.

37. Udgaonkar, J. B.; Baldwin, R. L. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **1988,** *335* (6192), 694-699.

38. Karplus, M.; Weaver, D. L. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* **1994,** *3* (4), 650-668.

39. Bashford, D.; Cohen, F. E.; Karplus, M.; Kuntz, I. D.; Weaver, D. L. Diffusion-collision model for the folding kinetics of myoglobin. *Proteins* **1988,** *4* (3), 211-227.

40. Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. Is burst hydrophobic collapse necessary for protein folding? *Biochemistry* **1995,** *34* (9), 3066-3076.

41. Agashe, V. R.; Shastry, M. C.; Udgaonkar, J. B. Initial hydrophobic collapse in the folding of barstar. *Nature* **1995,** *377* (6551), 754-757.

42. Fersht, A. R.; Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **2002,** *108* (4), 573-582.

43. Fersht, A. R. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* **1995,** *92* (24), 10869-10873.

44. Daggett, V.; Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem Sci* **2003,** *28* (1), 18-25.

45. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995,** *21* (3), 167-195.

46. Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nat Struct Biol* **1997,** *4* (1), 10-19.

47. Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **2000,** *25* (7), 331-339.

48. Pande, V. S.; Grosberg, A.; Tanaka, T.; Rokhsar, D. S. Pathways for protein folding: is a new view needed? *Curr Opin Struct Biol* **1998,** *8* (1), 68-79.

49. Ozkan, S. B.; Dill, K. A.; Bahar, I. Fast-folding protein kinetics, hidden intermediates, and the sequential stabilization model. *Protein Sci* **2002,** *11* (8), 1958-1970.

50. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991,** *254* (5038), 1598-1603.

51. Frauenfelder, H.; Parak, F.; Young, R. D. Conformational substates in proteins. *Annu Rev Biophys Biophys Chem* **1988,** *17*, 451-479.

52. Hegler, J. A.; Weinkam, P.; Wolynes, P. G. The spectrum of biomolecular states and motions. *HFSP J* **2008,** *2* (6), 307-313.

53. Cooper, A. Thermodynamic fluctuations in protein molecules. *Proc Natl Acad Sci U S A* **1976,** *73* (8), 2740-2741.

54. Cooper, A. Protein fluctuations and the thermodynamic uncertainty principle. *Prog Biophys Mol Biol* **1984,** *44* (3), 181-214.

55. Iwao, O.; Hideki, T.; Peter, G. W. Large local energy fluctuations in water. II. Cooperative motions and fluctuations. *The Journal of Chemical Physics* **1988,** *89* (9), 5852-5860.

56. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007,** *450* (7172), 964-972.

57. Lubchenko, V.; Wolynes, P. G.; Frauenfelder, H. Mosaic energy landscapes of liquids and the control of protein conformational dynamics by glass-forming solvents. *J Phys Chem B* **2005,** *109* (15), 7488-7499.

58. Angell, C. A. The old problems of glass and the glass transition, and the many new twists. *Proc Natl Acad Sci U S A* **1995,** *92* (15), 6675-6682.

59. Young, R. D.; Fenimore, P. W. Coupling of protein and environment fluctuations. *Biochim Biophys Acta* **2011,** *1814* (8), 916-921.

60. Fenimore, P. W.; Frauenfelder, H.; Magazu, S.; McMahon, B. H.; Mezei, F.; Migliardo, F.; Young, R. D.; Stroe, I. Concepts and problems in protein dynamics. *Chem Phys* **2013,** *424*, 2-6.

61. Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroe, I. R.; Swenson, J.; Young, R. D. A unified model of protein dynamics. *Proc Natl Acad Sci U S A* **2009,** *106* (13), 5129-5134.

62. Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Parak, F. G. Slaving: solvent fluctuations dominate protein dynamics and functions. *Proc Natl Acad Sci U S A* **2002,** *99* (25), 16047-16051.

63. Chen, G.; Fenimore, P.; Frauenfelder, H.; Mezei, F.; Swenson, J.; Young, R. Protein fluctuations explored by inelastic neutron scattering and dielectric relaxation spectroscopy. *Philos Mag* **2008,** *88* (33-35), 3877-3883.

64. Richards, F. M. The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* **1974,** *82* (1), 1-14.

65. Chothia, C. Structural invariants in protein folding. *Nature* **1975,** *254* (5498), 304-308.

66. Richards, F. M. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* **1977,** *6*, 151-176.

67. Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem Rev* **2016,** *116* (11), 6516-6551.

68. Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem Rev* **2016,** *116* (11), 6391-6423.

69. Skliros, A.; Zimmermann, M. T.; Chakraborty, D.; Saraswathi, S.; Katebi, A. R.; Leelananda, S. P.; Kloczkowski, A.; Jernigan, R. L. The importance of slow motions for protein functional loops. *Phys Biol* **2012,** *9* (1), 014001.

70. Juszczak, L. J.; Zhang, Z. Y.; Wu, L.; Gottfried, D. S.; Eads, D. D. Rapid loop dynamics of Yersinia protein tyrosine phosphatases. *Biochemistry* **1997,** *36* (8), 2227-2236.

71. Hu, X.; Stebbins, C. E. Dynamics of the WPD loop of the Yersinia protein tyrosine phosphatase. *Biophys J* **2006,** *91* (3), 948-956.

72. Qin, H.; Lim, L.; Song, J. Protein dynamics at Eph receptor-ligand interfaces as revealed by crystallography, NMR and MD simulations. *BMC Biophys* **2012,** *5*, 2.

73. Venkitakrishnan, R. P.; Zaborowski, E.; McElheny, D.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry* **2004,** *43* (51), 16046-16055.

74. Gokhale, R. S.; Khosla, C. Role of linkers in communication between protein modules. *Curr Opin Chem Biol* **2000,** *4* (1), 22-27.

75. Wriggers, W.; Chakravarty, S.; Jennings, P. A. Control of protein functional dynamics by peptide linkers. *Biopolymers* **2005,** *80* (6), 736-746.

76. Reddy Chichili, V. P.; Kumar, V.; Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci* **2013,** *22* (2), 153-167.

77. Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci* **2002,** *11* (2), 184-197.

78. McHaourab, H. S.; Oh, K. J.; Fang, C. J.; Hubbell, W. L. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry* **1997,** *36* (2), 307-316.

79. Yirdaw, R. B.; McHaourab, H. S. Direct observation of T4 lysozyme hinge-bending motion by fluorescence correlation spectroscopy. *Biophys J* **2012,** *103* (7), 1525-1536.

80. Wen, B.; Peng, J.; Zuo, X.; Gong, Q.; Zhang, Z. Characterization of protein flexibility using small-angle x-ray scattering and amplified collective motion simulations. *Biophys J* **2014,** *107* (4), 956-964.

81. de Groot, B. L.; Hayward, S.; van Aalten, D. M.; Amadei, A.; Berendsen, H. J. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins* **1998,** *31* (2), 116-127.

82. Gerstein, M.; Anderson, B. F.; Norris, G. E.; Baker, E. N.; Lesk, A. M.; Chothia, C. Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces. *J Mol Biol* **1993,** *234* (2), 357-372.

83. Mizutani, K.; Toyoda, M.; Mikami, B. X-ray structures of transferrins and related proteins. *Biochim Biophys Acta* **2012,** *1820* (3), 203-211.

84. Baker, E. N.; Baker, H. M. A structural framework for understanding the multifunctional character of lactoferrin. *Biochimie* **2009,** *91* (1), 3-10.

85. Liu, Z.; Gong, Z.; Jiang, W. X.; Yang, J.; Zhu, W. K.; Guo, D. C.; Zhang, W. P.; Liu, M. L.; Tang, C. Lys63-linked ubiquitin chain adopts multiple conformational states for specific target recognition. *Elife* **2015,** *4*.

86. Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys Rev Lett* **1998,** *80* (12), 2733-2736.

87. Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **2006,** *313* (5793), 1638-1642.

88. Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007,** *450* (7171), 838-844.

89. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007,** *450* (7171), 913-916.

90. Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **2001,** *291* (5512), 2429-2433.

91. Buck, M.; Rosen, M. K. Structural biology. Flipping a switch. *Science* **2001,** *291* (5512), 2329-2330.

92. Fenwick, R. B.; Esteban-Martin, S.; Richter, B.; Lee, D.; Walter, K. F.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J Am Chem Soc* **2011,** *133* (27), 10336-10339.

93. Smith, C. A.; Ban, D.; Pratihar, S.; Giller, K.; Paulat, M.; Becker, S.; Griesinger, C.; Lee, D.; de Groot, B. L. Allosteric switch regulates protein-protein binding through collective motion. *Proc Natl Acad Sci U S A* **2016,** *113* (12), 3269-3274.

94. Sanchez-Medina, C.; Sekhar, A.; Vallurupalli, P.; Cerminara, M.; Munoz, V.; Kay, L. E. Probing the free energy landscape of the fast-folding gpW protein by relaxation dispersion NMR. *J Am Chem Soc* **2014,** *136* (20), 7444-7451.

95. Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **2005,** *433* (7022), 128-132.

96. Bowman, G. R.; Geissler, P. L. Extensive conformational heterogeneity within protein cores. *J Phys Chem B* **2014,** *118* (24), 6417-6423.

97. DuBay, K. H.; Bowman, G. R.; Geissler, P. L. Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Acc Chem Res* **2015,** *48* (4), 1098-1105.

98. Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. Protein states and proteinquakes. *Proc Natl Acad Sci U S A* **1985,** *82* (15), 5000-5004.

99. Schrödinger, E. *What is life? The physical aspect of the living cell*. The University Press; The Macmillan company: Cambridge Eng. New York, 1962

100. Richards, F. M.; Lim, W. A. An analysis of packing in the protein folding problem. *Q Rev Biophys* **1993,** *26* (4), 423-498.

101. Zhou, Y.; Vitkup, D.; Karplus, M. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J Mol Biol* **1999,** *285* (4), 1371-1375.

102. Igumenova, T. I.; Frederick, K. K.; Wand, A. J. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem Rev* **2006,** *106* (5), 1672-1699.

103. DuBay, K. H.; Geissler, P. L. Calculation of proteins' total side-chain torsional entropy and its influence on protein-ligand interactions. *J Mol Biol* **2009,** *391* (2), 484-497.

104. Shakhnovich, E. I.; Finkelstein, A. V. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* **1989,** *28* (10), 1667-1680.

105. Finkelstein, A. V.; Shakhnovich, E. I. Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* **1989,** *28* (10), 1681-1694.

106. Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* **1995,** *375* (6531), 513-515.

107. Hoeltzli, S. D.; Frieden, C. Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-^{19}F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc Natl Acad Sci U S A* **1995,** *92* (20), 9318-9322.

108. Jha, S. K.; Udgaonkar, J. B. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* **2009,** *106* (30), 12289-12294.

109. Jha, S. K.; Marqusee, S. Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* **2014,** *111* (13), 4856-4861.

110. Sarkar, S. S.; Udgaonkar, J. B.; Krishnamoorthy, G. Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* **2013,** *105* (10), 2392-2402.

111. Rami, B. R.; Udgaonkar, J. B. Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* **2002,** *41* (6), 1710-1716.

112. Acharya, N.; Mishra, P.; Jha, S. K. A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Phys Chem Chem Phys* **2017,** *19* (44), 30207-30216.

113. Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* **2016,** *7* (1), 173-179.

114. Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* **2012,** *134* (20), 8543-8550.

115. Reiner, A.; Henklein, P.; Kiefhaber, T. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* **2010,** *107* (11), 4955-4960.

116. de Oliveira, G. A.; Silva, J. L. A hypothesis to reconcile the physical and chemical unfolding of proteins. *Proc Natl Acad Sci U S A* **2015,** *112* (21), E2775-2784.

117. Creamer, T. P. Side- chain conformational entropy in protein unfolded states. *Proteins* **2000,** *40* (3), 443-450.

118. Makhatadze, G. I.; Privalov, P. L. On the entropy of protein folding. *Protein Sci* **1996,** *5* (3), 507-510.

119. Némethy, G.; Leach, S.; Scheraga, H. A. The influence of amino acid side chains on the free energy of helix-coil transitions1. *J Phys Chem* **1966,** *70* (4), 998-1004.

120. Doig, A. J.; Sternberg, M. J. Side- chain conformational entropy in protein folding. *Protein Sci* **1995,** *4* (11), 2247-2251.

121. Sandberg, W. S.; Terwilliger, T. C. Influence of interior packing and hydrophobicity on the stability of a protein. *Science* **1989,** *245* (4913), 54-57.

122. Kellis Jr, J. T.; Nyberg, K.; Fersht, A. R. Energetics of complementary side chain packing in a protein hydrophobic core. *Biochemistry* **1989,** *28* (11), 4914-4922.

123. Behe, M. J.; Lattman, E. E.; Rose, G. D. The protein-folding problem: the native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci U S A* **1991,** *88* (10), 4195-4199.

124. Bhattacharyya, S.; Varadarajan, R. Packing in molten globules and native states. *Curr Opin Struct Biol* **2013,** *23* (1), 11-21.

125. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **1959,** *14*, 1-63.

126. Tanford, C. Protein denaturation. *Adv Protein Chem* **1968,** *23*, 121-282.

127. Privalov, P. L. Stability of proteins small globular proteins. *Adv Protein Chem* **1979,** *33*, 167-241.

128. Privalov, P. L.; Gill, S. J. Stability of protein structure and hydrophobic interaction. *Adv Protein Chem* **1988,** *39*, 191-234.

129. Kumar, S.; Nussinov, R. Fluctuations in ion pairs and their stabilities in proteins. *Proteins* **2001,** *43* (4), 433-454.

130. Alexov, E. Role of the protein side- chain fluctuations on the strength of pair-wise electrostatic interactions: Comparing experimental with computed pKas. *Proteins* **2003,** *50* (1), 94-103.

131. Kay, L. E. NMR studies of protein structure and dynamics. *J Magn Reson* **2005,** *173* (2), 193-207.

132. Stetz, M. A.; Caro, J. A.; Kotaru, S.; Yao, X.; Marques, B. S.; Valentine, K. G.; Wand, A. J. Characterization of Internal Protein Dynamics and Conformational Entropy by NMR Relaxation. *Methods Enzymol* **2019,** *615*, 237-284.

133. Boehr, D. D.; Dyson, H. J.; Wright, P. E. An NMR perspective on enzyme dynamics. *Chem Rev* **2006,** *106* (8), 3055-3079.

134. Mittermaier, A.; Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **2006,** *312* (5771), 224-228.

135. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **2007,** *448* (7151), 325-329.

136. Cousin, S. F.; Kadeřávek, P.; Bolik-Coulon, N.; Gu, Y.; Charlier, C.; Carlier, L.; Bruschweiler-Li, L.; Marquardsen, T.; Tyburn, J.-M.; Brüschweiler, R. Time-resolved protein side-chain motions unraveled by high-resolution relaxometry and molecular dynamics simulations. *J Am Chem Soc* **2018,** *140* (41), 13456-13465.

137. Korzhnev, D. M.; Kloiber, K.; Kanelis, V.; Tugarinov, V.; Kay, L. E. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* **2004,** *126* (12), 3964-3973.

138. Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. Cross-correlated relaxation enhanced 1H− 13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **2003,** *125* (34), 10420-10428.

139. Tugarinov, V.; Ollerenshaw, J. E.; Kay, L. E. Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: an application to an 82-kDa enzyme. *J Am Chem Soc* **2005,** *127* (22), 8214-8225.

140. Fenwick, R. B.; van den Bedem, H.; Fraser, J. S.; Wright, P. E. Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc Natl Acad Sci U S A* **2014,** *111* (4), E445-E454.

141. Kasinath, V.; Valentine, K. G.; Wand, A. J. A 13C labeling strategy reveals a range of aromatic side chain motion in calmodulin. *J Am Chem Soc* **2013,** *135* (26), 9560-9563.

142. Teilum, K.; Brath, U.; Lundström, P.; Akke, M. Biosynthetic 13C labeling of aromatic side chains in proteins for NMR relaxation measurements. *J Am Chem Soc* **2006,** *128* (8), 2506-2507.

143. Di Pietrantonio, C.; Pandey, A.; Gould, J.; Hasabnis, A.; Prosser, R. S. Understanding protein function through an ensemble description: characterization of functional states by 19F NMR. *Methods Enzymol* **2019,** *615*, 103-130.

144. Nguyen, D.; Chen, C.; Pettitt, B. M.; Iwahara, J. NMR Methods for characterizing the basic side chains of proteins: electrostatic interactions, hydrogen bonds, and conformational dynamics. *Methods Enzymol* **2019,** *615*, 285-332.

145. Lewandowski, J. R. Advances in solid-state relaxation methodology for probing site-specific protein dynamics. *Accounts of chemical research* **2013,** *46* (9), 2018-2027.

146. Agarwal, V.; Xue, Y.; Reif, B.; Skrynnikov, N. R. Protein side-chain dynamics as observed by solution-and solid-state NMR spectroscopy: a similarity revealed. *Journal of the American Chemical Society* **2008,** *130* (49), 16611-16621.

147. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L. Direct observation of hierarchical protein dynamics. *Science* **2015,** *348* (6234), 578-581.

148. Schuler, B.; Hofmann, H. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr Opin Struct Biol* **2013,** *23* (1), 36-47.

149. Huang, F.; Rajagopalan, S.; Settanni, G.; Marsh, R. J.; Armoogum, D. A.; Nicolaou, N.; Bain, A. J.; Lerner, E.; Haas, E.; Ying, L. Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* **2009,** *106* (49), 20758-20763.

150. Otosu, T.; Ishii, K.; Tahara, T. Microsecond protein dynamics observed at the single-molecule level. *Nat Commun* **2015,** *6* (1), 1-9.

151. Otosu, T.; Ishii, K.; Oikawa, H.; Arai, M.; Takahashi, S.; Tahara, T. Highly heterogeneous nature of the native and unfolded states of the B domain of protein a revealed by two-dimensional fluorescence lifetime correlation spectroscopy. *J Phys Chem B* **2017,** *121* (22), 5463-5473.

152. Jha, S. K.; Udgaonkar, J. B. Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* **2007,** *282* (52), 37479-37491.

153. Jha, S. K.; Dasgupta, A.; Malhotra, P.; Udgaonkar, J. B. Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry. *Biochemistry* **2011,** *50* (15), 3062-3074.

154. Ha, J.-H.; Loh, S. N. Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat Struct Biol* **1998,** *5* (8), 730-737.

155. Bernstein, R.; Schmidt, K. L.; Harbury, P. B.; Marqusee, S. Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proc Natl Acad Sci U S A* **2011,** *108* (26), 10532-10537.

156. Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **2002,** *9* (9), 646-652.

157. MacKerell Jr, A. D. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* **2004,** *25* (13), 1584-1604.

158. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010,** *330* (6002), 341-346.

159. Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. *Proteins* **2014,** *82* (2), 195-215.

160. DuBay, K. H.; Bothma, J. P.; Geissler, P. L. Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLoS Comput Biol* **2011,** *7* (9), e1002168.

161. Wu, S.; Jun Lee, C.; Pedersen, L. G. Analysis on long-range residue–residue communication using molecular dynamics. *Proteins* **2014,** *82* (11), 2896-2901.

162. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **1894,** *27* (3), 2985-2993.

163. Koshland Jr, D. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci U S A* **1958,** *44* (2), 98.

164. Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **2000,** *9* (1), 10-19.

165. Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A* **2005,** *102* (52), 18908-18913.

166. Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; De Groot, B. L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *science* **2008,** *320* (5882), 1471-1475.

167. Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **2009,** *5* (11), 789-796.

168. Diarra Dit Konte, N.; Krepl, M.; Damberger, F. F.; Ripin, N.; Duss, O.; Sponer, J.; Allain, F. H. Aromatic side-chain conformational switch on the surface of the RNA Recognition Motif enables RNA discrimination. *Nat Commun* **2017,** *8* (1), 654.

169. Law, A. B.; Fuentes, E. J.; Lee, A. L. Conservation of side-chain dynamics within a protein family. *J Am Chem Soc* **2009,** *131* (18), 6322-6323.

170. Tzeng, S.-R.; Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **2009,** *462* (7271), 368-372.

171. Tzeng, S.-R.; Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **2012,** *488* (7410), 236-240.

172. Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. Dynamically driven protein allostery. *Nat Struct Mol Biol* **2006,** *13* (9), 831-838.

173. Capdevila, D. A.; Braymer, J. J.; Edmonds, K. A.; Wu, H.; Giedroc, D. P. Entropy redistribution controls allostery in a metalloregulatory protein. *Proc Natl Acad Sci U S A* **2017,** *114* (17), 4424-4429.

174. Capdevila, D. A.; Edmonds, K. A.; Campanello, G. C.; Wu, H.; Gonzalez-Gutierrez, G.; Giedroc, D. P. Functional role of solvent entropy and conformational entropy of metal binding in a dynamically driven allosteric system. *J Am Chem Soc* **2018,** *140* (29), 9108-9119.

Chapter 2.

An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multi-Domain Protein

Reproduced with the permission of Mishra, P., and Jha, S. K. ***(2017)*** *An alternatively packed dry molten globule-like intermediate in the native state ensemble of a multidomain protein.* ***J. Phys. Chem. B****, 121, 40, 9336–9347.*

2.1 Introduction

The extent of conformational heterogeneity in the side-chain packing in the native (N) state ensemble of proteins is poorly understood.[1-12] Crystal structures of proteins project a time-averaged snapshot of the N state and show that the hydrophobic side-chain residues are packed tightly in the core and the packing densities are close to crystalline amino acids and solids.[1-4] In contrast, a body of theoretical work[5, 7, 11-13] suggests that the interior of folded proteins is dynamic in nature. It has been postulated that the low and the high energy conformations of the proteins with alternative side-chain packing arrangements exist at equilibrium in the folded protein ensemble and their populations follow Boltzmann probability distribution.[11-12, 14] Fluctuations between these different N-like forms could play an important role in many biological functions like ligand-binding,[15] catalysis[14] and signaling.[16] However, a high-resolution experimental characterization of these alternatively packed N-like conformations due to side-chain flexibility in the N state ensemble remains a significant challenge.

It has been proposed recently that the N state of proteins could be an ensemble of loosely packed dry molten globular forms.[10, 17-19] Dry molten globules (DMG) are hypothesized to be expanded forms of the N state which retain N-like global secondary and tertiary structure but have loose and molten side-chain packing and dehydrated cores. It has also been suggested using phase diagrams that during unfolding, proteins possibly sample an ensemble of DMG-like states near the N state basin and unfold via a DMG-like transition state.[20-21] Experimental data on DMG-like states are, however, very limited.[17, 22-27] Structural and thermodynamic characterization of DMG-like intermediate (I) states can provide important insights on the relative contributions of the van der Waals (vdW) interactions and the hydrophobic effect in protein stability.[17] The conformational entropy of DMG-like states is higher compared to the N states, which could also enable them to undergo dynamic volume fluctuations in the N states required for allosteric functions.[18, 28] However, it is not clear how mobile and different are the side-chain packing interactions in the core of the DMG compared to the N state.[10, 18] It also remains to be understood whether DMG-like properties can be localized only in the selective regions of the protein structure. DMG have mainly been observed as transient kinetic intermediates during the unfolding of few proteins,[17, 22-25, 27] and it has been difficult to identify them as a subset of the N state ensemble under equilibrium conditions.[26, 29-30] In particular, the detection of the rotation or movement of a protein side-chain residue in a member of the N

state ensemble, which would constitute the most direct evidence in the support of heterogeneity of side-chain packing in the N state ensemble, has remained challenging till date.

In this study, we have site-specifically probed the energetics and structural composition of the N state ensemble of a multi-domain protein, human serum albumin (HSA), under two different N-like environmental conditions, i.e, at pH 5 and pH 7, using tools of thermodynamics and a battery of site-specific and high-resolution spectroscopic probes, including fluorescence resonance energy transfer (FRET), dynamic fluorescence quenching, red-edge excitation shift (REES) and near- and far-UV circular dichroism (CD). The application of multiple spectroscopic probes allowed us to dissect the conformational heterogeneity of the N state ensemble in terms of distinct structural events characterized by changes in protein expansion, vdW packing, core hydration, solvation dynamics of the hydrophobic core and secondary and tertiary structure. We show that at pH 5, but not at pH 7, the N state ensemble is in equilibrium with an expanded DMG-like I state which has similar global secondary and tertiary structure as the N state, but has a dry and molten core with an alternative inter-domain side-chain packing arrangement. We use these observations to estimate the energetic contribution of inter-domain side-chain packing in protein stability.

2.2 Materials and Methods

Spectroscopic methods and instruments

All the absorption and fluorescence emission spectra were measured using a Perkin Elmer lambda 650 UV/Vis spectrophotometer and a Perkin Elmer fluorescence spectrometer LS 55, respectively, using a cuvette of path length 1 cm. All the CD studies were done on Jasco J-815 CD spectrometer. For far- and near-UV CD, the path lengths of the quartz cells used were 1 mm and 1 cm, respectively. Background signal corrections were done for all the fluorescence and CD spectra by subtracting the buffer signals. Abbe refractometer from Rajdhani Scientific Instruments Co. (Model: RSR-2) was used to measure the refractive index of solutions wherever required.

Chemicals, reagents, buffers, and experimental conditions

HSA (99% pure, fatty acid and globulin free) and urea (ultrapure grade) were purchased from Alfa Aesar. 5-((((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was obtained from Life Technologies. Guanidine hydrochloride (GdmCl),

acrylamide and all other chemicals of highest purity grade were purchased from Sigma. All the chemicals were used as received without any further purification. HSA concentration was determined by measuring the absorbance at 280 nm and using a molar extinction coefficient[31] of 36500 $M^{-1}cm^{-1}$.

For all the experiments, 20 mM phosphate and 20 mM acetate buffers were used as pH 7 and pH 5 native buffers, respectively. The unfolding buffers were composed of 9 M urea and 20 mM phosphate/acetate for the respective pH. For the I state, the buffer composition was 1.6 M urea and 20 mM acetate at pH 5. The urea and GdmCl concentrations were determined from refractive index measurements.[32] For pH titrations, the universal buffer was composed of 20 mM acetate, 20 mM phosphate and 20 mM borate and the pH was adjusted with HCl/NaOH.

Preparation of 1, 5-IAEDANS labeled HSA

We used the standard protocol to label HSA with 1,5-IAEDANS at the sole free cysteine, C34, as reported earlier.[30, 33] In brief, the calculated volume of GdmCl unfolded HSA was added to a 20-fold molar excess of 1,5-IAEDANS in the unfolding buffer (6 M GdmCl and 20 mM Tris at pH 8). The mixture was then continuously stirred in dark at the room temperature for 4 hrs. After the completion of the labeling reaction, the reaction mixture was diluted 10-fold with the refolding buffer (20 mM phosphate at pH 7) and kept overnight at 4 °C. It was then concentrated to 2.5 mL using a centrifugal concentrator (GE healthcare) and passed through a PD-10 desalting column (GE healthcare) to separate the labeled HSA from free dye and GdmCl. The labeled protein concentration was determined by measuring the absorbance at 337 nm and using the molar extinction coefficient (ε_{337})[34-35] of 4500 $M^{-1}cm^{-1}$. The total protein concentration was determined by measuring the absorbance at 280 nm and using a molar extinction coefficient (ε_{280})[31] of 36500 $M^{-1}cm^{-1}$. It should be noted that 1,5-IAEDANS has a molar extinction coefficient of around 900-1100 M^{-1} cm^{-1} at 280 nm[36-37], which is 35-40 fold less compared to that of the intrinsic ε_{280} of the protein (ε_{280} = 36500 $M^{-1}cm^{-1}$). Hence, the contribution of 1,5-IAEDANS to the total absorbance at 280 nm will be only 2.5-3%, if the protein is completely labeled with 1,5-IAEDANS. For example, for a typical absorbance of 0.2 at 280 nm, at which all the experiments have been performed, the contribution of 1,5-IAEDANS for the fully labeled protein will be 0.005, which is negligible. The percentage of labeling was calculated by dividing the labeled protein concentration by the total protein concentration and multiplying the ratio by 100. The protein was observed to be > 95% labeled. It should be noted that our assumption of the negligible contribution of the

absorbance of 1,5-IAEDANS at 280 nm gave the lower limit on the percentage of labeling and the actual percentage will be slightly (2-3%) higher.

pH-induced equilibrium unfolding transition monitored by fluorescence and CD

For the fluorescence monitored pH titrations, the protein concentration in each sample was 4-5 µM. The samples were kept overnight at different pH at the room temperature. The fluorescence spectra for the protein samples were acquired by exciting the sole tryptophan (W214) at 295 nm and collecting the emission from 310 nm to 420 nm. Nominal slit widths for excitation and emission were used. The scan speed was 100 nm/min. The far- and near-UV CD spectra were monitored in the wavelength range of 190-250 nm and 250-310 nm, respectively. The protein concentration for the far-UV CD experiment was 4-5 µM, whereas, for the near-UV CD experiment, it was ~20 µM. For each CD spectrum, the data pitch, data integration time, bandwidth and scan speed were 1 nm, 1 s, 2 nm and 100 nm/min, respectively. Three scans were averaged for each CD spectrum. The mean residual ellipticity (MRE) values were calculated from the CD signal (MRE = CD/(10 × no of peptide bonds× [protein] × path length of the cuvette)).

Analysis of the pH titration

During the pH-induced unfolding transition,[30] HSA transforms from the fast-migrating (F) form to the N form between pH 3.5 to pH 4.8. Between pH 4.8 to pH 8.5, the protein remains in the N form, followed by a transition to the basic (B) form beyond pH 8.5.

The far-UV CD monitored pH titration was analyzed using a model in which the structural transition was coupled to a single deprotonation step because the change in CD signal with respect to pH was observed to occur in a single sigmoidal manner. The data were fitted to equation 2.1:[30]

$$Y_{obs} = \frac{Y_F + Y_N 10^{(pH-pH_{m1})}}{1 + 10^{(pH-pH_{m1})}} \qquad (eq\ 2.1)$$

where, Y_{obs} is the far-UV CD signal at 222 nm at a particular pH; Y_F and Y_N are the signals of the F and the N forms, respectively, and pH_{m1} is the midpoint of the observed titration.

Since the change in fluorescence signal with respect to pH occurred in two sigmoidal steps, the data were fitted to equation 2.2. This equation is based on a model in which the

structural transitions from F to N form and from N to B form are each coupled to one deprotonation step.[30]

$$Y_{obs} = \frac{Y_F + Y_N 10^{(pH-pH_{m1})}}{1 + 10^{(pH-pH_{m1})}} + \frac{Y_B + Y_N 10^{(pH_{m2}-pH)}}{1 + 10^{(pH_{m2}-pH)}} \qquad (eq\ 2.2)$$

where, Y_{obs} is the fluorescence signal at 340 nm at a particular pH value; Y_F, Y_N, and Y_B , are the signals of the F, the N and B forms, respectively; and pH_{m1} and pH_{m2}, respectively, are the midpoints of the F to N and N to B transitions.

Urea-induced equilibrium unfolding transitions monitored by fluorescence and far-UV CD

For the urea-induced equilibrium unfolding transitions at pH 7 and pH 5, the protein samples (4-5 µM) were incubated overnight at different concentrations of urea at the room temperature. The equilibrium unfolding was monitored by the change in fluorescence of W214 at 328 nm upon excitation at 295 nm. For far-UV CD, it was monitored by the change in CD signal at 222 nm.

Analysis of the two-state equilibrium unfolding transitions at pH 7

For pH 7, the fluorescence and the far-UV CD monitored equilibrium unfolding transitions were analyzed using a two-state, N⇌U model. The data were fitted to equation 2.3:[38]

$$y_{obs} = \frac{y_N + y_U\, e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 2.3)$$

where, y_{obs} is the observed fluorescence/CD signal; y_N and y_U are the signals of the N and the unfolded (U) states, respectively; ΔG_{NU} is the free energy of unfolding of N⇌U transition. It is assumed that ΔG_{NU} has a linear dependence on denaturant concentration ([D]) and is given by the following equation:

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \qquad (eq\ 2.4)$$

where, $\Delta G_{NU}^{H_2O}$ and m_{NU} are the standard free energy at 0 M urea and slope of the N⇌U transition, respectively.

The fractions of the native (f_N) and the unfolded (f_U) states at any particular denaturant concentration at pH 7 were determined using equations 2.5 and 2.6, respectively.

$$f_N = \frac{1}{1 + e\frac{-\Delta G_{NU}}{RT}} \qquad (eq\ 2.5)$$

$$f_U = \frac{e\frac{-\Delta G_{NU}}{RT}}{1 + e\frac{-\Delta G_{NU}}{RT}} \qquad (eq\ 2.6)$$

Analysis of the three-state equilibrium unfolding transitions at pH 5

For pH 5, the fluorescence and the far-UV CD monitored equilibrium unfolding transitions were analyzed using a three-state, N⇌I⇌U model. For the global fit, all the data were fitted to equation 2.7:[39-40]

$$y_{obs} = \frac{y_N + (y_N + (y_U - y_N)z_I)e\frac{-\Delta G_{NI}}{RT} + y_U\, e\frac{-\Delta G_{NU}}{RT}}{1 + e\frac{-\Delta G_{NI}}{RT} + e\frac{-\Delta G_{NU}}{RT}} \qquad (eq\ 2.7)$$

where, y_N and y_U are the signals of the N and the U states, respectively; ΔG_{NI} and ΔG_{NU} are the free energies of the unfolding of N⇌I and N⇌U transitions, respectively. z_I, which measures the optical signal of I (y_I), relative to y_N and y_U, is given by:

$$z_I = \frac{y_I - y_N}{y_U - y_N} \qquad (eq\ 2.8)$$

It is assumed that all the free energies have a linear dependence on $[D]$ and are given by the following equations:

$$\Delta G_{NI} = \Delta G_{NI}^{H_2O} + m_{NI}[D] \qquad (eq\ 2.9)$$

$$\Delta G_{IU} = \Delta G_{IU}^{H_2O} + m_{IU}[D] \qquad (eq\ 2.10)$$

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \qquad (eq\ 2.11)$$

where, $\Delta G_{NI}^{H_2O}$, $\Delta G_{IU}^{H_2O}$ and $\Delta G_{NU}^{H_2O}$ are the standard free energies of N⇌I, I⇌U and N⇌U transitions, respectively, and m_{NI}, m_{IU} and m_{NU} are the respective slopes associated with the above mentioned transitions. ΔG_{NU} and m_{NU} can also be expressed as : $\Delta G_{NU} = \Delta G_{NI} + \Delta G_{IU}$ and $m_{NU} = m_{NI} + m_{IU}$, respectively.

The fractional change in signal, f_{app} was determined from y_{obs} (as measured in equation 2.7) by:[40]

$$f_{app} = \frac{y_{obs} - (y_N + m_N[D])}{(y_U + m_U[D]) - (y_N + m_N[D])} \qquad (eq\ 2.12)$$

In the above equation, y_N and y_U are assumed to have a linear dependence on $[D]$, that are given by the slopes m_N and m_U, respectively. The data obtained upon the transformation using equation 2.12 were fitted to equation 2.13.

$$f_{app} = \frac{z_I \times e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 2.13)$$

The fractions of the native (f_N), the intermediate (f_I) and the unfolded (f_U) states at pH 5, as a function of the $[D]$ were calculated using the following equations:

$$f_N = \frac{1}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 2.14)$$

$$f_I = \frac{e^{\frac{-\Delta G_{NI}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 2.15)$$

$$f_U = \frac{e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 2.16)$$

All the thermodynamic parameters derived from the global two-state fit for the data at pH 7 and three-state fit for the data at pH 5 are mentioned in Table 2.1.

FRET between W214 and C34-IAEDANS and determination of the Forster's distance

We determined the inter-molecular distance between W214 and C34-IAEDANS using the FRET methodology as described earlier.[30] For FRET measurements, all the fluorescence spectra of the labeled and unlabeled protein in the N, the I and the U states at pH 5 were collected in an identical manner with the excitation wavelength of 295 nm.

The efficiency of energy transfer, E, was determined according to equation 2.17:[41]

$$E = 1 - \frac{F_{DA}}{F_D} \qquad (eq\ 2.17)$$

where, F_{DA} and F_D are the donor fluorescence intensities in the presence and the absence of the acceptor, respectively.

The donor-acceptor distance, *R*, is related to the efficiency of energy transfer, *E*, by:[41]

$$R = R_0[\frac{1-E}{E}]^{\frac{1}{6}} \qquad (eq\ 2.18)$$

where, R_0 is the Forster's distance at which the energy transfer is 50%. The R_0 was determined by equation 2.19.[25, 30, 41]

$$R_0 = 0.211[Q_D J \kappa^2 n^{-4}]^{\frac{1}{6}} \qquad (eq\ 2.19)$$

In the above equation, Q_D is the quantum yield of the donor fluorescence, *J* is the overlap integral, κ^2 is the orientation factor and *n* is the refractive index of the medium.

Q_D at pH 7 was taken as 0.31.[30] Since Q_D is directly proportional to the area under the fluorescence emission spectrum,[41] we used the value for the Q_D at pH 7 and the ratio of the area under the fluorescence emission spectrum at pH 5 to the area under the fluorescence emission spectrum at pH 7, and determined the value of Q_D at pH 5 to be 0.30. Similarly, the value of Q_D for the I state was determined to be 0.27.

The overlap integral, *J*, expresses the spectral overlap between the donor emission and the acceptor absorption, and is given by:[25, 41]

$$J = \frac{\int F(\lambda)\varepsilon(\lambda)\lambda^4 d\lambda}{\int F(\lambda) d\lambda} \qquad (eq\ 2.20)$$

where, $F(\lambda)$ is the fluorescence emission spectra of W214 and $\varepsilon(\lambda)$ is the absorption spectra of C34-IAEDANS. The refractive index, *n*, of the medium at pH 5 was determined to be 1.3315 ± 0.0005, whereas, for the I state, it was 1.3451 ± 0.0005. The value of the orientation factor, κ^2, was taken to be 2/3, assuming that the donor and the acceptor are randomly oriented with respect to each other. The values of R_0, *J*, and other FRET parameters are listed in Table 2.2.

Acrylamide quenching experiments

The solvent accessibility assay was performed by using acrylamide, a neutral quencher of W214 fluorescence. All the protein samples under different conditions were incubated

overnight at the room temperature. Calculated volumes from a freshly prepared 2 M acrylamide stock solution were added to the protein samples just before collecting the spectra. The protein was excited at 295 nm and the emission spectra were collected from 310-420 nm.

We used the Stern-Volmer equation[41] to analyze the quenching of fluorescence intensity at the emission maxima, i.e.,

$$\frac{F_0}{F} = 1 + K_{SV}[Q] \qquad (eq\ 2.21)$$

where, F_0 and F are the fluorescence intensities at the emission maxima at 0 M and any particular acrylamide concentration $[Q]$, respectively; and K_{SV} is the Stern-Volmer constant which is given by:[41]

$$K_{SV} = \tau_0 k_q \qquad (eq\ 2.22)$$

In the above equation, τ_0 and k_q are the intensity averaged fluorescence lifetime and bimolecular quenching constants, respectively. We determined the values of τ_0 for W214 in the N and the U states at pH 7 and pH 5 from the fluorescence lifetime data reported in the previous studies.[42-43] The fluorescence lifetime decays of W214 fit to a sum of three exponentials. Using the three components of the fluorescence lifetimes and their relative amplitudes, we calculated the values of τ_0 using the following equation:[30, 41]

$$\tau_0 = \frac{\sum_i \alpha_i \tau_i^2}{\sum_i \alpha_i \tau_i} \qquad (eq\ 2.23)$$

where, α_i is the amplitude of the i^{th} fluorescence lifetime τ_i such that $\sum_i \alpha_i = 1$ and i = 1-3. The calculated values of τ_0 are listed in Table 2.3.

For the I state, τ_0 was estimated as per the following procedure. Using the fluorescence lifetime data reported in the previous studies,[42-43] we first calculated the amplitude averaged fluorescence lifetime,[41] τ_m, using the following equation:[30, 41]

$$\tau_m = \sum \alpha_i \tau_i \qquad (eq\ 2.24)$$

where, α_i is the amplitude of the i^{th} fluorescence lifetime τ_i such that $\sum_i \alpha_i = 1$ and i = 1-3. Since, quantum yield, Q_D, is directly proportional to τ_m [41] at any particular condition, we used the known values of Q_D and τ_m for the N state at pH 7 (0.31 and 6.2 ns, respectively) and the value of Q_D for the I state (0.27) and calculated the value of τ_m for the I state to be 5.4 ns. In

the previously reported studies,[42-43] we observed that the ratio of τ_0 with respect to τ_m for all the conditions is 1.11. Due to this, we assumed that for the I state also, this ratio will hold good and we calculated the value of τ_0, for the I state to be 6.0 ns.

The values of k_q were calculated from the values of K_{SV} and τ_0, using equation 2.22. We observed that for the U states at pH 7 and pH 5, the values of k_q decreased by 1.4 fold in the presence of 9 M urea. The values of k_q were corrected for the viscosity of 9 M urea. [30] All the values of k_q are mentioned in Table 2.3.

REES experiments

The fluorescence emission spectra for the REES experiments were collected after excitation at different wavelengths ranging from 295 nm to 305 nm. The protein concentration used were 8-15 µM. The spectra were recorded with a scan speed of 80 nm/min and averaged over three scans. A slit width of 5-8 nm was used for excitation. The error bars for the values of the emission maximum were estimated from two independent measurements.

2.3 Results and Discussion

2.3.1 pH dependence of the spectroscopic properties of the N state ensemble of HSA

The three-dimensional structure of HSA consists of twenty eight helices packed and distributed over its three domains (Figure 2.1A). The three domains of the protein are held together by a tightly packed inter-domain cluster of hydrophobic atoms of residues L198, S202, F206, A210, W214, A217, V343, V344, L347, N458, L481, V482 and R484. The sole tryptophan residue of the protein, W214, is located in a helical segment of domain II (Figure 2.1A). During pH titration of the protein, we observed that the fluorescence of W214 remains constant between pH 4.8 to pH 8.5, but decreases in a sigmoidal manner below pH 4.8 and above pH 8.5 (Figure 2.1B). We also monitored the changes in the secondary structure of the protein during pH titration by measuring the changes in far-UV CD signal at 222 nm (Figure 2.1B). We observed that the secondary structure of the protein decreases marginally below pH 4.8, but remains almost constant from pH 4.8 to pH 12. These results indicate that HSA undergoes an acid-induced structural transition below pH 4.8 and a base-induced structural transition above pH 8.5, as reported in previous studies.[44-45] The acid-induced structural transition is accompanied by changes in both the tertiary and the secondary structure of the protein, but the base-induced transition is only accompanied by changes in the tertiary

structure. Most importantly, these results suggest that HSA maintains N-like secondary and tertiary structure between pH 4.8 and 8.5, which constitute its N state ensemble.

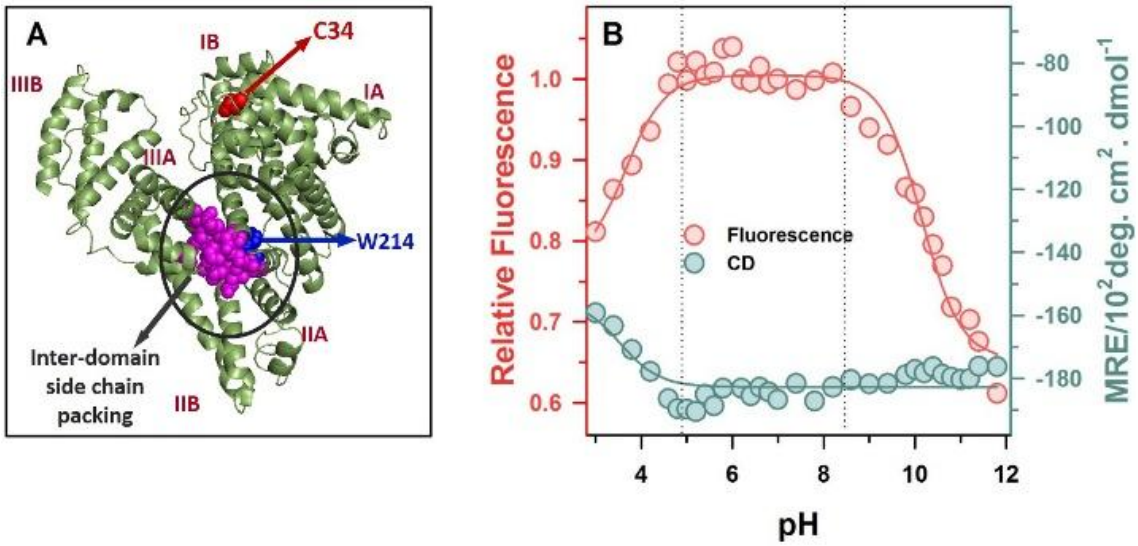


Figure 2.1*. HSA maintains N-like conformation between pH 4.8 to pH 8.5. (A) The location of W214 is shown which is involved in the inter-domain side-chain packing in the three domain structure of the protein (drawn from PDB file 1AO6 using the program PyMOL). The location of C34 in domain I is also shown. (B) Mean residue ellipticity (MRE) of the protein at 222 nm and fluorescence emission of W214 at 340 nm are shown as a function of pH according to the right and the left y-axis, respectively. The green and the red line through the data represent a fit to a two-state (equation 2.1) and a three-state (equation 2.2) pH titration model, respectively.*

2.3.2 Thermodynamic characterization of the N state ensemble under different N-like conditions

We measured the structural stability of the N state ensemble at two different N-like environmental conditions, pH 5 and pH 7, by urea denaturation experiments (Figure 2.2). The fluorescence Stokes shift of the tryptophan residue is very sensitive to the polarity of its surrounding medium. W214 is partially buried in the hydrophobic core at the interface of domain I and II in the N state (Figure 2.1A). Accordingly, we observed that, both at pH 5 and pH 7, the mean value of the wavelength of maximum fluorescence emission (λ_{max}^{em}) in the N state is 340 nm (Figure 2.2A and 2.2A: inset). The λ_{max}^{em} shifts to red to 350 nm in the U state in 9 M urea, both at pH 5 and pH 7, indicating the complete hydration of W214 upon unfolding (Figure 2.2A, 2.2A: inset and 2.2B). At pH 7, the λ_{max}^{em} of W214 changes in apparently two-

state fashion as a function of [urea], indicating that W214 broadly experiences only N-like and U-like solvation during unfolding (Figure 2.2A: inset and 2.2B). However, at pH 5, we observed that λ_{max}^{em} of W214 changes in two distinct sigmoidal steps during urea-induced unfolding (Figure 2.2A and 2.2B). At pH 5, the λ_{max}^{em} first shifts to blue to 330 nm and plateaus around 1.6 M urea, which later shifts to red to 350 nm upon complete unfolding in 9 M urea. These results indicate that urea-induced unfolding at pH 5 occurs *via* an I state in which tertiary structural changes at the interface of domain I and II allow the movement of W214 from a partially exposed to a highly hydrophobic environment.

Because λ_{max}^{em} is thermodynamically an intensive property, we monitored the changes in fluorescence intensity of W214 as a function of [urea] to determine the changes in free energies and populations of the N, the I and the U states during N⇌I⇌U transition. It is important to note that the observed fluorescence intensity at any wavelength is a measure of the fluorescence intensity of all the different species present in the equilibrium mixture and their fractional population. Therefore, by monitoring fluorescence intensity at any wavelength, we are actually monitoring how the population of different species is changing as a function of [urea]. We observed that fluorescence intensity of W214 during urea-induced equilibrium unfolding transition changes in a single sigmoidal step at pH 7, but in two distinct sigmoidal steps at pH 5 (Figure 2.2C). We analyzed our data according to a two-state model at pH 7 and a three-state model at pH 5 (Figure 2.2C: inset). The thermodynamic parameters derived from the global analysis of the data are listed in Table 2.1. We observed that the standard free energy of unfolding during N⇌U transition , $\Delta G_{NU}^{H_2O}$, at pH 5 is 5.3 ± 0.7 kcal mol^{-1} and at pH 7 is 5.4 ± 0.3 kcal mol^{-1}, respectively. The $\Delta G_{NU}^{H_2O}$ at both the pH are similar (within error), indicating that the structural stability of the N state is same at the two N-like conditions. The value of $\Delta G_{NI}^{H_2O}$, the standard free energy of unfolding during N⇌I transition, is 2.1 ± 0.3 kcal mol^{-1}, indicating that 35-40% of the change in the total free energy of unfolding occurs during the partial unfolding of the N state to the I state.

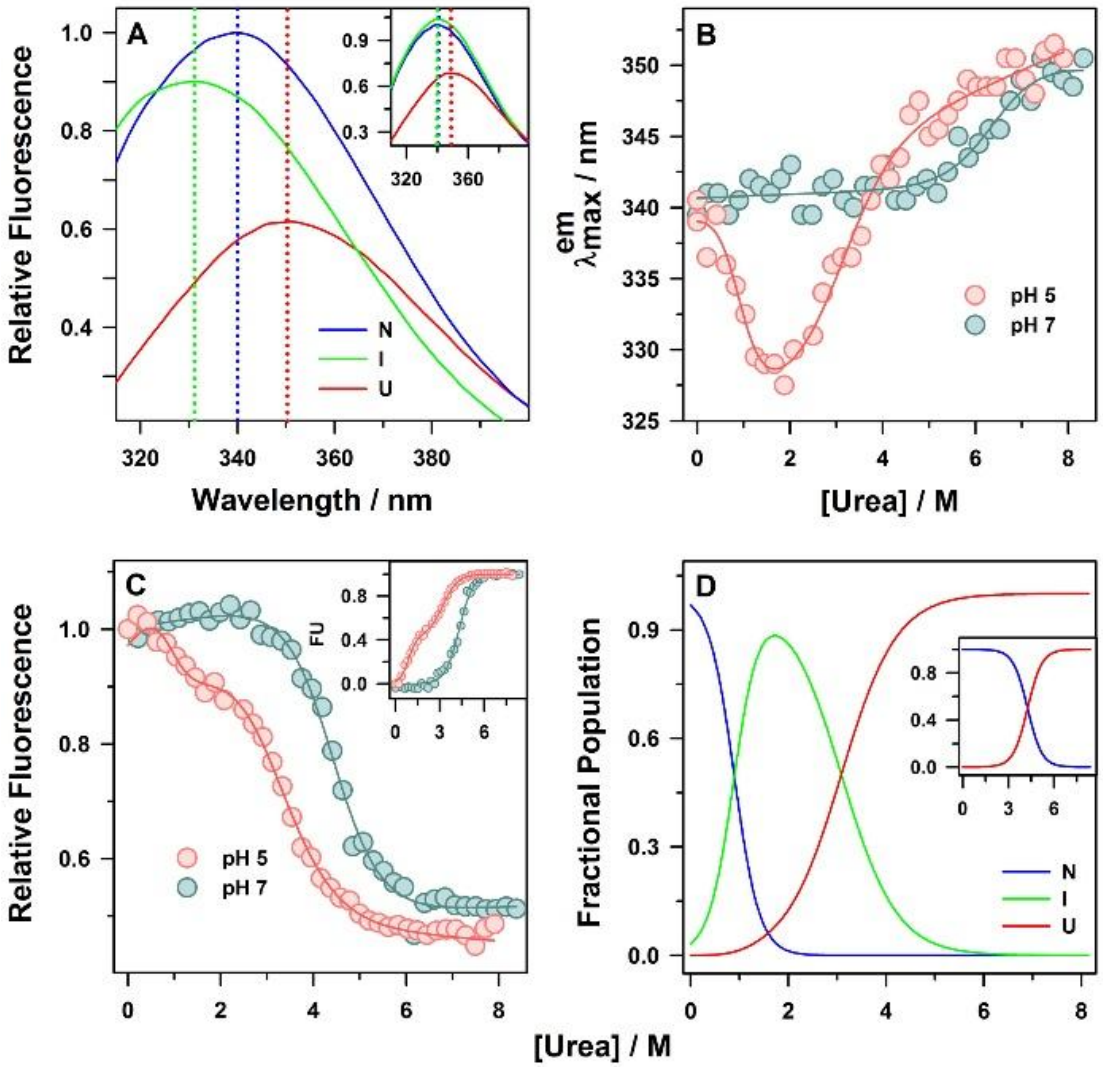


Figure 2.2. *Urea-induced structural transition of HSA at pH 5 and pH 7. (A) Panel A and its inset, respectively, show the fluorescence emission spectra of the protein at pH 5 and pH 7 in 0 M (solid blue line), 1.6 M (solid green line) and 9 M (solid red line) urea. The dotted lines represent the wavelength of maximum fluorescence in the respective conditions. (B) Changes in the wavelength of maximum fluorescence emission (λ^{em}_{max}) are plotted as a function of [urea]. (C) Equilibrium unfolding transition monitored by the change in W214 fluorescence at 328 nm. The inset shows the apparent fraction of the unfolded protein (FU) as a function of [urea]. (D) The changes in the population of the N, the I and the U states at pH 5 and the N and the U states at pH 7 (inset) are plotted as a function of [urea] (equation 2.9 to 2.11, 2.5 and 2.6,). The blue and the red lines through the data in panels B and C and in the inset of panel C are fits to a two-state and a three-state model, respectively. The axes of the insets in panels A, C and D, wherever not labeled, are same as that of the main panel.*

Table 2.1. Comparison of thermodynamic parameters for urea-induced equilibrium unfolding of HSA at pH 5 and pH 7 by fluorescence intensity (flu), CD and λ_{max}^{em}

parameters[a]	**flu / CD**[b] **(pH 7)**	**flu / CD**[b] **(pH 5)**	$\boldsymbol{\lambda_{max}^{em}}$ [c] **(pH 5)**
$\Delta \boldsymbol{G_{NI}^{H_2O}}$ (kcal mol^{-1})	-	2.1 ± 0.3	2.1 ± 1.2
$\Delta \boldsymbol{G_{IU}^{H_2O}}$ (kcal mol^{-1})	-	3.2 ± 0.8	3.2 ± 1.4
$\Delta \boldsymbol{G_{NU}^{H_2O}}$ (kcal mol^{-1})	5.4 ± 0.3	5.3 ± 0.7	5.3 ± 0.7
$\boldsymbol{m_{NI}}$ (kcal mol^{-1} M^{-1})	-	-2.3 ± 0.4	-2.6 ± 3.0
$\boldsymbol{m_{IU}}$ (kcal mol^{-1} M^{-1})	-	-1.0 ± 0.5	-1.0 ± 4.1
$\boldsymbol{m_{NU}}$ (kcal mol^{-1} M^{-1})	-1.2 ± 0.1	-3.3 ± 0.3	-3.6 ± 2.9
$\boldsymbol{z_I}$ [d]	-	0.4 ± 0.2	0.1 ±0.8

[a] All the thermodynamic parameters for pH 5 and pH 7 were determined, respectively, from the three-state (N⇌I⇌U) and the two-state (N⇌U) analysis of urea-induced equilibrium unfolding experiments. A global analysis was performed as discussed in Section 2.2. All the parameters shown in this table can be used to generate all the fits shown in Figures 2.2C, 2.2C inset, and 2.3A inset.

[b] For pH 5, the error bars shown were estimated from three independent measurements whereas, for pH 7, the errors shown were resulted from five independent measurements.

[c] Although λ_{max}^{em} is an intensive property and not directly proportional to the amount of the molecule, we fitted the pH 5 data in Figure 2.2B using the three-state equilibrium unfolding model, and constraining the value of $\Delta G_{NU}^{H_2O}$. Very surprisingly, we observed that the thermodynamic parameters are quite similar to the equilibrium unfolding curve monitored by fluorescence and far-UV CD.

[d] z_I is a measure of the optical signal of I relative to the optical signals of the N and the U states.

We determined the changes in the populations of the N, the I and the U states as a function of [urea] by using the thermodynamic parameters for N⇌I, I⇌U and N⇌U transitions (Figure 2.2D). We found that the I state is maximally populated to the extent of 85 ± 4% at ~1.6 M urea, where the population of the N state is 8 ± 6% and the U state is 7 ± 3%, respectively. No intermediate state was observed to be populated at pH 7 (Figure 2.2D: inset). Very interestingly, we observed that around 5% of the I state is populated under N-like conditions at pH 5, i.e., at 0 M urea. This result indicates that the N state ensemble at pH 5 is structurally and energetically heterogeneous.

2.3.3 Comparison of the global structure of the N and the I state

We maximally populated the I state at 1.6 M urea (pH 5) and characterized its structure in terms of secondary and tertiary structural content, side-chain packing, core hydration and solvation dynamics of the protein side-chains in its core. First, we compared the global secondary structure and stability of the I state and the N state ensemble by far-UV CD spectroscopy. We observed that the far-UV CD spectrum of the N state and the U state at pH 5 are identical to the N state and U state at pH 7 (Figure 2.3A). The mean residue ellipticity (MRE) at 222 nm is a measure of the secondary structural content in proteins. We observed that the mean value of MRE at 222 nm for the I, the N and the U states are -18850 deg cm^2 $dmol^{-1}$, -20650 deg cm^2 $dmol^{-1}$ and -4950 deg cm^2 $dmol^{-1}$, respectively. The change in MRE during N⇌I transition is only 10% of the total change in MRE between the N and the U states. Because 7 ± 3% of unfolded molecules are also present at 1.6 M urea at pH 5 (Figure 2.2D), this result indicates that the secondary structure of the I state is similar to the N state, and that the protein-protein hydrogen bonds (H-bonds) are not replaced by protein-water H-bonds.

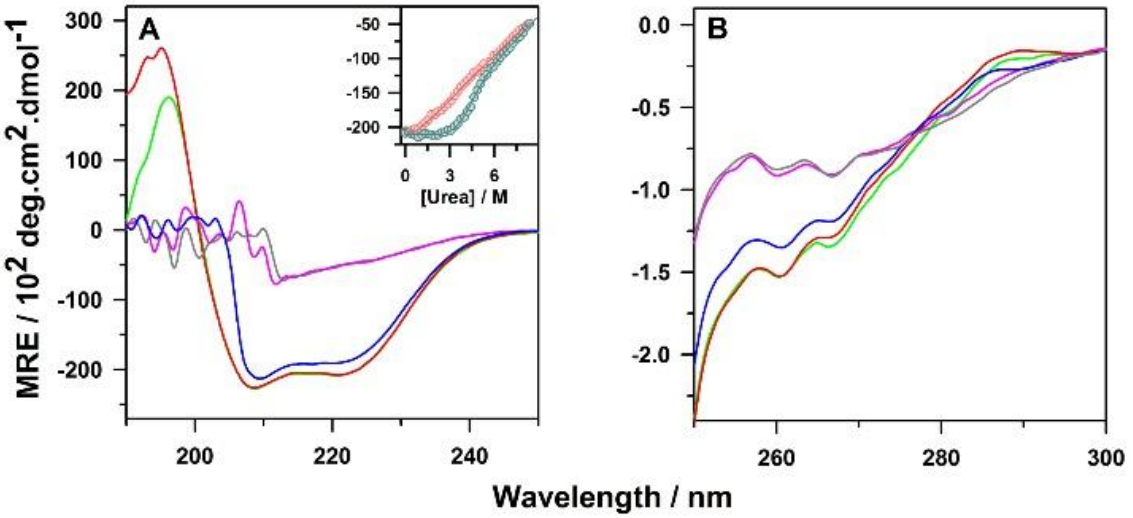


***Figure 2.3**. Global structural analysis of HSA at pH 5 and pH 7. (A) Changes in MRE as a function of wavelength in the peptide region (secondary structure) and (B) in the aromatic region (tertiary structure). In both the panels, the signal of the N state at pH 5 and pH 7 are shown by the green line and the dark red line, respectively; the signal of the U state at pH 5 and pH 7 are shown by the gray line and the pink line, respectively, and the signal of the I state is shown by the dark blue line. The inset in panel A shows the changes in MRE at 222 nm as a function of [urea] at pH 5 (red circles) and pH 7 (cyan circles). The blue and the red lines through the data in the inset of panel A are fits to a two-state and a three-state model, respectively. The y-axis of the inset in panel A has the same label as that of the main panel.*

We also measured the stability of the secondary structure of the N state ensemble at pH 5 and pH 7. We observed that the stability of the N state ensemble at pH 7 as measured by far-UV CD is similar to that of the fluorescence (Figure 2.3A: inset; Figure 2.2C and Table 2.1). For both fluorescence and CD measured unfolding transitions, the midpoint of transition is 4.5 M [urea]. At pH 5, we observed a three-state urea-induced equilibrium unfolding transition (Figure 2.3A: inset). However, the change in MRE between the N and the I states is very small and unfolding of the secondary structure in the I state appears to be gradual and noncooperative. Nevertheless, the equilibrium unfolding curve can be satisfactorily fitted to a three-state model using the similar parameters as that of the fluorescence monitored equilibrium unfolding curve (Table 2.1).

We next compared the global tertiary structure of the I state and the N state ensemble by near-UV CD spectroscopy (Figure 2.3B). The near-UV CD spectrum of HSA in the N state has absorption bands mainly in the 255 to 270 nm region (Figure 2.3B), indicating that it primarily reports on the chirality created by the asymmetric packing of its 31 phenylalanine residues in the hydrophobic core. The near-UV CD spectra of the N state and the U state at pH 5 are very similar to the N state and the U state at pH 7. We observed that MRE at 261 nm for the I, the N and the U states are -135 ±15 deg cm^2 $dmol^{-1}$, -150 ± 15 deg cm^2 $dmol^{-1}$ and -85 ± 15 deg cm^2 $dmol^{-1}$, respectively, and around 10-20% of the total change in MRE occurs during N⇌I transition. Accounting for the decrease in MRE due to the presence of 7 ± 3% of unfolded molecules, these results suggest that the global tertiary structure in the I state and the N state are only marginally different. This is a surprising result in the light of the large blue shift of the λ_{max}^{em} of W214 in the I state (Figure 2.2A and 2.2B), which indicates that the side-chain packing

interactions involving W214 must be broken in the I state. In the crystal structure of HSA, we observed that 30 of the 31 phenylalanine residues are mainly distributed in the core of the three domains and only one phenylalanine residue, F206, is spatially located in the inter-domain region near W214 and participates in the inter-domain side-chain packing (Figure 2.4A). Hence, the observation that the λ_{max}^{em} of W214 is highly blue shifted in the I state but its near-UV CD spectrum is similar to the N state suggests that the intra-domain side-chain packing is intact in the I state but the inter-domain side-chain packing is disrupted.

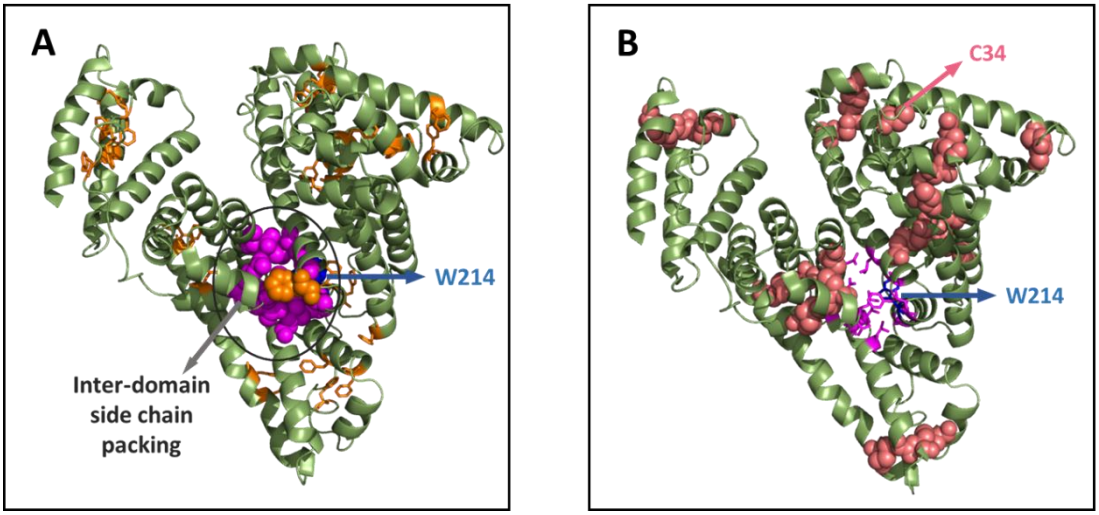


Figure 2.4. *The structure of HSA showing all the phenylalanine residues and cysteine residues. (A)The phenylalanine residues are shown in orange colored sticks except F206 (orange sphere) and the residues participating in the inter-domain side-chain packing are shown in magenta spheres and includes W214 (blue sphere) and F206 (orange sphere). (B) All the cysteine residues are shown in pink colored spheres and the residues participating in the inter-domain side-chain packing are shown in magenta sticks and includes W214 in the blue stick. Except C34, all the cysteine residues are involved in intra-domain disulfide bonds. The disulfide stitches provide stability to the intra-domain regions, but not the inter-domain region. These structures were drawn from PDB file 1AO6 using the program PyMOL.*

2.3.4 Inter-domain expansion in the I state as revealed by site-specific FRET measurements

The disruption of inter-domain side-chain packing must lead to the movement of the protein domains away from each other. To probe this further, we measured the inter-domain distances between domain I and II in the N state and the I state using FRET. We used W214 which is located in the middle of a helix in domain II as the donor (D) fluorophore. HSA has

a single free cysteine residue, C34, which is located at the beginning of helix 3 in domain I (HSA has 35 cysteine residues in total, 34 of which form disulfide bonds; Figure 2.4B). We labeled the thiol moiety of C34 with the fluorescent dye 1,5-IAEDANS, and used it as the acceptor (A) fluorophore. We have shown previously that C34 can be quantitatively labeled with 1,5-IAEDANS[30] and that W214 and C34- IAEDANS form a FRET pair.[30] The emission spectrum of W214 significantly overlaps with the absorbance spectrum of C34-IAEDANS (Figure 2.5A and 2.5B). The secondary structure and thermodynamic stability of the unlabeled (HSA) and the 1,5-IAEDANS labeled (HSA-IAEDANS) proteins are also similar,[30] and hence the data on HSA and HSA-IAEDANS can be compared directly. C34-IAEDANS quenches the fluorescence of W214 to a larger extent in the N state compared to the U state (Figure 2.6A). This result is expected because C34-IAEDANS and W214 are located nearer to each other in the N state but far away in the U state. We observed that the extent of quenching of W214 fluorescence by C34-IAEDANS in the I state is significantly less than that in the N state (Figure 2.6A and 2.6B).

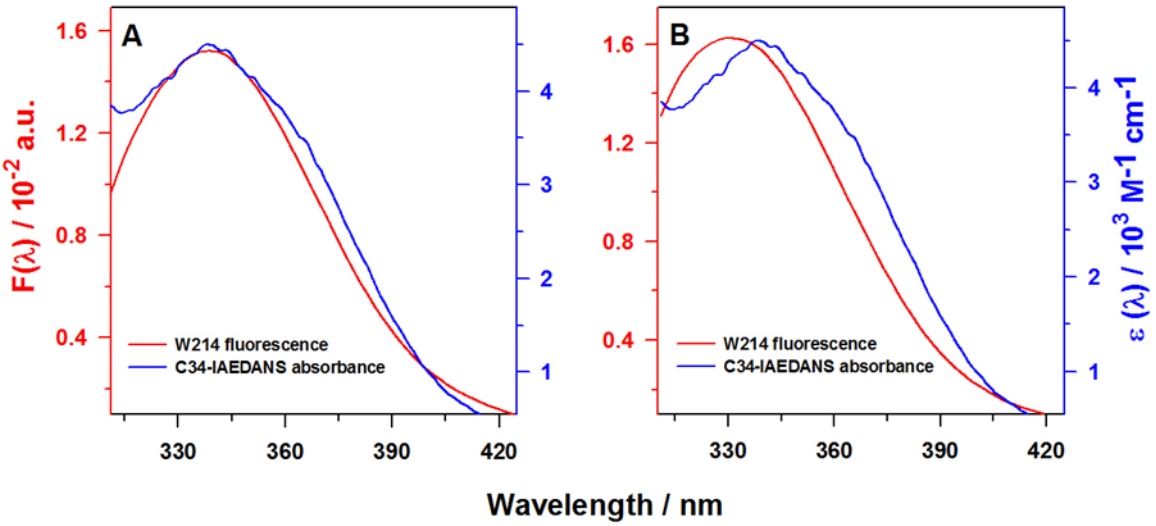


Figure 2.5. *Determination of the overlap integral, J, in the N state at pH 5. (A) and in the I state (B). In all the figures, the fluorescence emission spectra of W214 and the absorption spectra of C34-IAEDANS are shown according to the left Y-axis and the right Y-axis, respectively. Each fluorescence spectrum has been normalized so that the total area under the spectrum is unity. Each absorption spectrum has been divided by the respective molar protein concentration to obtain ε(λ).*

We used the data on the fluorescence intensity of the donor in the absence and the presence of the acceptor to quantitate the D-A distance in the N state and the I state using equation 2.18. The values of Forster's distance are determined experimentally and found to be 25.7 Å in the N state and 25.0 Å in the I state (Figure 2.5B and Table 2.2). We observed that the D-A distance in the N state is 29.3 ± 0.1 Å and that in the I state is 30.3 ± 0.1 Å. The D-A distance in the I state is only 1 Å larger than the N state. This difference, however, is significant because of the following four reasons: (i) If a D-A pair is nearer to each other, the energy transfer by FRET results in the decrease of the fluorescence emission of the donor fluorophore (quenching) but an increase in the fluorescence emission of the acceptor fluorophore. We observed that the fluorescence of C34-IAEDANS in the N state is higher than that in the I state upon excitation at 295 nm (Figure 2.6B: inset). The change in fluorescence intensity between the N and the I states is ~33% of the total change in the fluorescence intensity between the N and the U state. These results indicate that extent of energy transfer is more in the N state than the I state and least in the U state, and hence the D-A distance in the N state is less than the I state and the largest in the U state; (ii) The measured D-A distances in the N and the I states are very near to R_0 and hence lie in the most sensitive region of the FRET measurements (between 0.5 R_0 to 1.5 R_0); (iii) We determined the D-A distances in the N and I states by 3-5 separate measurements and observed that the standard error in D-A distances is only 0.1 Å (Table 2.2); (iv) Because most of the secondary and tertiary structures of the protein are intact in the I state, a large change in inter-atomic distances is not even expected. Hence, the results of the FRET experiments imply that domain I and domain II have moved away from each other by 1 Å in the I state compared to the N state, along the axis connecting W214 and C34-IAEDANS. Because the strength of vdW interactions varies steeply with inter-atomic distances, these results indicate that the inter-domain packing interactions between domain I and II must be loose in the I state. The loosening of side-chain packing appears sufficient to allow the movement of the W214 side-chain to a more hydrophobic region of the domain-domain interface in the I state. It should be kept in mind, however, that the loosely packed N-like intermediate states could also be compact in size than the N states, as has been reported recently for HP35.[19]

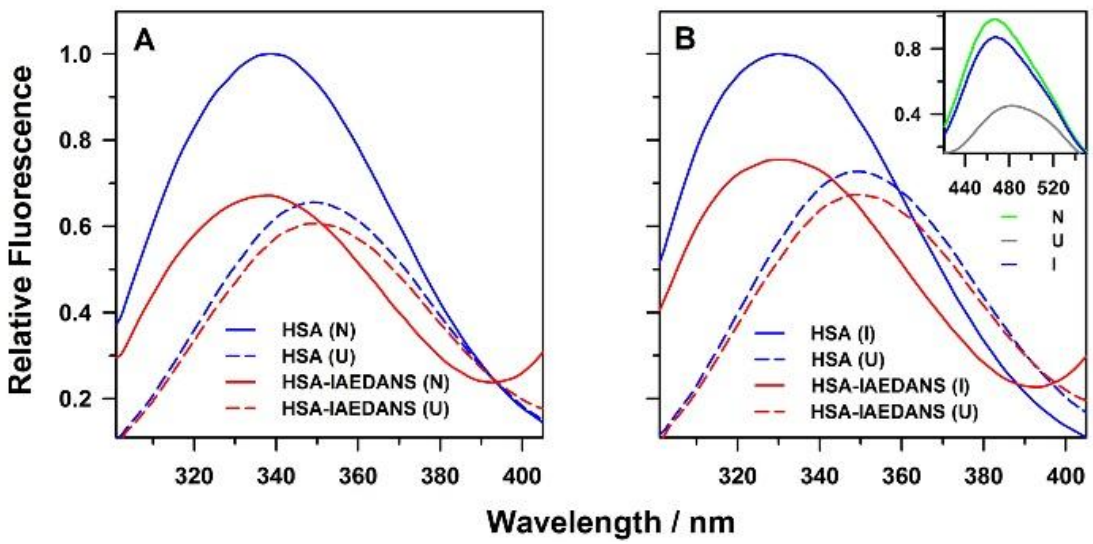


Figure 2.6*. The inter-domain region in the I state is expanded. The distance between W214 and C34-IAEDANS was monitored by FRET. Fluorescence emission spectra of W214 in HSA and HSA-IAEDANS at pH 5 are shown in panel (A) for the N and the U states and in panel (B) for the I and the U states. The inset in panel (B) shows the fluorescence emission spectra of C34-IAEDANS in the N, the I and the U states upon excitation at 295 nm. The axes labels of the inset in panel B are same as that of the main panel.*

Table 2.2. Values of FRET efficiency (E), quantum yield (Q_D), overlap integral (J), Förster's distance (R_0), and D-A distance (R) for FRET between W214 and C34-IAEDANS pair

Conditions	$^{*}E$	$^{\S}Q_D$	$^{\dagger}J$/ $10^{13}M^{-1}cm^{-1}$ nm^4	$^{\ddagger}R_0$ / Å	$^{\ddagger}R$ / Å
N (pH 5)	0.31 ± 0.005	0.30 ± 0.005	5.1	25.68	29.3 ± 0.13
I	0.24 ± 0.004	0.27 ± 0.005	5.1	25.03	30.3 ± 0.12

$^{*}E$ was determined using the fluorescence spectra in Figures 2.6A and 2.6B using equation 2.17.

$^{\S}Q_D$ was determined as described in the Section 2.2.

$^{\dagger}J$ was calculated using equation 2.20.

‡ R_0 was calculated using equation 2.19 with the following values: $\kappa^2 = 2/3$; $n = 1.332$ (for the N state) and 1.346 (for the I state).

‡ R was calculated using equation 2.18.

Note: The errors reported in the values of different parameters denote the standard error estimated from three or more separate measurements.

2.3.5 Dynamic fluorescence quenching experiments reveal that the inter-domain region is dry

We observed that the disruption of side-chain packing between domains I and II is not accompanied by hydration of the intervening hydrophobic core. The λ_{max}^{em} of W214 in the I state is 330 nm indicating that its immediate surrounding is hydrophobic in nature (Figure 2.2A and 2.2B). We used dynamic fluorescence quenching experiments[41, 46-47] to understand whether loose structure around W214 allows it to undergo structural fluctuations and come in molecular contact with the quencher molecules present in water. Acrylamide is a neutral collisional quencher of tryptophan fluorescence.[41] We measured the fluorescence intensity of W214 in the absence and in the presence of different concentrations of acrylamide in the N, the I and the U states (Figure 2.7A) and used the Stern-Volmer equation[41] to determine the values of Stern-Volmer constant, K_{SV}, and the bimolecular rate constant, k_q, for the formation of the molecular contact between W214 and acrylamide in the photo-excited state of W214. Physically, K_{SV} signifies the change of fluorescence intensity of a fluorophore as a function [quencher] and k_q is a measure of the rate of molecular contact between the fluorophore and the quencher. The values of K_{SV} in the U states at pH 5 and pH 7 are 9.22 M^{-1} and 9.26 M^{-1}, respectively. The values of K_{SV} in the N states at pH 5 and pH 7 are 4.51 M^{-1} and 6.88 M^{-1}, respectively. The value of K_{SV} in the I state is 5.02 $M^{-1,}$ which is very similar to the value observed in the N state ensemble at pH 5.

We used these values of K_{SV} to determine the values of k_q. We observed that the values of k_q in the U state at pH 5 and pH 7, where W214 is completely exposed to water, are 4.20 x 10^9 $M^{-1}s^{-1}$ and 4.02 x 10^9 $M^{-1}s^{-1}$, respectively (Figure 2.7A: inset). The values of k_q decrease to 0.69 x 10^9 $M^{-1}s^{-1}$ and 0.98 x 10^9 $M^{-1}s^{-1}$ in the N states at pH 5 and pH 7, respectively, due to the protection provided by the protein structure around W214. The mean value of k_q is ~5-fold smaller in the N state compared to the U state. However, the errors associated with the determination of the values of K_{SV} and τ_0 are very small (Table 2.3). Hence, the errors in the

values of k_q are unlikely to be more than ± 10% and ~5-fold dynamic range in the value of k_q is quite significant. We observed that the value of k_q in the I state is 0.73 x 10^9 $M^{-1}s^{-1}$ which is similar to that in the N state (Figure 2.7A: inset). These results indicate that in the I state and the N state, W214 is protected against water-solvation to the similar extent. Hence, the loose side-chain packing in the inter-domain region in the I state does not expose hydrophobic side-chains to water molecules, and the inter-domain region appears to be like a dry globule.

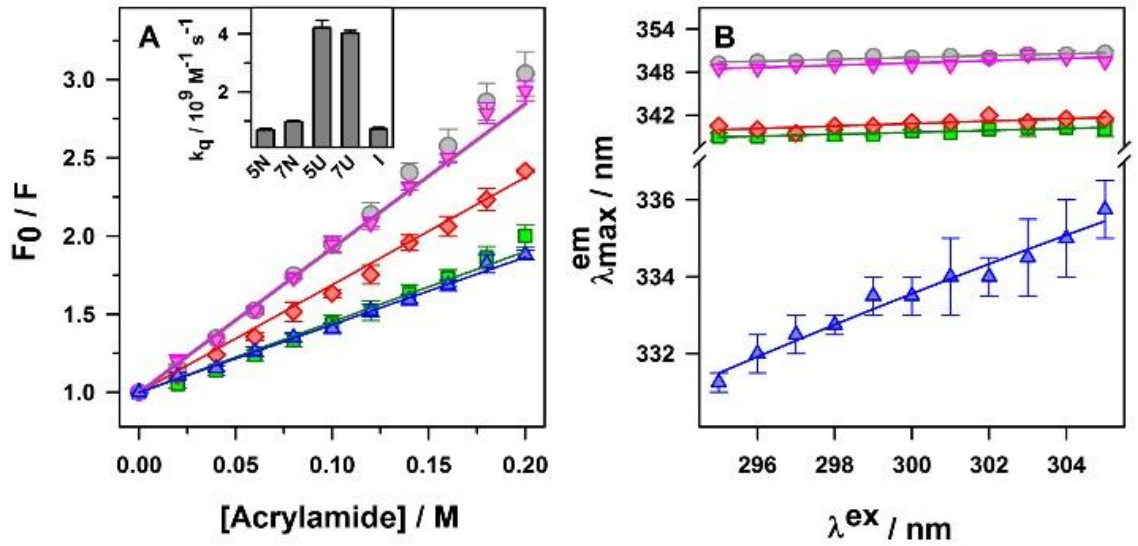


Figure 2.7*. The inter-domain region in the I state has a hydrophobic and molten interior. (A) Stern-Volmer plots for quenching of fluorescence of W214 by acrylamide in the N (green squares), the I (dark blue triangles) and the U (gray circles) states at pH 5, and the N (red diamonds) and the U (pink inverted triangles) states at pH 7. The solid lines through the data are the least square fits to equation 2.21. The values of bimolecular quenching constant, k_q, are compared in the inset. (B) The wavelength of maximum fluorescence emission (λ_{max}^{em}) of W214 is dependent on the wavelength of excitation (λ^{ex}) in the I state (dark blue triangles) but is largely independent in the N (green squares) and the U (gray circles) states at pH 5 and the N (red diamonds) and the U (pink inverted triangles) states at pH 7. In panel B, the solid lines through the data are drawn to guide the eyes.*

Table 2.3. Values of Stern-Volmer constants (K_{sv}), intensity averaged fluorescence lifetimes (τ_0) and bimolecular quenching rate constants (k_q) for W214

Conditions	K_{sv} /M^{-1}	$^{*}\tau_0$ /ns	k_q /10^9 $M^{-1}s^{-1}$
pH 5 N	4.51 ± 0.13	6.8	0.69 ± 0.06
pH 7 N	6.88 ± 0.05	6.9	0.98 ± 0.04
pH 5 U	9.22 ± 0.16	3.3	4.20 ± 0.25[§]
pH 7 U	9.26 ± 0.09	3.3	4.02 ± 0.10[§]
Intermediate	5.02 ± 0.12	6.0	0.73 ± 0.04

[*] These values are calculated from the fluorescence lifetime data reported in previous studies.[42-43]

[§] These values are corrected for the effects of viscosity of 9 M urea, as described previously.[30]

Note: The errors shown for the above parameters denote the standard error estimated from three or more separate measurements.

2.3.6 The solvation dynamics of the side-chains in the inter-domain region is heterogeneous

It has been predicted that the hydrophobic core of dry globules might resemble a molten liquid due to the disruption of the side-chain packing interactions.[10, 17-18] However, it has been very difficult to probe the degree of mobility of side-chains in the core of the dry globules. We investigated the solvation dynamics of the protein matrix around W214 in the I state using REES experiments.[41, 48] In general, the fluorescence emission spectrum of a fluorophore is independent of the excitation wavelength (λ^{ex}) because of the faster relaxation (compared to the ns fluorescence timescale) of the solvent molecules surrounding it. However, when the distribution of solvent molecules around the fluorophore is heterogeneous (leading to a broad distribution of solute-solvent interaction energies) and the dynamics of solvent molecules is much slower than the fluorescence timescale (like in a viscous solvent), the fluorescence emission spectrum becomes dependent on the λ^{ex}, when excited around the red edge of the absorption spectrum.[41, 48] We observed that the λ_{max}^{em} of W214 depends very strongly on the λ^{ex} in the I state but shows no dependence in the N state and the U state, respectively (Figure 2.8A and 2.8B, Figure 2.7B). The λ_{max}^{em} of W214 in the I state shifts to 336 nm from 330 nm, when it is excited at 305 nm compared to 295 nm (Figure 2.7B). In contrast, the λ_{max}^{em} of W214 remains constant at 340 nm and 350 nm in the N state and the U state, respectively (Figure

2.7B, Figure 2.8), when the λ^{ex} is changed from 295 nm to 305 nm. These results indicate that water molecules surrounding W214 in the U state and the complex solvation environment around it in the N state is highly dynamic. In the I state, W214 is buried in the hydrophobic core at the interface of domains I and II (Figures 2.2A and 2.2B). Hence, the large magnitude (6 nm) of REES observed in the dry globular I state indicates that the solvation environment created by the dipoles of the protein matrix and side-chains around W214 in the inter-domain region is (i) highly heterogeneous (ii) relaxes much slower than the ns fluorescence timescale of W214 (most probably in the tens of µs) and (iii) is similar to a molten and viscous liquid.[48]

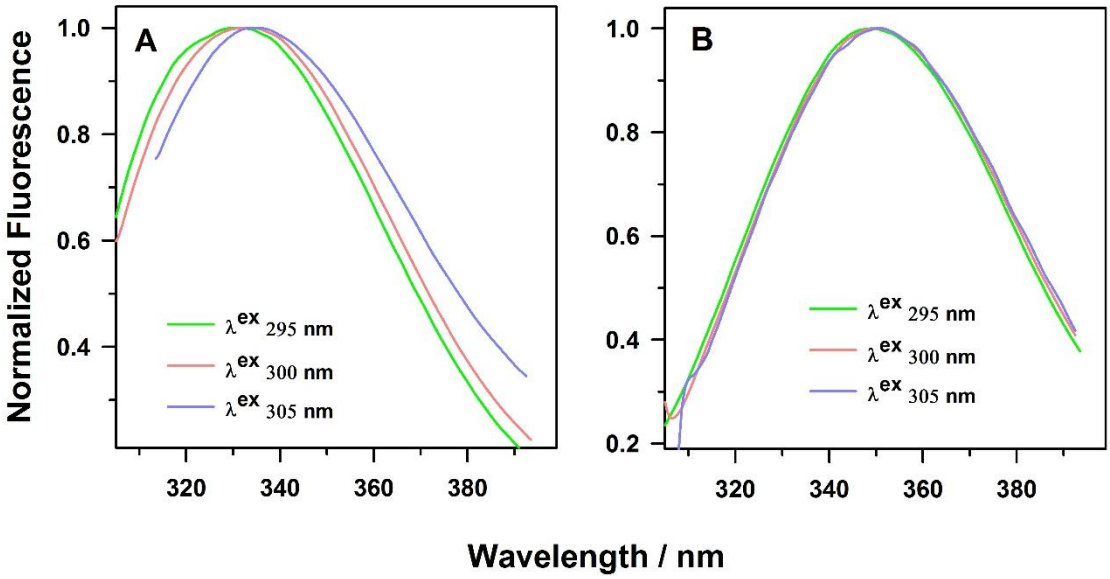


Figure 2.8. *Red-edge excitation shift of W214. Panel A and B, respectively, show the representative fluorescence emission spectra of W214 in the I state and the U state at pH 5, when excited at different wavelengths of excitation, λ^{ex}. For comparison, all the spectra were normalized to 1 at their emission maximum.*

The above results suggest that the I state is like a DMG in which the intra-domain packing remains similar to the N state but inter-domain packing is loose. It should be noted that the three-dimensional structure of HSA contains a network of 34 disulfide bonds that are distributed all over the protein structure except in the inter-domain region (Figure 2.4B). The intra-domain region in all the three domains appears to be stapled by disulfide bonds. It is highly likely that these disulfide stitches stabilize the structure in the intra-domain region

compared to the inter-domain region in both the N and the I states and contribute in restricting the intra-domain side-chain dynamics.

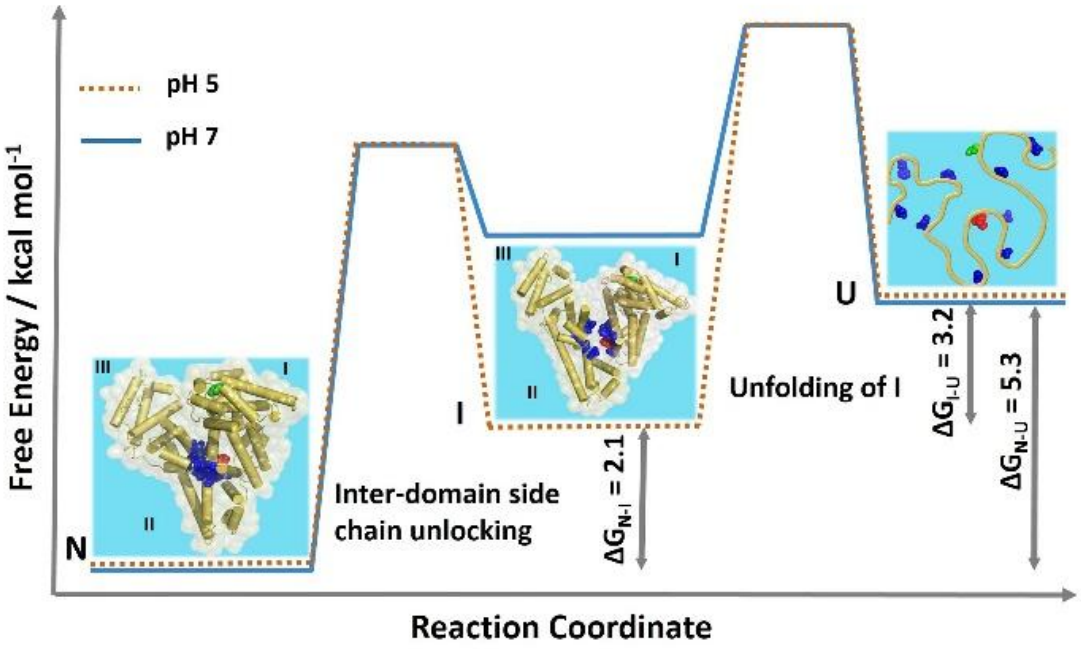


Figure 2.9*. Model energy diagram. The schematic figure shows the relative energies of the N, the I and the U states. In the I state, the protein is expanded and the side-chain packing is disrupted in the inter-domain region allowing W214 to move to a hydrophobic cavity. These events occur without the hydration of the intervening hydrophobic region. Structural models of the N and the I states are drawn from the PDB file 1AO6 using the program PyMOL. In the N and the I states, yellow cylinders represent α-helical secondary structure of the protein. In the U state yellow curved line shows the protein backbone. The white boundary around the protein in the N and the I states shows the solvent accessible surface area of the protein. The light blue color represents water molecules. The side-chain residues involved in inter-domain side-chain packing are shown in dark blue color. The side-chains of W214 and C34 are shown in red and green spheres, respectively.*

2.3.7 Inter-domain packing contributes significantly to protein stability

The DMG-like I state has N-like secondary structure and core hydrophobicity, but vdW packing in the inter-domain region is loose. Hence, the difference in thermodynamic stability of the I state compared to the N state can reveal the contribution of inter-domain packing interactions in the stability of this protein (Figure 2.9). As discussed above in Figure 2.2C and Table 2.1, $\Delta G_{NU}^{H_2O}$ is 5.3 kcal mol^{-1} and $\Delta G_{NI}^{H_2O}$ is 2.1 kcal mol^{-1}, indicating that a very large

fraction (35-40%) of the total change in free energy occurs during N⇌I transition. One possible explanation for this large change in free energy is that some side-chain packing interactions are loose in parts of the protein other than the interface of domains I and II. For example, our structural model (Figure 2.9) shows that the movement of domains I and II away from each other could also disrupt some packing interactions between domains I and III. However, the decrease in stability of the I state is not due to the disruption of the intra-domain packing interactions as indicated by the results of the CD experiments (Figures 2.3A and 2.3B). Hence, our thermodynamic measurements indicate that inter-domain side-chain packing interactions significantly contribute to the stability of the native protein and that small changes in the packing interactions could significantly destabilize the native protein. Because, the DMG-like I state is 2.1 kcal mol^{-1} less stable than the N state (Figure 2.9), it appears that the enthalpic cost of the disruption of inter-domain packing interactions during N⇌I transition is only partially compensated by the increase in conformational entropy of the side-chains.

2.3.8 N⇌I transition represents the early stages of chemical denaturation

A three-state fit of the equilibrium unfolding transition at pH 5 also yielded the *m*-values (slope of the linear denaturant dependence of the free energies) for N⇌I and N⇌U transitions (Figure 2.2C and Table 2.1). We observed that the values of m_{NU} and m_{NI} are 3.3 ± 0.3 kcal mol^{-1} M^{-1} and 2.3 ± 0.4 kcal mol^{-1} M^{-1}, respectively. *m*-value changes during structural transitions of proteins are usually interpreted as the change in the solvent accessible surface area during the transition. The value of m_{NI} corresponds to 65-70% of the total change in *m*-value during the complete unfolding of the protein (m_{NU}). Hence, these results suggest that 65-70% of the protein core is accessible to water in the DMG-like I state. This result is surprising in view of the results of dynamic fluorescence quenching (Figure 2.7A) and far-UV CD (Figure 2.3A) experiments which suggest that in the I state, W214 has N-like solvent accessibility and secondary structure and that the protein-protein H-bonds are not replaced by protein-water H-bonds. It is also important to note that during N⇌I transition, W214 moves to a more hydrophobic (and not hydrated) environment compared to the N state (Figure 2.2A and 2.2B). One possible origin of the large change in *m*-values during N⇌I transition could be the direct interaction of the urea molecules with the protein surface during the initial stages of denaturation. It has been shown, both by computer simulations[49-51] and by experiments,[52-54] that denaturants like urea and guanidine hydrochloride unfold proteins by a two-step direct interaction mechanism. In the first step, they preferentially interact with the protein surface

compared to water in the first solvation cell and loosen the protein structure to populate DMG-like structures. Water molecules enter the protein core in the second step to completely unfold the protein. It has been shown for a few proteins, including RNase H[52] and a variant of CTPR protein,[55] that chemical denaturants modulate the equilibrium between N and DMG and that significant changes in the *m*-value occur during N⇌DMG transition. Taken together, these results suggest that the N⇌I transition observed in this study represents the early stages of chemical denaturation and that the large changes in the *m*-value arise due to the direct interaction of urea molecules with the native HSA, shifting the equilibrium towards the formation of the DMG-like I state. Hence, one should be careful in the interpretation of *m*-value changes when equilibrium intermediates are present on the folding pathways of proteins.

2.3.9 pH dependent modulation of the N state heterogeneity

We observed that the I state is only populated at pH 5 but not at pH 7. This is interesting because the structure and the thermodynamic stabilities of the N and the U states at pH 5 and pH 7 are almost identical (Figure 2.1 and 2.2, Table 2.1). Although we observe that the equilibrium unfolding transition at pH 7 is apparently two-state, it is possible that the energy of the I state is higher than the U state at pH 7 and hence it is not populated (Figure 2.9). The decrease in pH to 5 stabilizes the I state considerably to be populated at equilibrium but does not affect the energy of the N state and the U state (Figure 2.9). This proposal is supported by observations on many proteins including, barstar,[39] RNase H[56] and SH3 domain of PI3 kinase[57] where it has been shown that small changes in environmental conditions like single-site mutation and changes in ionic strength can populate high energy intermediate states and convert a two-state folder into a three-state or multi-state folder. Hence, our observation suggests that the side-chain conformational heterogeneity of the N state ensemble can be modulated by changing the external environmental conditions.

2.4 Conclusion

In summary, we probed the conformational heterogeneity of the side-chain packing in the native state ensemble of HSA, a multi-domain protein, under two different N-like conditions, pH 5 and pH 7, where the protein has a similar structure and thermodynamic stability. We observed that the N state ensemble at pH 5 contains a DMG-like I state, which is not populated at pH 7. The I state gets populated to 85 ± 4% at 1.6 M urea. Far- and near-UV CD experiments show that the I state has N-like secondary and tertiary structures. Fluorescence

measurements on W214, which is involved in the inter-domain side-chain packing and spatially located at the interface of the three domains of the protein, show that inter-domain side-chain packing is loose in the I state and W214 has moved to a more hydrophobic environment compared to the N state. FRET experiments show that the I state has a larger inter-domain distance between domain I and domain II, compared to the N state. Dynamic fluorescence quenching experiments suggest that the inter-domain region in the I state is as dry as the native protein. REES experiments on W214 show that the solvation environment created by the protein matrix and the side-chains around it is like a molten and viscous liquid. A comparison of the thermodynamic stability of the I state and the N state reveals that the inter-domain side-chain packing contributes significantly (35-40%) to the stability of the native protein. Our results indicate that the desolvation of the hydrophobic core and development of side-chain packing are two distinct structural events during protein folding and are not coupled processes as commonly believed. Our results support the hypothesis that the N states of proteins might be an ensemble of loosely and alternatively packed DMG-like states[10, 17-18] which allow enough dynamics and volume fluctuations in the N states required for allosteric functions.[28, 58] In the case of HSA, this mechanism of dynamic allostery might have physiological and functional importance. The main function of HSA in human serum is to bind and transport ligands of different shapes, sizes and chemical nature (hormones, fatty acids, bilirubin etc.). It is possible that the side-chain conformational heterogeneity in the N state ensemble of HSA facilitates fluctuations and inter-conversion between different DMG-like alternatively packed N states required for binding to a multitude of its ligand partners. In this context, our result that the heterogeneity of the N state ensemble is dependent upon pH suggests that the equilibrium between the N and the DMG forms of HSA can be modulated by the changes in the physiological conditions.

2.5 References

1. Richards, F. M. The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* **1974,** *82* (1), 1-14.

2. Chothia, C. Structural invariants in protein folding. *Nature* **1975,** *254* (5498), 304-308.

3. Chothia, C.; Janin, J. Principles of protein-protein recognition. *Nature* **1975,** *256* (5520), 705-708.

4. Richards, F. M. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* **1977,** *6*, 151-176.

5. McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977,** *267* (5612), 585-590.

6. Wuthrich, K.; Wagner, G. Internal Motion in Globular Proteins. *Trends Biochem Sci* **1978,** *3* (10), 227-230.

7. Zhou, Y.; Vitkup, D.; Karplus, M. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J Mol Biol* **1999,** *285* (4), 1371-1375.

8. Jha, S. K.; Udgaonkar, J. B. Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* **2007,** *282* (52), 37479-37491.

9. Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; Garcia-Moreno, B. E.; Royer, C. A. Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci U S A* **2012,** *109* (18), 6945-6950.

10. Bhattacharyya, S.; Varadarajan, R. Packing in molten globules and native states. *Curr Opin Struct Biol* **2013,** *23* (1), 11-21.

11. Bowman, G. R.; Geissler, P. L. Extensive conformational heterogeneity within protein cores. *J Phys Chem B* **2014,** *118* (24), 6417-6423.

12. DuBay, K. H.; Bowman, G. R.; Geissler, P. L. Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Acc Chem Res* **2015,** *48* (4), 1098-1105.

13. Elber, R.; Karplus, M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science* **1987,** *235* (4786), 318-321.

14. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007,** *450* (7172), 964-972.

15. Tzeng, S. R.; Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **2012,** *488* (7410), 236-240.

16. Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* **2010,** *6* (5), 352-358.

17. Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* **2010,** *78* (13), 2725-2737.

18. Baldwin, R. L.; Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* **2013,** *23* (1), 4-10.

19. Neumaier, S.; Kiefhaber, T. Redefining the dry molten globule state of proteins. *J Mol Biol* **2014,** *426* (13), 2520-2528.

20. Shakhnovich, E. I.; Finkelstein, A. V. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* **1989,** *28* (10), 1667-1680.

21. Finkelstein, A. V.; Shakhnovich, E. I. Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* **1989,** *28* (10), 1681-1694.

22. Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* **1995,** *375* (6531), 513-515.

23. Hoeltzli, S. D.; Frieden, C. Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-19F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc Natl Acad Sci U S A* **1995,** *92* (20), 9318-9322.

24. Jha, S. K.; Dhar, D.; Krishnamoorthy, G.; Udgaonkar, J. B. Continuous dissolution of structure during the unfolding of a small protein. *Proc Natl Acad Sci U S A* **2009,** *106* (27), 11113-11118.

25. Jha, S. K.; Udgaonkar, J. B. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* **2009,** *106* (30), 12289-12294.

26. Reiner, A.; Henklein, P.; Kiefhaber, T. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* **2010,** *107* (11), 4955-4960.

27. Sarkar, S. S.; Udgaonkar, J. B.; Krishnamoorthy, G. Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* **2013,** *105* (10), 2392-2402.

28. Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J Am Chem Soc* **2017,** *139* (10), 3599-3602.

29. Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* **2012,** *134* (20), 8543-8550.

30. Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* **2016,** *7* (1), 173-179.

31. Painter, L.; Harding, M. M.; Beeby, P. J. Synthesis and interaction with human serum albumin of the first 3,18-disubstituted derivative of bilirubin. *J Chem Soc, Perkin Trans 1* **1998,** (18), 3041-3044.

32. Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **1986,** *131*, 266-280.

33. Jha, A.; Udgaonkar, J. B.; Krishnamoorthy, G. Characterization of the heterogeneity and specificity of interpolypeptide interactions in amyloid protofibrils by measurement of site-specific fluorescence anisotropy decay kinetics. *J Mol Biol* **2009,** *393* (3), 735-752.

34. Lillo, M. P.; Beechem, J. M.; Szpikowska, B. K.; Sherman, M. A.; Mas, M. T. Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: a steady-state and time-resolved study of yeast phosphoglycerate kinase. *Biochemistry* **1997,** *36* (37), 11261-11272.

35. Saxena, A. M.; Udgaonkar, J. B.; Krishnamoorthy, G. Characterization of intra-molecular distances and site-specific dynamics in chemically unfolded barstar: evidence for denaturant-dependent non-random structure. *J Mol Biol* **2006,** *359* (1), 174-189.

36. Hudson, E. N.; Weber, G. Synthesis and characterization of two fluorescent sulfhydryl reagents. *Biochemistry* **1973,** *12* (21), 4154-4161.

37. Atanasiu, C.; Su, T. J.; Sturrock, S. S.; Dryden, D. T. Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res* **2002,** *30* (18), 3936-3944.

38. Street, T. O.; Courtemanche, N.; Barrick, D. Protein folding and stability using denaturants. *Methods Cell Biol* **2008,** *84*, 295-325.

39. Nath, U.; Udgaonkar, J. B. Perturbation of a tertiary hydrogen bond in barstar by mutagenesis of the sole His residue to Gln leads to accumulation of at least one equilibrium folding intermediate. *Biochemistry* **1995,** *34* (5), 1702-1713.

40. Wani, A. H.; Udgaonkar, J. B. Revealing a concealed intermediate that forms after the rate-limiting step of refolding of the SH3 domain of PI3 kinase. *J Mol Biol* **2009,** *387* (2), 348-362.

41. Lakowicz, J. R. *Principles of fluorescence spectroscopy*. Springer: Singapore, 2006.

42. Amiri, M.; Jankeje, K.; Albani, J. R. Origin of fluorescence lifetimes in human serum albumin. Studies on native and denatured protein. *J. Fluoresc.* **2010,** *20* (3), 651-656.

43. Amiri, M.; Jankeje, K.; Albani, J. R. Characterization of human serum albumin forms with pH. Fluorescence lifetime studies. *J Pharm Biomed Anal* **2010,** *51* (5), 1097-1102.

44. Carter, D. C.; Ho, J. X. Structure of serum albumin. *Adv Protein Chem* **1994,** *45*, 153-203.

45. Dockal, M.; Carter, D. C.; Ruker, F. Conformational transitions of the three recombinant domains of human serum albumin depending on pH. *J Biol Chem* **2000,** *275* (5), 3042-3050.

46. Eftink, M. R.; Ghiron, C. A. Dynamics of a protein matrix revealed by fluorescence quenching. *Proc Natl Acad Sci U S A* **1975,** *72* (9), 3290-3294.

47. Strambini, G. B.; Gonnelli, M. Fluorescence quenching of buried Trp residues by acrylamide does not require penetration of the protein fold. *J Phys Chem B* **2010,** *114* (2), 1089-1093.

48. Demchenko, A. P. Site-selective Red-Edge effects. *Methods Enzymol* **2008,** *450*, 59-78.

49. Mountain, R. D.; Thirumalai, D. Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc* **2003,** *125* (7), 1950-1957.

50. Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci U S A* **2008,** *105* (44), 16928-16933.

51. Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein folding: from theory to practice. *Curr Opin Struct Biol* **2013,** *23* (1), 22-29.

52. Jha, S. K.; Marqusee, S. Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* **2014,** *111* (13), 4856-4861.

53. Dasgupta, A.; Udgaonkar, J. B.; Das, P. Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J Phys Chem B* **2014,** *118* (24), 6380-6392.

54. de Oliveira, G. A.; Silva, J. L. A hypothesis to reconcile the physical and chemical unfolding of proteins. *Proc Natl Acad Sci U S A* **2015,** *112* (21), E2775-E2784.

55. Cohen, S. S.; Riven, I.; Cortajarena, A. L.; De Rosa, L.; D'Andrea, L. D.; Regan, L.; Haran, G. Probing the Molecular Origin of Native-State Flexibility in Repeat Proteins. *J Am Chem Soc* **2015,** *137* (32), 10367-10373.

56. Spudich, G. M.; Miller, E. J.; Marqusee, S. Destabilization of the *Escherichia coli* RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol* **2004,** *335* (2), 609-618.

57. Dasgupta, A.; Udgaonkar, J. B. Four-state folding of a SH3 domain: salt-induced modulation of the stabilities of the intermediates and native state. *Biochemistry* **2012,** *51* (23), 4723-4734.

58. Cooper, A. Thermodynamic fluctuations in protein molecules. *Proc Natl Acad Sci U S A* **1976,** *73* (8), 2740-2741.

Chapter 3.

Slow Motion Protein Dance Visualized Using Red Edge Excitation Shift of a Buried Fluorophore

Reproduced with the permission of Mishra, P., and Jha, S. K. ***(2019)*** *Slow motion protein dance visualized using red-edge excitation shift of a buried fluorophore.* ***J. Phys. Chem. B,*** *123, 6, 1256–1264.*

3.1 Introduction

The extent and degree of conformational heterogeneity in side-chain packing inside the hydrophobic interior of proteins are poorly understood. A body of theoretical work predicts a spectrum of possibilities for the dynamic properties of the protein core including (i) proteins could be like aperiodic solids which lack the periodic regularity of crystals;[1] (ii) proteins may behave like surface-molten solids with liquid-like surface while interior could be solid-like;[2] (iii) the side-chains in the core of proteins could undergo substantial structural fluctuations and the core could be like a dense fluid.[3-4] In contrast, crystal structures of proteins project a static image of the native (N) state, where the interior of proteins are tightly packed with packing densities as high as 0.7-0.8,[5-7] which is very similar to the solid crystals of small organic compounds. Recent studies propose that the N state ensemble contains loosely packed forms termed dry molten globules (DMG)[8-11] in which tight side-chain packing is ruptured but the core remains dehydrated. DMG-like states have expanded and dynamic cores with high conformational entropy and are important for entropy driven protein functions like dynamic allostery.[12-13] Experimentally DMG like states, however, have been majorly observed as kinetic intermediates[14-16] and it has remained elusive to detect and characterize them under equilibrium conditions in the N state ensemble.[8, 11, 17] Thermodynamically, it is possible for proteins to undergo large fluctuations around the mean values of their physical properties like internal energy, enthalpy and molecular volume to give rise to DMG like protein structures because of their small size.[18] However, it has remained extremely challenging to understand the precise molecular and temporal nature of fluctuations and how small and rapid movements of protein side-chains in the hydrophobic core may combine to give an overall change in mean configuration.

There is a scarcity of experimental and spectroscopic methods which can yield dynamic information (on µs to ns timescale) on side-chain packing in the native state ensemble of proteins. NMR methods like dynamic order parameter of methyl side-chains and relaxation dispersion experiments have been employed for a few proteins[19] including calmodulin[20-21] and cyclophilin A[22-23] and side-chain dynamics in ps to µs timescale have been measured. However, NMR methods are limited to small, soluble proteins and cannot be applied to large, multi-domain and aggregation-prone proteins, due to the requirement of millimolar concentrations for NMR studies. Thiol-disulphide exchange is an elegant alternative method that has provided temporal information on the side-chain packing of proteins such as

apmyoglobin[24], barstar[25], monellin[26] and RNAse H.[27] A major disadvantage of this method, however, is that, undesirable side reactions such as disulfide scrambling and air oxidation of thiols to disulfide compete with the thiol-disulfide exchange reaction.

Fluorescence based methods are highly sensitive (require μM quantities of proteins) and are methods of choice for large multi-domain proteins. Red edge excitation shift (REES) studies of a fluorophore buried in the core of a protein can potentially give dynamic information on dipoles created by polar and polarizable side-chains of the hydrophobic core, if they solvate the fluorophore on nano second and slower timescale.[28-29] The phenomenon of REES is marked by a red shift in the wavelength of the maximum fluorescence emission (λ_{max}^{em}), when excited at the red edge of the excitation spectrum of the fluorophore[28, 30-31] (Figure 3.1). In general, the λ_{max}^{em} for a fluorophore is independent of the wavelength of excitation (λ^{ex}).[31] However, in the cases where the environment surrounding the fluorophore is heterogeneous and motionally restricted, this principle is not followed and REES is observed. The molecular basis of REES is governed by the heterogeneity of the fluorophore–solvent interactions and the rate at which the solvent dipoles reorient themselves around the fluorophore in the excited state, as schematically illustrated in Figure 3.1. For the illustration, a model[30, 32] is considered in which there exists a wide distribution of the fluorophore-solvent interaction energy in the ground state and after a vertical transition to the Franck-Codon (F) excited state, the fluorescence emission with a decay rate τ_f, can occur from two distinct solvation states: (i) a completely solvent-relaxed, R state, where reorientation of solvent dipoles around the fluorophores with a time-constant of τ_s is complete, and (ii) a partially solvent-relaxed, R′ state, where solvent-dipole reorientation around the fluorophore has not completely occurred. Because of the heterogeneous distribution of the fluorophore-solvent interaction energy in the ground state, photo-excitation of the ensemble at the center (λ_C) or the red-edge (λ_R) of the absorbance spectrum photoselects a different sub-population of the molecules to the excited state. In the case of a dynamically solvated fluorophore (Figure 3.1A), the rate of the solvent relaxation is faster than the fluorescence timescale ($\tau_s < \tau_f$) which allows the solvent dipoles to realign around the excited fluorophore before the emission. As a result, irrespective of the excitation at the λ_C or the λ_R, the fluorescence emission takes place from the completely solvent-relaxed R state. Thus, the emission spectra becomes independent of the λ^{ex}. On the other hand, if the fluorophore is placed in a heterogeneous and restricted solvation environment, as in the core of a protein, the solvent relaxation time becomes slower than the rate of the fluorescence emission i.e., $\tau_s > \tau_f$ (Figure 3.1B). This does not permit the solvent dipolar realignment before

the fluorescence emission, and in this case, the emission occurs from a partially solvent-relaxed, R′ state. For the low-energy λ_R excitation, the orientations of the solvent dipoles are such that the energy difference between the ground and the R′ state is smaller than that for the high-energy λ_C excitation. As a result, the fluorescence spectra exhibit red spectral shift for the λ_R excitation compared to the λ_C excitation.

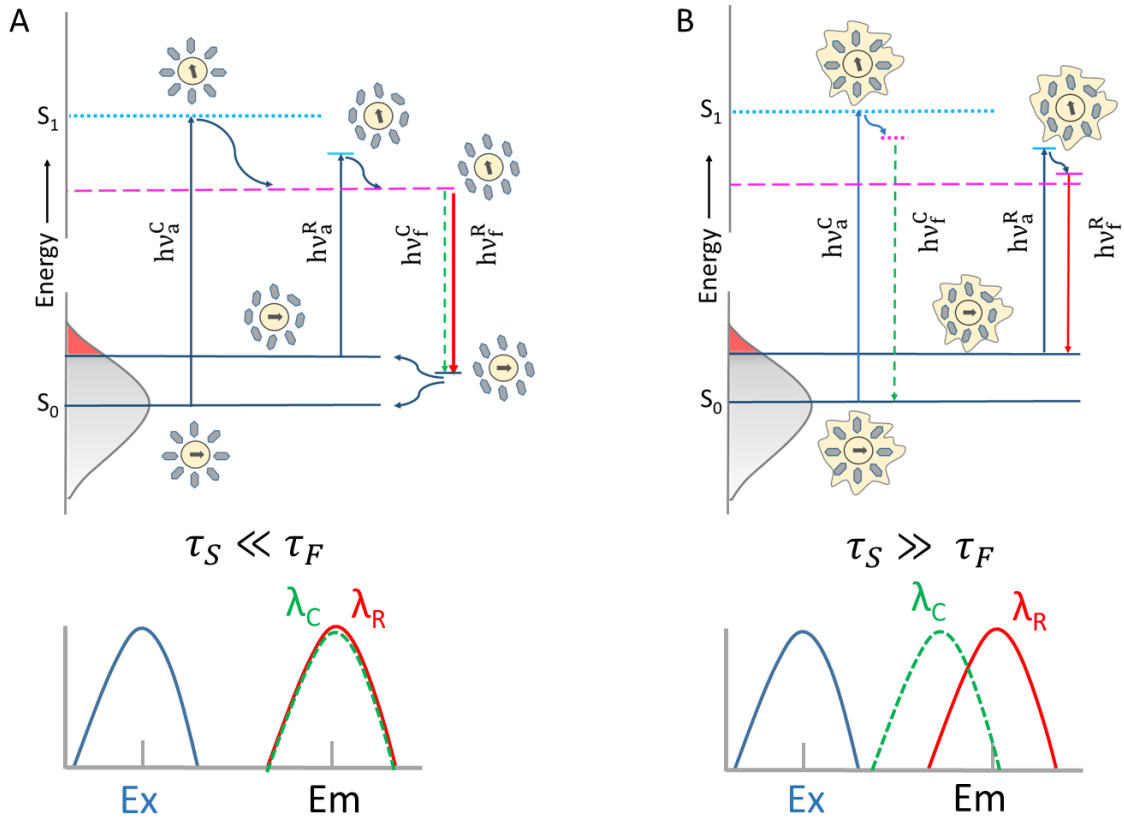


Figure 3.1. *Schematic description of REES. Comparative effects of the central and the red edge excitations (top) on the emission spectra (bottom) of a fluorophore are shown. (A) The timescale of the solvent relaxation, τ_s << the fluorescence timescale, τ_f, (B) τ_s >> τ_f. In both the panels, the inner yellow circle and the thick grey arrow inside it represent the fluorophore and its dipole moment vector, respectively. The grey hexagons and their directions denote the orientations of the solvent molecules. The shaded grey distribution with the red tail represents the distribution of the solute-solvent interaction energy in the ground singlet state (S_0). In the excited singlet state (S_1), the dotted and solid cyan lines show the Franck-Condon (F) states for the center and the red-edge excitations, respectively. The dashed pink lines indicate the complete solvent-relaxed (R) states. The incomplete solvent relaxed states (R') for the center and the red-edge excitations are presented by the dotted and the solid pink lines, respectively. The vertical upwards blue arrows represent the excitation energies (absorbance) and the*

vertically downwards dashed green and solid red arrows represent the fluorescence emission energies. In the bottom panels, the green dashed lines and the solid red lines, respectively, represent the emission spectra when excited at the center (λ_C) and the red-edge (λ_R) of the absorbance spectrum (the solid blue lines). In the panel B (top), the crooked line around the grey hexagons denotes a restricted solvation environment, as inside the core of a protein.

One of the unique features of REES is that it provides information about the heterogeneity and the dynamics of the environment, in contrast to other fluorescence techniques that report about the fluorophore itself. This unique feature of REES makes it a reliable method to study the dense fluidity of the side-chain residues inside the protein core. Proteins with heterogeneous side-chain packing can give rise to an ensemble of DMG-like conformations in equilibrium with the N state.[9] If the polar and buried side-chain residues inside the core of the DMG solvate a suitably placed fluorophore on a timescale much slower than the fluorescence time-scale, REES can provide useful information on the equilibrium conformational states of a protein. REES has been applied mainly to study the solvation dynamics of water molecules in a complex solvation environment.[28-29, 33] The dynamics of interfacial water molecules in model membranes is one such example.[32, 34-35] In the context of proteins, REES has been utilized to understand the conformational changes in the native[30, 36-37], the unfolded (U) [38] and the molten globule [11, 39-40] states. Extrinsic fluorescent probes such as 2-(p-toluidinylnaphthalene)-6-sulfonate (TNS) have been employed to monitor REES in multiple proteins including β-lactoglobulin, β-casein and serum albumins.[30] In the case of erythroid spectrin[38], the application of REES revealed the presence of residual structure in the unfolded state. However, its application to study the dynamics of the protein core is highly limited.[28-29]

In this study, we have utilized REES of a buried fluorophore to study the dynamics and conformational heterogeneity of core packing in the native state ensemble of a multi-domain protein, Human serum albumin (HSA). We observed that the N state of HSA is heterogeneous and in a subpopulation of protein molecules, some of the side-chain residues in the protein core have different orientations and packing than the mean population. Our results indicate that the N state of HSA is an ensemble of DMG-like states.

3.2 Materials and Methods

Spectroscopic methods and instruments

We used a Perkin Elmer lambda 650UV/Vis spectrophotometer and a Perkin Elmer fluorescence spectrometer LS 55, respectively, to measure all the absorption and fluorescence emission spectra. In all the cases, cuvette of path length 1 cm was used. All the circular dichroism (CD) spectra were obtained using a Jasco J-815 CD spectrometer. Here, cuvettes of path length 1 mm and 1 cm were used for far and near-UV CD, respectively. For all the spectra, background signal corrections were done by subtracting the buffer signals.

Chemicals, reagents, buffers, and experimental conditions

HSA (99% pure, fatty acid and globulin free) and urea (ultrapure grade) were procured from Alfa Aesar. 5-((((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was purchased from Life Technologies. Guanidine hydrochloride (GdmCl) was obtained from Sigma. All other chemicals of the highest purity grade were procured from HiMedia and used as received without any further purification. We determined the concentration of HSA by measuring the absorbance at 280 nm and using a molar extinction coefficient[41] of 36500 $M^{-1}cm^{-1}$.

20 mM acetate, 20 mM phosphate and 20 mM Tris-HCl were used as pH 5, pH 7 and pH 8 native buffers, respectively. The unfolding buffers for each pH consisted of 9 M urea in their respective native buffers. We determined the concentrations of urea and GdmCl. from refractive index measurements[42], using an Abbe refractometer from Rajdhani Scientific Instruments Co. (Model: RSR-2).

Preparation of 1, 5-IAEDANS labeled HSA

HSA was labeled with 1,5-IAEDANS at the sole free cysteine, C34, as described earlier[11]. In short, the protein was labeled with 15 fold molar excess of dye in 6 M GdmCl and 20 mM Tris-HCl at pH 8. The reaction mixture was kept in dark for 4 hrs. The protein was then refolded and the labeling reaction was quenched by diluting the reaction mixture into 20 fold volume excess of refolding buffer (20 mM phosphate buffer, pH 7) followed by overnight incubation at 4 ºC. The labeled protein was then separated from the excess free dye by a PD-10 column. The percentage of labeling was estimated as reported previously[11]. The labeled protein (HSA-IAEDANS) was found to be >95% labeled with the dye.

pH dependence of protein structure monitored by fluorescence and CD

For pH dependence studies of protein structure, all the samples were incubated overnight at different pH. For fluorescence measurements, the protein concentration was ~4-5 µM. The aromatic residues were excited at 280 nm and the emission was collected from 310 nm to 420 nm. C34-IAEDANS was excited at 337 nm and the emission spectrum was collected from 350 nm to 550 nm. Slit widths for excitation and emission were kept nominal. The scan speed was 100 nm/min.

For far-UV and near-UV CD measurements, the protein concentrations were 4-5 µM and ~20 µM, respectively. The far-UV CD experiment was monitored in the wavelength range of 190-250 nm, whereas for the near-UV CD experiment, it was 250-310 nm. Each spectrum was averaged over three scans. For each CD spectrum, the data pitch, data integration time, bandwidth and scan speed were 1 nm, 1 s, 2 nm and 100 nm/min, respectively.

Urea-induced equilibrium unfolding transitions monitored by fluorescence

For the urea-induced equilibrium unfolding transitions for HSA and HSA-IAEDANS, the protein samples (~4-5 µM) were incubated overnight in different concentrations of urea at room temperature at pH 7.0. The equilibrium unfolding was monitored by the change in fluorescence of W214 at 340 nm upon excitation at 295 nm. The transitions were analyzed using a two-state, N⇌U model. The data were fitted to the equation 3.1:[11, 43]

$$y_{obs} = \frac{y_N + y_U\, e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 3.1)$$

where, y_{obs} is the observed fluorescence signal; $y_N$and y_U are the signals of the native and unfolded states, respectively; ΔG_{NU} is the free energy of unfolding of N to U. It is assumed that ΔG_{NU} is linearly dependent on denaturant concentration, [D] and is given by :

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \qquad (eq\ 3.2)$$

where, $\Delta G_{NU}^{H_2O}$ and m_{NU} are the standard free energy at 0 M urea and slope of the N⇌U transition, respectively.

Time-resolved fluorescence measurements

An excitation laser source of 370 nm was used to measure the time-resolved decay kinetics of C34-IAEDANS. The emission decays were collected at 480 nm with 10,000 counts

at the peak. An aqueous solution of the milk powder was used to measure the instrument response function (IRF), which was found to be ~ 196 ps. We used the DAS6 analysis software supplied by Horiba Scientific and deconvulated and fitted the fluorescence decays to a sum of two-three exponentials as per the following equation:[31]

$$I(t) = \sum_i \alpha_i \, e^{-\frac{\tau}{\tau_i}} \quad , \qquad i = 1 \; to \; 3 \qquad (eq\ 3.3)$$

In the above equation, I (t) corresponds to the fluorescence intensity at time t and α_i is the amplitude of the i^{th} fluorescence lifetime, τ_i. The mean fluorescence lifetime (τ_m) was calculated using the following equation:

$$\tau_m = \sum \alpha_i \tau_i \qquad (eq\ 3.4)$$

such that, $\sum_i \alpha_i = 1$ and i = 1 to 3.

REES experiments

For C34-IAEDANS, the fluorescence emission spectra were collected by exciting from 337 nm to 410 nm, whereas, for W214, the excitation wavelengths ranged from 295 nm to 305 nm. The protein concentration was kept between ~ 8-15 µM. Nominal slit widths for both excitation and emission were used. The scan speed was 50 nm/min. Each spectrum was averaged over three scans. For all the samples, the buffer signal was corrected for solvent Raman peak. The emission maxima were calculated both by the first derivative as well as the manual inspection of the fluorescence spectrum and were found to be similar. The error bars for the values of emission maximum were estimated from the standard deviations of at least three independent measurements.

3.3 Results and Discussion

3.3.1 C34 is buried inside the protein core and is solvated by polar and polarizable residues

HSA is the most abundant protein found in blood plasma having the major function of transportation of fatty acids, hormones and other compounds through the blood stream. It is an all helical, three-domain protein where each domain is divided into two subdomains (Figure 3.2A). It has 35 cysteine residues distributed over its three domains. Out of these 35 cysteines, 34 form disulfide bonds and the sole free cysteine (C34) is buried in the hydrophobic core of

domain I (Figure 3.2A). From the crystal structure, it was revealed that inside the core of domain I, C34 is solvated by the side-chains of residues A28, L31, Q33, P35, D38, H39, L42, L74, V77, Y84, Y140 and R144 (Figure 3.2B). Out of the 12 amino acid residues solvating the C34 in the protein core, 2 are basic (H39 and R144), 1 is acidic (D38) and 3 are polar but uncharged (Q33, Y84 and Y140) in nature. The percentage burial of all these residues is 95 ± 5; except Q33, P35 and D38, which are around 60% buried. We also calculated the water accessibility of C34 by ASAview[44] and found it to be ~99% buried.

3.3.2 The structure and stability of HSA does not change upon labeling C34 with 1,5-IAEDANS

In order to probe the core dynamics of the protein, we labeled C34 with a highly sensitive fluorophore, 1,5-IAEDANS. It is important to determine that the labeling does not change the structural properties as well as the thermodynamic stability of HSA. We first compared the global secondary structure of the labeled (HSA-IAEDANS) and the unlabeled proteins by far-UV CD spectroscopy (Figure 3.2C). The mean residual ellipticity (MRE) for HSA and HSA-IAEDANS at pH 7 at 222 nm were -18670 deg cm^2 $dmol^{-1}$ and -18262 deg cm^2 $dmol^{-1}$, respectively, which conveyed that both the labeled and the unlabeled HSA have similar secondary structure. We next investigated the effect of labeling on the thermodynamic stability of HSA by urea-induced unfolding experiments (Figure 3.2D). We observed that the standard free energy of unfolding, $\Delta G_{NU}^{H_2O}$, for HSA was 5.17 ± 0.15 kcal mol^{-1} and for HSA-IAEDANS was 5.11 ± 0.18 kcal mol^{-1}. The similar values of $\Delta G_{NU}^{H_2O}$ for both HSA and HSA-IAEDANS indicated that the structural stability of the N state did not change upon labeling. Similarly, the m-value, which is a measure of the solvent-exposed surface area of the protein, remained similar for both HSA (1.15 ± 0.03 kcal mol^{-1} M^{-1}) and HSA-IAEDANS (1.15 ± 0.10 kcal mol^{-1} M^{-1}).

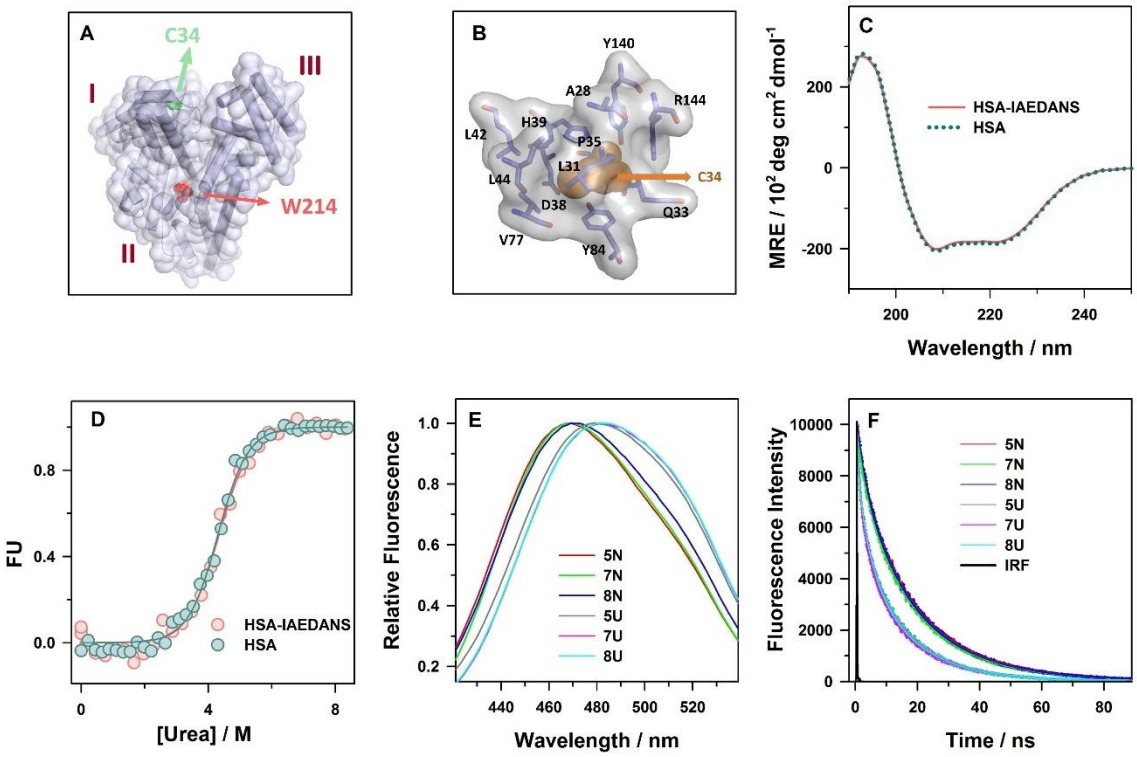


Figure 3.2. *C34 and C34-IAEDANS are buried in the protein structure in the N state and HSA and HSA-IAEDANS have similar structure and stability. (A) The location of C34 and W214 is shown in the crystal structure of HSA (drawn from PDB file 1AO6 using the program PyMOL). In the three-domain structure of the protein, C34 is completely buried in the core of domain I and W214 is partially buried in the inter-domain region. (B) The structure of HSA showing the side-chain residues solvating C34 in the core of domain I. The atoms of C34 are shown in orange spheres. The side-chain of the amino acids solvating C34 are shown in blue sticks. The background light grey surface around the amino acid residues shows the water accessible surface area of the residues burying C34. (C) Mean residue ellipticity (MRE) of HSA and HSA-IAEDANS are shown in the far-UV range in the blue dotted and the solid red lines, respectively. (D) The change in the fraction of unfolded molecules (FU) as a function of [urea] is shown for HSA (blue circles) and HSA-IAEDANS (red circles). The solid blue line through the circles is a fit to a two-state, N⇌U model. (E) The fluorescence emission spectra of HSA-IAEDANS in the N and the U states are shown at different pH (indicated in the figure). Each spectrum was normalized at its maximum intensity. (F) The fluorescence lifetime decay of C34-IAEDANS in the N and the U states at different pH as indicated in the figure are shown (IRF: Instrument Response Function).*

3.3.3 Burial of C34-IAEDANS in the core as revealed by steady-state and time-resolved fluorescence studies

All the above data revealed that the labeling of C34 did not change the global secondary structure and thermodynamic stability of HSA. Hence, C34-IAEDANS can be used as a reporter to probe the core dynamics of the protein in the N state. HSA is known to maintain N-like conformation in the pH range of 4.8 to 8.5[11, 40] which form its N state ensemble. In one of our previous studies[11], we have shown that the N state ensemble of HSA contains a DMG-like intermediate with an alternate side-chain packing arrangement and that the pH modulates the N⇌DMG equilibrium. The population of the DMG increases as the pH of the environment decreases from 7 to 5. In order to further understand the equilibrium dynamics of the N state ensemble, we performed REES experiments at different pH varying from pH 5 to pH 8. It is, however, important to ensure that the probe C34-IAEDANS is buried in the protein core at all the pH. Fluorescence Stokes' shift and fluorescence lifetimes are sensitive measures of the solvation environment of a fluorophore. We, therefore, monitored the λ_{max}^{em} and fluorescence lifetime as a function of pH in the N and urea-unfolded forms at pH 5, pH 7 and pH 8 to probe the burial of C34-IAEDANS. The λ_{max}^{em} of C34-IAEDANS in the solvent exposed U state at all the pH is 483 nm and in the N state is 469 nm (Figure 3.2E). Since around 50% of the residues solvating C34 are polar or charged as revealed by the crystal structure (Figure 3.2B), the blue shift is not expected to be very large as in the case of a complete hydrophobic environment. Nevertheless, the observation that C34-IAEDANS shows a large Stokes' shift in the U state which significantly blue-shifts in the folded state at all the pH indicated that C34-IAEDANS is buried in the hydrophobic core in the N state. Similarly, at all the pH, the N states exhibited a larger fluorescence lifetime (~19 ns) compared to the U states (~15 ns) in the time resolved fluorescence measurements (Figure 3.2F) further supporting the burial of C34-IAEDANS in the pH range 5-8. In previous studies,[40, 45] we have measured the solvent accessibility of C34-IAEDANS in the pH range 2-11, using dynamic fluorescence quenching experiments with acrylamide as the quencher. The quenching experiments suggested that C34-IAEDANS is similarly buried in the protein structure at all the pH measured. The results of the Stokes' shift and the fluorescence lifetime experiments in the current study support our previous observations.

3.3.4 REES as a reporter of heterogeneity and dynamics of the protein core

In order to understand the dynamics of core residues in the N states, we excited the HSA-IAEDANS at different wavelengths from the center to the red edge of the absorbance spectrum of C34-IAEDANS (Figure 3.3A). We observed that λ_{max}^{em} varies as a function of λ^{ex} in the N state ensemble (Figure 3.3B). For example, the representative fluorescence emission spectra at pH 7, as shown in Figure 3.3B, displayed a λ_{max}^{em} of 469 nm when excited at 337 nm. However, upon excitation at 410 nm, λ_{max}^{em} shifts by 6 nm to 475 nm. Interestingly, this effect was not observed in the U state (Figure 3.3B: inset), where the observed λ_{max}^{em} remained at ~483 nm irrespective of any excitation between 337 nm - 410 nm. When we plotted the λ_{max}^{em} as a function of λ^{ex} at different pH, we observed that the N state ensemble showed a typical parabolic curve (Figure 3.3C). When excited from the center up to the red edge (in the wavelength range of 337 nm – 370 nm) of the absorbance spectrum, λ_{max}^{em} remained unchanged. Upon exciting at different wavelengths on the red edge (in the wavelength range of 371 nm – 410 nm), the λ_{max}^{em} steeply shifted towards red. However, this phenomenon was absent in the U states at different pH (Figure 3.3C) where only a very small change in λ_{max}^{em} (1-2 nm) as a function of λ^{ex} was observed (Figure 3.3D). In the U forms, the protein structure is disrupted and C34-IAEDANS is solvated by bulk water molecules (Figure 3.2E and 3.2F). Because, the solvation environment is highly dynamic in this case and the timescale of solvent relaxation (~ps) is faster than the fluorescence timescale (~ns), very little REES is observed (Figure 3.1). In contrast, in the N forms, C34-IAEDANS is buried in the protein core and is solvated by the dipolar side-chains of the protein matrix (Figure 3.2B). The high extent of the REES observed in N forms (Figure 3.3C and 3.3D) indicates that the photo-excitation at the red edge selectively excites a sub-population of protein molecules in which the fluorophore has a different solvation environment than the mean of the population distribution (Figure 3.1), and the side-chains solvating the fluorophore reorient themselves on a timescale slower than the ~19 ns fluorescence lifetime of C34-IAEDANS. As C34-IAEDANS is majorly solvated by the side-chains of buried polar residues (Q33, D38, H39, Y84, Y140, R144) (Figure 3.2B), these results indicate that in the N forms, the orientations of these side-chain residues are different in the two population. Hence, our results suggest that the core packing of HSA is heterogeneous and the side-chains solvating C34-IAEDANS have relatively restricted motions, similar to a dense viscous liquid.

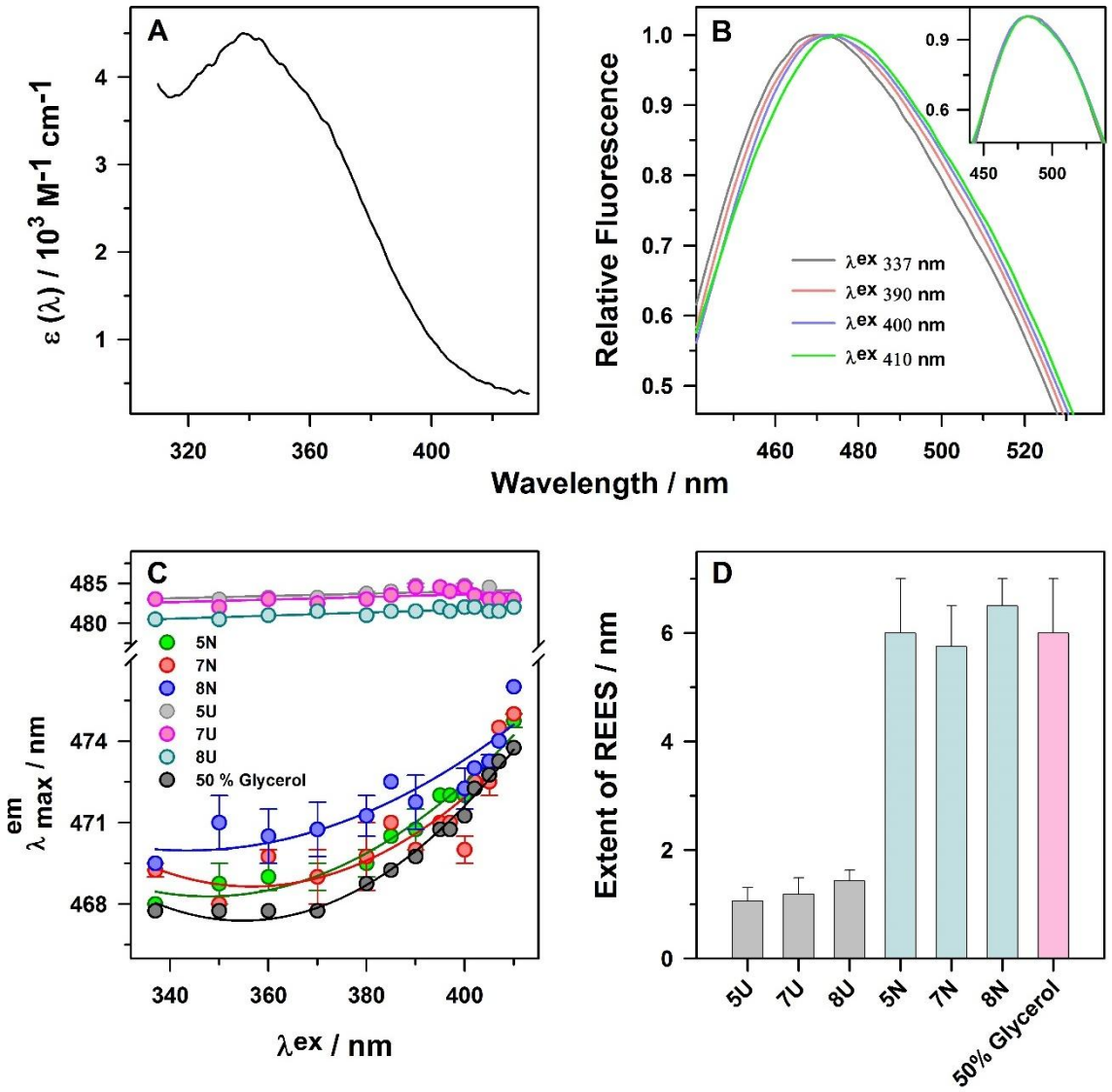


Figure 3.3. *Core dynamics and heterogeneity of HSA as revealed by REES. (A) The molar absorptivity coefficient spectrum, ε (λ) of C34-IAEDANS is shown in panel A. The absorption spectrum has been divided by the molar protein concentration to obtain ε (λ). (B) The fluorescence emission spectra of C34-IAEDANS when excited at different wavelengths in the red edge regions in the N and the U states at pH 7 are shown in panel B and its inset, respectively. (C) Panel C shows the dependence of the wavelength of maximum fluorescence emission (λ_{max}^{em}) on the wavelength of excitation (λ^{ex}) of C34-IAEDANS in the N and the U states at pH 5, pH 7 and pH 8 and in the presence of 50% glycerol at pH 7. (D) The extent of REES as observed in C34-IAEDANS at all the U and N states at pH 5, pH 7 and pH 8 and in the presence of 50% glycerol at pH 7.*

3.3.5 The observed REES of C34-IAEDANS is independent of the viscosity of the exterior solvent

One possible reason for the observed REES of C34-IAEDANS is that it might be measuring the slow dynamics of one or few ordered water molecules around it. This is, however, not the case due to the following two reasons; (i) in the crystal structure of HSA, no water molecules could be seen near C34 which is present in the core; (ii) most importantly, if the REES is due to the interfacial water molecules, then the extent of REES will depend upon the nature and viscosity of the exterior solvent (water in this case), as has been shown previously[34]. We examined this possibility by changing the viscosity and polarity of the exterior solvent with 50 % glycerol/ pH 7 buffer mixture in the N state. We observed that the λ_{max}^{em} as a function of λ^{ex} showed a similar parabolic trend in both the presence and the absence of 50 % glycerol in the external solvent (Figure 3.3C). The extent of REES remained similar (~6 nm) in the absence and the presence of 50% glycerol (Figure 3.3D). This result strongly indicates that the observed REES is due to the constrained environment of the protein matrix and not because of the dynamics of the interfacial water molecules around C34-IAEDANS.

3.3.6 The REES of a partially solvent exposed residue, W214, depends upon the viscosity of the exterior solvent

The sole tryptophan residue, W214, of HSA is located at the interface of domain I and domain II and is partially exposed (~45%) to the exterior solvent as revealed by the crystal structure (Figure 3.2A). In previous studies, we[11, 40] and others[46] have shown that W214 shows a small but distinct amount of REES due to the presence of ordered water molecules around it at the protein-water interface. We took the case of W214 as a positive control to check whether the REES due to interfacial water molecules could be modulated by the change in the viscosity and polarity of the solvent. For W214, in the N state at pH 7, the extent of REES was ~2 nm in response to the excitation from 295 nm – 305 nm (Figure 3.4A and 3.4B). However, upon changing the exterior solvent to 50% glycerol/ pH 7 buffer mixture, the extent of REES dramatically increased to ~6 nm for the same excitation range (Figure 3.4C and 3.4D). These results show that the amount of REES is markedly perturbed by changing the solvent to 50% glycerol/ pH 7 buffer mixture for a fluorophore near the solvent exposed surface of the protein. Hence, the absence of any modulation of the extent of REES in the case of C34-IAEDANS, by increasing the viscosity of exterior solvent, indicate that the solvation environment of C34-IAEDANS is the protein matrix and not water and the observed REES is due to the

heterogeneity in the orientations of the protein side-chains and their slower dipolar relaxation compared to the fluorescence timescale.

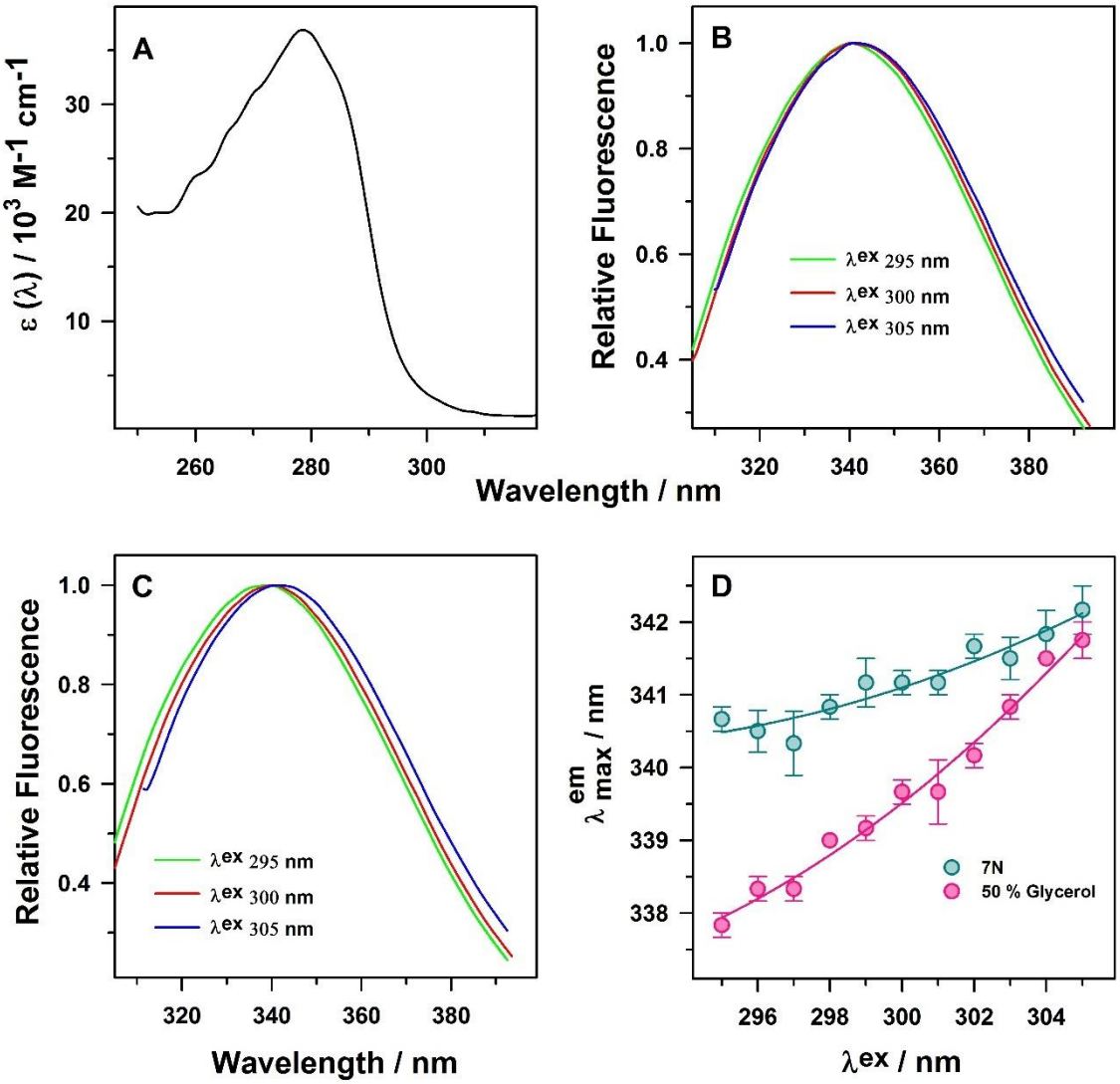


Figure 3.4*. Change in viscosity of exterior solvent modulates the REES of W214. (A) The molar absorptivity coefficient spectrum, ε (λ) of W214. (B) The representative fluorescence emission spectra of W214 at pH 7 when excited at different wavelengths of excitation at the red edge of its absorbance spectrum. (C) Panel C shows the representative fluorescence emission spectra of W214 in the presence of 50% glycerol at pH 7 upon excitation at different wavelengths of the red edge of its absorbance spectrum. For comparison, all the spectra were normalized to 1 at their emission maximum. (D) The dependence of* λ_{max}^{em} *on* λ^{ex} *of W214 in the N state and in the presence of 50% glycerol at pH 7.*

3.3.7 pH as a modulator of heterogeneity of the core

Although the extent of REES for C34-IAEDANS remained the same at all the pH for the N forms (Figure 3.3D), the functional dependence with respect to pH appeared to be marginally different (Figure 3.3C). One possible reason for this could be the heterogeneity in the N state ensemble. As discussed above and reported earlier,[11] pH is known to modulate the equilibrium between different DMG like near-native states which together form the N state ensemble. The changes in structural heterogeneity due to the side-chain packing of neighboring residues around the fluorophores, at different pH, could give rise to this marginally different functional dependence. Another factor that might contribute here is the pKa of neighboring polar and charged residues. The C34-IAEDANS which is the reporter of the protein core is being solvated mostly by the side-chains of charged and polar residues, namely, Q33, D38, H39, Y84, Y140 and R144. These residues might have a different pKa in the buried form in the N state than in the water-exposed unfolded state. D38 and H39 could possibly titrate in the pH range 5 to 8, and that could be another reason for the marginally different trend in REES observed for the N state ensemble at the different pH.

3.3.8 HSA has a compact and highly helical secondary structure but a flexible and loosely packed tertiary structure

Above results indicate that side-chain residues inside the hydrophobic core of HSA are heterogeneously oriented with possibly alternate packing arrangements and that the population distribution might be modulated by pH[11]. We observed that this heterogeneous distribution of side-chain packing does not loosen and expand the N forms enough to allow the penetration of water molecules inside the protein core. HSA has 18 tyrosine residues distributed all over its 3 domains along with the sole tryptophan residue W214. The absorbance spectrum of tryptophan overlaps with the emission spectrum of tyrosine and they form a fluorescence resonance energy transfer (FRET) pair[40]. Figure 3.5A shows the fluorescence spectra of HSA in the N and the U forms in the pH range 5 to 8, after excitation at 280 nm that excites both the tyrosine and tryptophan residues. The U form, at each pH, exhibited similar fluorescence spectra with two peaks with λ_{max}^{em} at 300 nm and 350 nm, corresponding to the emission due to the tyrosine and tryptophan residues, respectively. In contrast, the N forms at each pH exhibited similar fluorescence spectra, with the single peak with λ_{max}^{em} at 340 nm due to the emission of W214 buried in the protein structure. The peak at 300 nm due to the tyrosine residues is absent as the fluorescence emission of tyrosine residues is heavily quenched due to the spatial proximity to

W214 in the N form.[40] These results suggest that at all the pH the structure in the N form is compact. Site-specific FRET experiments in a previous study also showed similar results.[11] The observation that the λ_{max}^{em} of W214 in the N form is similar at each pH and shifted to blue by ~10 nm compared to the U form indicate that the core is not substantially hydrated in the alternatively packed N forms and it behaves like a DMG.

We observed that HSA has a highly helical but a loosely packed three-dimensional structure. Figure 3.5B shows the far-UV CD spectrum of HSA in the N and U forms in the pH range 5 to 8. The N forms at pH each have similar far-UV CD spectra with two minimums at 208 nm and 222 nm. The values of MRE at 208 nm is -22149 ± 913 deg cm^2 $dmol^{-1}$ and at 222 nm is -20331 ± 833 deg cm^2 $dmol^{-1}$ indicating a highly helical structure.[47] The inset in Figure 3.5B shows the near-UV CD spectra of HSA in the N and the U forms in the pH range 5 to 8. HSA absorbs weakly in the near-UV region and its near-UV CD spectrum, at each pH, is similar and shows absorbance bands mainly in the 255 to 270 nm region, due to the asymmetric packing of its 31 phenylalanine residues in the hydrophobic core. The absorbance in 275-295 nm region due to the tyrosine and tryptophan residues is almost negligible. The value of MRE at 261 nm in the N form is -121 ± 3 deg cm^2 $dmol^{-1}$ that is only slightly higher than that in the U form (-72 ± 10 deg cm^2 $dmol^{-1}$). The low values of MRE in the near-UV CD spectrum indicate that the three-dimensional structure of HSA is loosely packed. Hence, these results indicate that HSA has a highly helical structure but a flexible and loose tertiary structure. The loosely packed structure of HSA is likely to give rise to the heterogeneous distribution of side-chain packing indicated by the REES experiments.

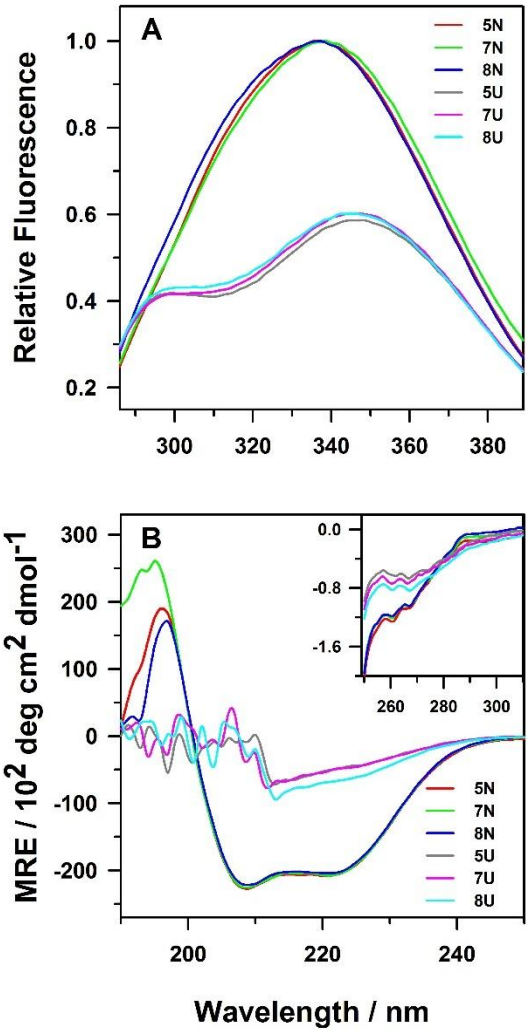


Figure 3.5*: Global structural properties of HSA. (A) Fluorescence emission spectra of the protein at pH 5, pH 7 and pH 8 in both the N and the U forms, when excited at 280 nm are shown. (B) Panel B and its inset show changes in MRE with respect to wavelength as a measure of global secondary and tertiary structures, respectively, in the N and the U forms at pH 5, pH 7 and pH 8.*

3.3.9 The N state ensemble of HSA contains a number of DMG-like states

The N state ensemble of HSA has a flexible structure (Figure 3.5B: inset) and it contains sub-population of protein molecules in which the orientation and packing of some of the side-chain residues in the protein core are different than the mean population (Figure 3.3 and Figure 3.1). These results indicate that the N state ensemble of HSA contains multiple DMG-like states. Although DMG-like states have been originally postulated to be hypothetical transition states,[48-49] they have been observed as partially unfolded intermediate states on the native-side of the free-energy barrier of some proteins, both in simulations[10, 50] and in experiments.[14-16, 40,

[45, 51-54] It has, however, been challenging to detect them in the N-state ensemble. Notably, DMG-like states have been observed in the N-state ensemble of the small protein HP35 by triplet-triplet energy transfer method.[8, 17] Our observation that N state remains in equilibrium with DMG-like conformations with heterogeneous side-chain orientations might be useful for multiple functions of HSA.[12] For example, HSA performs a multitude of functions in which binding and transportation of ligands hold maximum significance. The ligands are of a variety of chemical nature, shape and size. The presence of multiple high energy binding sites could facilitate the binding by shifting the N⇌DMG equilibrium by preferentially stabilizing one particular high energy state with respect to others. We observed that the side-chains solvating C34-IAEDANS relax on a ns or slower timescale. It is possible that many of these small fluctuations of side-chains could couple to give large thermodynamic fluctuations required for conformational conversion and subsequent allosteric and entropy driven functions.[12-13] These results also indicate that the core of HSA is like a molten dense liquid.

3.4 Conclusion

In summary, in this study, we explored the heterogeneity and dynamics of side-chain packing in the hydrophobic core of the native state ensemble of HSA using REES, a spectroscopic method that is well developed to study the dynamics of solvent molecules. Crystal structure and solvent accessibility calculations showed that C34 is buried in protein structure and surrounded by side-chains of polar and charged amino acid residues. We labeled C34 with a highly sensitive fluorescent dye 1,5-IAEDANS. The immediate solvation environment of C34-IAEDANS is the side-chains of the polar and charged residues buried in the protein core that is well supported by fluorescence Stokes' shift and lifetime studies. REES experiments on HSA-IAEDANS revealed that the λ_{max}^{em} varies in a seemingly parabolic fashion as a function of λ^{ex} in the N state ensemble, whereas no such functional dependence was observed in the U forms. The extent of REES for C34-IAEDANS in the N state ensemble was observed to be independent of the viscosity of the exterior solvent. Global structural probes showed that the protein maintains a highly helical structure, but the aromatic side-chain residues in the core are loosely packed and the core is devoid of water molecules. All these observations indicated that the side-chain residues inside the core have a heterogeneous distribution of spatial arrangement in the N state ensemble. These side-chain residues fluctuate much slower than the fluorescence timescale of C34-IAEDANS and hence the core behaves like a molten dense liquid. These results together suggest that the N state of HSA is an ensemble

of DMG-like states. HSA binds to a variety of ligands of different molecular structures and inter-conversion between different DMG-like states in the N state ensemble could be crucial to perform its functions.

3.5 References

1. Schrödinger, E. *What is life? The physical aspect of the living cell*. The University press; The Macmillan company: Cambridge England, 1945.
2. Zhou, Y.; Vitkup, D.; Karplus, M. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J Mol Biol* **1999,** *285* (4), 1371-1375.
3. Bowman, G. R.; Geissler, P. L. Extensive conformational heterogeneity within protein cores. *J Phys Chem B* **2014,** *118* (24), 6417-6423.
4. DuBay, K. H.; Bowman, G. R.; Geissler, P. L. Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Acc Chem Res* **2015,** *48* (4), 1098-1105.
5. Richards, F. M. The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* **1974,** *82* (1), 1-14.
6. Chothia, C. Structural invariants in protein folding. *Nature* **1975,** *254* (5498), 304-308.
7. Richards, F. M. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* **1977,** *6*, 151-176.
8. Reiner, A.; Henklein, P.; Kiefhaber, T. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* **2010,** *107* (11), 4955-4960.
9. Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* **2010,** *78* (13), 2725-2737.
10. Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein folding: from theory to practice. *Curr Opin Struct Biol* **2013,** *23* (1), 22-29.
11. Mishra, P.; Jha, S. K. An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein. *J Phys Chem B* **2017,** *121* (40), 9336-9347.

12. Baldwin, R. L.; Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* **2013,** *23* (1), 4-10.

13. Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J Am Chem Soc* **2017,** *139* (10), 3599-3602.

14. Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* **1995,** *375* (6531), 513-515.

15. Jha, S. K.; Udgaonkar, J. B. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* **2009,** *106* (30), 12289-12294.

16. Sarkar, S. S.; Udgaonkar, J. B.; Krishnamoorthy, G. Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* **2013,** *105* (10), 2392-2402.

17. Neumaier, S.; Kiefhaber, T. Redefining the dry molten globule state of proteins. *J Mol Biol* **2014,** *426* (13), 2520-2528.

18. Cooper, A. Thermodynamic fluctuations in protein molecules. *Proc Natl Acad Sci U S A* **1976,** *73* (8), 2740-2741.

19. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007,** *450* (7172), 964-972.

20. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **2007,** *448* (7151), 325-329.

21. Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* **2010,** *6* (5), 352-358.

22. Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. Enzyme dynamics during catalysis. *Science* **2002,** *295* (5559), 1520-1523.

23. Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005,** *438* (7064), 117-121.

24. Ha, J. H.; Loh, S. N. Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat Struct Biol* **1998,** *5* (8), 730-737.

25. Jha, S. K.; Udgaonkar, J. B. Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* **2007,** *282* (52), 37479-37491.

26. Jha, S. K.; Dasgupta, A.; Malhotra, P.; Udgaonkar, J. B. Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry. *Biochemistry* **2011,** *50* (15), 3062-3074.

27. Bernstein, R.; Schmidt, K. L.; Harbury, P. B.; Marqusee, S. Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proc Natl Acad Sci U S A* **2011,** *108* (26), 10532-10537.

28. Demchenko, A. P. The red-edge effects: 30 years of exploration. *Luminescence* **2002,** *17* (1), 19-42.

29. Chattopadhyay, A.; Haldar, S. Dynamic Insight into Protein Structure Utilizing Red Edge Excitation Shift. *Acc Chem Res* **2014,** *47* (1), 12-19.

30. Demchenko, A. P. On the nanosecond mobility in proteins. Edge excitation fluorescence red shift of protein-bound 2-(p-toluidinylnaphthalene)-6-sulfonate. *Biophys Chem* **1982,** *15* (2), 101-109.

31. Lakowicz, J. R. *Principles of fluorescence spectroscopy*. Springer: Singapore, 2006.

32. Lakowicz, J. R.; Keating-Nakamoto, S. Red-edge excitation of fluorescence and dynamic properties of proteins and membranes. *Biochemistry* **1984,** *23* (13), 3013-3021.

33. Arya, S.; Mukhopadhyay, S. Ordered water within the collapsed globules of an amyloidogenic intrinsically disordered protein. *J Phys Chem B* **2014,** *118* (31), 9191-9198.

34. Chattopadhyay, A.; Mukherjee, S. Red edge excitation shift of a deeply embedded membrane probe: Implications in water penetration in the bilayer. *J Phys Chem B* **1999,** *103* (38), 8180-8185.

35. Haldar, S.; Chaudhuri, A.; Chattopadhyay, A. Organization and Dynamics of Membrane Probes and Proteins Utilizing the Red Edge Excitation Shift. *J Phys Chem B* **2011,** *115* (19), 5693-5706.

36. Guha, S.; Rawat, S. S.; Chattopadhyay, A.; Bhattacharyya, B. Tubulin conformation and dynamics: a red edge excitation shift study. *Biochemistry* **1996,** *35* (41), 13426-13433.

37. Catici, D. A.; Amos, H. E.; Yang, Y.; van den Elsen, J. M.; Pudney, C. R. The red edge excitation shift phenomenon can be used to unmask protein structural ensembles: implications for NEMO-ubiquitin interactions. *FEBS J* **2016,** *283* (12), 2272-2284.

38. Chattopadhyay, A.; Rawat, S. S.; Kelkar, D. A.; Ray, S.; Chakrabarti, A. Organization and dynamics of tryptophan residues in erythroid spectrin: novel structural features of denatured spectrin revealed by the wavelength-selective fluorescence approach. *Protein Sci* **2003,** *12* (11), 2389-2403.

39. Chaudhuri, A.; Haldar, S.; Chattopadhyay, A. Organization and dynamics of tryptophans in the molten globule state of bovine α-lactalbumin utilizing wavelength-selective fluorescence approach: comparisons with native and denatured states. *Biochem Biophys Res Commun* **2010,** *394* (4), 1082-1086.

40. Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* **2016,** *7* (1), 173-179.

41. Painter, L.; Harding, M. M.; Beeby, P. J. Synthesis and interaction with human serum albumin of the first 3,18-disubstituted derivative of bilirubin. *J Chem Soc, Perkin Trans 1* **1998,** (18), 3041-3044.

42. Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **1986,** *131*, 266-280.

43. Street, T. O.; Courtemanche, N.; Barrick, D. Protein folding and stability using denaturants. *Methods Cell Biol* **2008,** *84*, 295-325.

44. Ahmad, S.; Gromiha, M.; Fawareh, H.; Sarai, A. ASAView: Database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics* **2004,** *5*, 51.

45. Acharya, N.; Mishra, P.; Jha, S. K. A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Phys Chem Chem Phys* **2017,** *19* (44), 30207-30216.

46. Demchenko, A. P. Red-edge-excitation fluorescence spectroscopy of single-tryptophan proteins. *Eur Biophys J* **1988,** *16* (2), 121-129.

47. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006,** *1* (6), 2876-2890.

48. Finkelstein, A. V.; Shakhnovich, E. I. Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* **1989,** *28* (10), 1681-1694.

49. Shakhnovich, E. I.; Finkelstein, A. V. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* **1989,** *28* (10), 1667-1680.

50. Mountain, R. D.; Thirumalai, D. Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc* **2003,** *125* (7), 1950-1957.

51. Hoeltzli, S. D.; Frieden, C. Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-19F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc Natl Acad Sci U S A* **1995,** *92* (20), 9318-9322.

52. Rami, B. R.; Udgaonkar, J. B. Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* **2002,** *41* (6), 1710-1716.

53. Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* **2012,** *134* (20), 8543-8550.

54. Jha, S. K.; Marqusee, S. Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* **2014,** *111* (13), 4856-4861.

Chapter 4.

A pH-dependent Protein Stability Switch Coupled to the Perturbed pKa of a Single Ionizable Residue

4.1 Introduction

The role of electrostatics in protein function, solubility and stability has long been recognized.[1-7] The major contributors of electrostatic forces in proteins are the charged states of individual ionizable residues. More often, ionizable amino acid residues are excluded from the interior of the folded proteins because of their intrinsic incompatibility with the hydrophobic environment of the core.[8-12] However, they occasionally get buried inside the protein core to perform certain crucial biological functions.[13-15] This energetically unfavorable process of burial, where ionizable amino acids get sequestered into the less polar protein core from bulk water, allows them to experience different microenvironments, bringing a shift in their pKa. The value of pKa, which gives the information about the equilibrium between the charged and neutral state of an ionizable amino acid, can be modulated by multiple factors. Apart from the dielectric constant of the environment of the ionizable residue, coulombic interactions and hydrogen bonding with nearby polar/ionic residues can affect the pKa of the residue.[16] The interactions that stabilize the buried ionizable residue are known to contribute favourably to protein stability by stabilizing the native state. However, the structural adaptations to facilitate the burial of an ionizable residue inside the core and its effect on protein stability are not fully understood.[17-20]

The stability of a protein is determined by the difference in the free energies of its native (N) and the unfolded (U) states. Therefore, the knowledge of the parameters that influence the stabilities of the N and the U states becomes crucial. pH, being a well-known modulator of electrostatic interaction, can affect the thermodynamic stabilities of the N and the U states in different ways. A buried ionizable amino acid can titrate differently as a function of pH in the folded and the unfolded conformations of the protein.[17, 19, 21] In addition, contrary to the initial notion, recent reports on the U states have emphasized on the presence of nonrandom conformers with local and long range interactions similar or different to their respective native structures.[22-24] These interactions shift the pKa of charged residues in the unfolded protein from its standard value. The interactions that stabilize the charged residue in the U state can contribute unfavourably to the thermodynamic stability of proteins. This disparity in the pKa values of the residue in the N and the U states, therefore, gives rise to a pH dependent protein stability. The site-specific measurement of the pKa values of the buried ionizable residue becomes essential for the detailed understanding of the thermodynamic stability of the protein.

Limited methods have been able to measure the pKa of an ionizable residue in the N and the U states of a protein reliably. Although theoretical studies have succeeded in calculating the pKa of the ionizable residues of a protein, the calculations are yet to attain the experimental accuracy.[25-26] The major complications arise in determining the molecular force field factors due to the heterogeneous environment experienced by the buried residues and generating the precise conformational unfolded state ensemble of the protein.[25-27]

Experimentally pKa can be determined by studying the coupling of thermodynamics of folding and unfolding of a protein to the ionization equilibria of a buried ionizable residue (Figure 4.1). For the illustration, the model protein in its N state is assumed to have an acidic amino acid (A-H) buried inside its hydrophobic core and there exists an equilibrium between the protonated / neutral and deprotonated / charged states of the acidic residue in the folded and unfolded conformations of the protein (Figure 4.1). The four species in equilibrium are represented as N^P, N^D, U^P and U^D in which N and U are native and unfolded conformations, respectively, and the superscripts P and D, respectively, denote the protonated and deprotonated states of the ionizable residue. K_{NU}^P and K_{NU}^D are the equilibrium constants of the equilibrium between the folded and unfolded conformations, respectively, in their protonated and deprotonated states. K_a^N is the acid dissociation constant of $N^P \rightleftharpoons N^D$, while K_a^U are the acid dissociation constant of the equilibrium between U^P and U^D. At any pH, the standard free energy of unfolding, which is a measure of the thermodynamic stability of the protein, $\Delta G_{NU}^{H_2O}$ is given by[17]:

$$\Delta G_{NU}^{H_2O} = \Delta G_{NU}^{H_2O,NE} + \Delta G_{NU}^{H_2O,pH} \qquad (eq\ 4.1)$$

where $\Delta G_{NU}^{H_2O,NE}$ and $\Delta G_{NU}^{H_2O,pH}$ are the non-electrostatic contribution (involving folding / unfolding equilibria) and electrostatic contribution (involving protonation / deprotonation equilibria) to $\Delta G_{NU}^{H_2O}$. The electrostatic contribution to the standard free energy can be calculated from the following equation[17]:

$$\Delta G_{NU}^{H_2O,pH} = -2.303RT\ log\left(\frac{[H^+] + K_a^U}{[H^+] + K_a^N}\right) \qquad (eq\ 4.2)$$

This thermodynamic cycle that links the equilibria between the charged and uncharged states of a protein to its folding / unfolding cycle has been implemented in various cases to determine the pKa of an ionizable residue in the N and the U states of a protein. NMR has been

the method of choice for most of these studies.[17, 19, 24, 28] Although an elegant approach, NMR can only be applied to small and soluble proteins and has limited applications for large, multi-domain and aggregation-prone proteins.

In this work, we have tried to understand the pH dependence of the stability of a multi-domain protein, human serum albumin (HSA). HSA is an all helical protein found abundantly in blood plasma and plays a distinct role in the transportation of metabolites, drugs and ions in the bloodstream and the maintenance of the pH of the plasma and extravascular fluids.[29-30] HSA maintains its native structure in the pH range of 5.0 – 9.0.[29, 31-32] However, it is also known to undergo several pH dependent conformational transitions. The N form of HSA converts to a fast migrating form (F) in the pH range of 5.0 – 3.5 followed by a transition into the acid expanded or the extended (E) form below pH 3.5.[29, 31, 33-34] Above pH 9.0, the N form transforms to a basic (B) form.[29, 31, 35] We explored the pH dependence of the thermodynamic stability of HSA in the N state ensemble, i.e., in the near neutral pH range of 6.0 - 9.0. This is an important pH range because the functional pH of HSA in plasma is ~7.4. The crystal structure of HSA reveals that it consists of three domains and each domain is divided into two sub-domains. A cluster of hydrophobic residues holds together the three domains of HSA in its N state.[32] The sole tryptophan residue (W214) is located in the inter-domain cluster (Figure 4.2A). We used the fluorescence method to understand the coupling of the folding / unfolding cycle of HSA with the ionization of the ionizable residues. We show that the pH dependent changes in the stability of HSA in the pH range of 6.0 – 9.0 are due to perturbed ionization of a single thiol group (C34). C34 is known to impart antioxidant properties to HSA as its reduced form is highly abundant in blood plasma.[36-38] This makes C34 an important target for reactive oxidizing species [37] and the determination of its pKa in the N and the U states of the protein becomes crucial considering its redox potential. We calculated the pKa of C34 in the N and the U states. Our results show that the pKa of the thiol is also perturbed in the U state in addition to the N state indicating that the charged interactions in the U state are also important for the thermodynamic stability of the proteins.

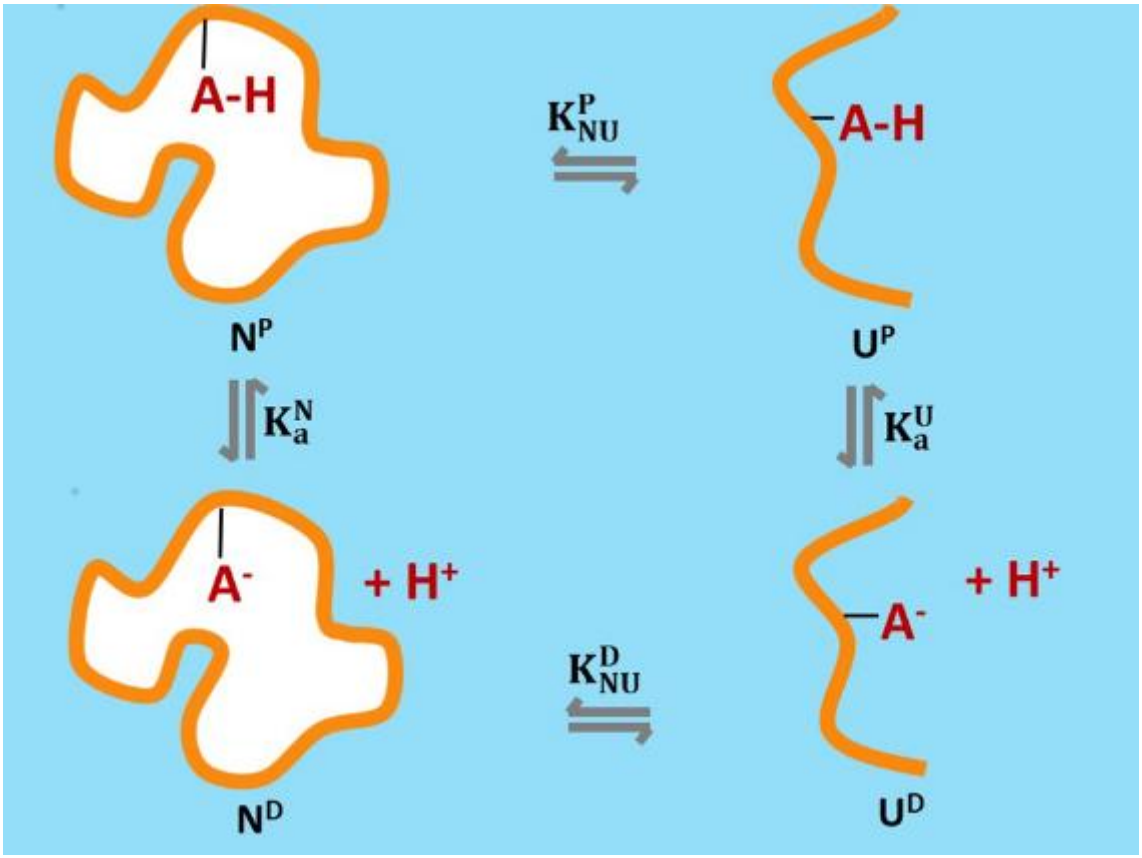


Figure 4.1. *The coupling of the folding / unfolding cycle of proteins with the ionization of an ionizable amino acid residue buried in the protein core. The orange lines in the figure represent the protein backbone. A-H is the protonated acidic ionizable residue in its neutral form buried inside the core (white patch) of the fully folded protein and exposed in its unfolded counterpart.* A^- *refers to the deprotonated / charged state of the acidic residue.* N^P *and* U^P*, respectively, refer to the folded and unfolded state of protein when the ionizable amino acid is protonated and* K^P_{NU} *is the equilibrium constant between them.* N^D *and* U^D *are, respectively, the folded and unfolded conformations of the protein when the ionizable amino acid is deprotonated/charged and* K^D_{NU} *refers to their equilibrium constant.* K^N_a *is the equilibrium acid dissociation constant between the charged and neutral states of the native protein. Similarly, the acid dissociation constant for the equilibrium between the charged and the neutral states of the unfolded protein is represented by* K^U_a*.*

4.2 Materials and Methods

Spectroscopic methods and instruments

Fluorescence experiments were performed on Fluoromax-4 spectrofluorometer from HORIBA Scientific, while absorption studies were done using UV 3200 spectrophotometer procured from LABINDIA Analytical. Quartz cuvettes of path length of 1 cm were used for signal acquisition for both fluorescence and absorption studies. Circular Dichroism (CD) studies were performed on Jasco J-815 CD spectrometer. For far-UV CD experiments, cuvettes of path lengths 1 mm were used. An Abbe refractometer from Rajdhani Scientific Instruments Co. (model: RSR-2) was used to calculate the denaturant concentration by measuring the refractive index.[39]

Chemicals and buffers

HSA (99% pure, fatty acid and globulin free) and urea (ultrapure grade) were sourced from Alfa Aesar. 5-((((2-Iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was purchased from Life Technologies. All other chemicals of the highest purity grade were purchased from HiMedia and used directly without further purification. HSA concentration was measured using an extinction coefficient of 36,500 M^{-1} cm^{-1} by measuring absorbance at 280 nm.[40] For all the pH titration studies, native buffers consisted of 20 mM MES, 20 mM phosphate and 20 mM Tris-HCl for pH 6.0, pH 6.5 - pH 7.5 and pH 8.0 - pH 9.0, respectively. The unfolding buffers constituted of 9 M urea in their respective native buffer.

1,5-IAEDANS labeled HSA preparation

We followed the previously reported protocol for the site-specific covalent labeling of HSA.[32] In brief, 20 fold molar excess of 1,5-IAEDANS was added to HSA, which was previously unfolded in 6 M Guanidine Hydrochloride (GdmCl) and 20 mM Tris-HCl at pH 8.0. The reaction mixture was stirred for 4 hours in the dark followed by refolding in 15 fold volume excess of 20 mM phosphate buffer at pH 7.0 with an overnight incubation at 4 ºC. Refolded protein was then concentrated by centrifugal concentrator (GE healthcare, molecular weight cutoff-30 kDa) to 2.5 mL. To remove the remaining unbound dye and GdmCl, the concentrated reaction mixture was passed through a PD-10 desalting column (GE healthcare). The percentage of labeling in the labeled protein (HSA-IAEDANS) was estimated as discussed previously[32] and was observed to be >95%.

Fluorescence and CD experiments

For fluorescence experiments, the protein concentration used was 4-6 µM. In HSA and HSA-IAEDANS, W214 was excited at 295 nm and the emission spectra were collected from

310 nm -550 nm. 1,5-IAEDANS attached to the C34 (C34-IAEDANS) was excited at 337 nm. The excitation slit widths for all the fluorescence experiments were kept between 1.0 nm to 1.5 nm, while the emission slit widths used were in the range of 8 – 12 nm.

For far-UV and near-UV CD measurements, 3-4 µM and ~20 µM of protein concentrations, respectively, were used. Scans were collected from 190 nm - 250 nm for far-UV and, from 250 nm - 310 nm for near-UV experiments. The data pitch, data integration time, scan speed and bandwidth were kept at 1 nm, 1 s, 50 nm/min and 2 nm, respectively, for all the experiments. Background signal corrections have been done for all the CD and fluorescence experiments by subtracting the buffer signal.

Urea induced equilibrium unfolding experiments

For urea-induced equilibrium unfolding experiments, protein samples (HSA and HSA-IAEDANS) were subjected to a gradient of urea concentration and incubated for 3 hours at room temperature in the pH range of 6.0 – 9.0. For fluorescence measured experiments, W214 and C34-IAEDANS were excited at 295 nm and 337 nm, respectively, and equilibrium unfolding was measured by the change in respective fluorescence signal at 340 nm and 469 nm, respectively. For far-UV CD measured equilibrium unfolding experiments, the signals were measured at 222 nm. The data were fitted to the following equation based on a two-state, N⇌U model.[32, 41]

$$y_{obs} = \frac{y_N + y_U\, e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 4.3)$$

In the above equation, y_{obs} is the observed fluorescence/CD signal; y_N and y_U are the signals of the N and the U states, respectively; ΔG_{NU} is the free energy of unfolding of N⇌U, which has a linear dependency on denaturant concentration and is given by:

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \qquad (eq\ 4.4)$$

where $\Delta G_{NU}^{H_2O}$ is the standard free energy of unfolding in 0 M urea and m_{NU} is the slope of the N⇌U transition.

The fractions of unfolded states, f_U, at any particular denaturant concentration were determined as per the equation mentioned below:

$$f_U = \frac{e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (eq\ 4.5)$$

Analysis of titration of thermodynamic parameters as a function of pH

Assuming that the single titratable group (C34) in HSA has different pKa values in the native (pK_a^N) and the unfolded states (pK_a^U), the dependency of the free energy of unfolding, $\Delta G_{NU}^{H_2O}$ on pH was analyzed using the following equation[17, 20]:

$$\Delta G_{NU}^{H_2O} = \Delta G_{NU}^{H_2O,NE} - 2.303RT\ log\left(\frac{1 + 10^{pK_a^U - pH}}{1 + 10^{pK_a^N - pH}}\right) \qquad (eq\ 4.6)$$

The above equation is a combined and modified form of equations 4.1 and 4.2 (see Section 4.1).

Fluorescence anisotropy experiments

We measured the anisotropy of C34-IAEDANS using FM4-Pol polarization accessory in the L format configuration on Fluoromax-4. The concentration of protein in all the samples was ~10 μM. The C34-IAEDANS was excited at 337 nm and emission was collected at 480 nm. The data were collected for 120 s with an excitation and an emission slit width of 1 nm and 10 nm, respectively. The integration time was 1 s. The fluorescence anisotropy, r, was calculated using the following equation:

$$r = \frac{I_{VV} - I_{VH}G}{I_{VV} + 2I_{VH}G} \qquad (eq\ 4.7)$$

where I is the emission intensity of the fluorophore and the subscripts V (vertical) and H (horizontal) refer to the orientation of polarizers with respect to the excitation beam (first subscript) and the emission beam (second subscript). The instrumental grating factor, G, was the ratio of I_{HV} to I_{HH}.

4.3 Results and Discussion

4.3.1 The N and U states of HSA have distinct spectroscopic properties

We first compared the fluorescence signal of W214 in the N and the U states of HSA at pH 7.0 and we observed that the signal from the U state was highly quenched and the wavelength of the maximum fluorescence emission (λ_{max}^{em}) in the U state (349 nm) was red

shifted as compared to the N state (340 nm), pertaining to the hydration of W214, which lead to the quenching of its fluorescence (Figure 4.2B). Similarly, the global secondary structure content of HSA in the N and U states were distinctly different, as revealed by the far-UV CD spectra with the global secondary structure being entirely lost in the U state (Figure 4.2B, inset). The mean residual ellipticity (MRE) at 222 nm for the N and the U states of HSA were -18,690 and -4,868 deg cm^2 $dmol^{-1}$, respectively. All these results conveyed that the N and the U states of HSA have distinct spectroscopic properties.

4.3.2 Thermodynamic stability of HSA is pH dependent

We utilized the difference in spectroscopic properties of the N and the U states of HSA to monitor their equilibrium population distribution as a function of urea concentration and the thermodynamic stability of HSA over a range of pH (pH 6.0 – pH 9.0). We first used W214 fluorescence as a measure of structure loss upon unfolding during the N⇌U transition. Urea induced single sigmoidal equilibrium unfolding curves for HSA were obtained in the pH range of 6.0 - 9.0. Two representative plots at pH 7.0 and pH 8.0 are shown in Figure 4.2C. We observed that the midpoint of the N⇌U transition (C_m) changed as a function of pH in the pH range of 7.0 - 8.0. We further compared the thermodynamic stability of the secondary structure of HSA as measured by far-UV CD at pH 7.0 and pH 8.0 (Figure 4.2C, inset). The slopy nature of the unfolded baseline in the equilibrium unfolding experiments monitored by far-UV CD limited the correct estimation of the thermodynamic parameters. Nevertheless, we observed that, at pH 7.0, the loss of the protein secondary structure was initiated ~3M [urea] and C_m was ~4.5 M [urea]. However, at pH 8.0, the unfolding of HSA started ~1.5 M [urea] and the observed C_m was ~3 M [urea]. These results were in very well correspondence with the W214 fluorescence measured unfolding transitions (Figure 4.2C). W214 fluorescence is also a measure of the change in the tertiary structure. These results suggested that the global secondary and tertiary structure of the protein dissolved similarly during urea-induced unfolding studies. The unfolding curves as measured by W214 fluorescence were converted into fractions of unfolded protein and were analyzed using a two-state N⇌U model (Figure 4.2D, Section 4.2). The mean value of standard free energy of unfolding of the N⇌U transition, $\Delta G_{NU}^{H_2O}$, at pH 6.5 was 4.8 kcal mol^{-1}, at pH 7.5 was 4.0 kcal mol^{-1} and at pH 8.0 was 3.1 kcal mol^{-1}. The mean value of the change in the solvent accessible surface area (m_{NU}), as measured by the slope of the N⇌U transition, was 1.00 kcal mol^{-1} M^{-1} at pH 6.5, 0.94 kcal mol^{-1} M^{-1} at pH 7.5 and 0.98 kcal mol^{-1} M^{-1} at pH 8.0. These results show that the thermodynamic

stability of HSA is dependent on pH and decreases upon increasing the pH above 6.5. However, the value of m_{NU} remains unperturbed.

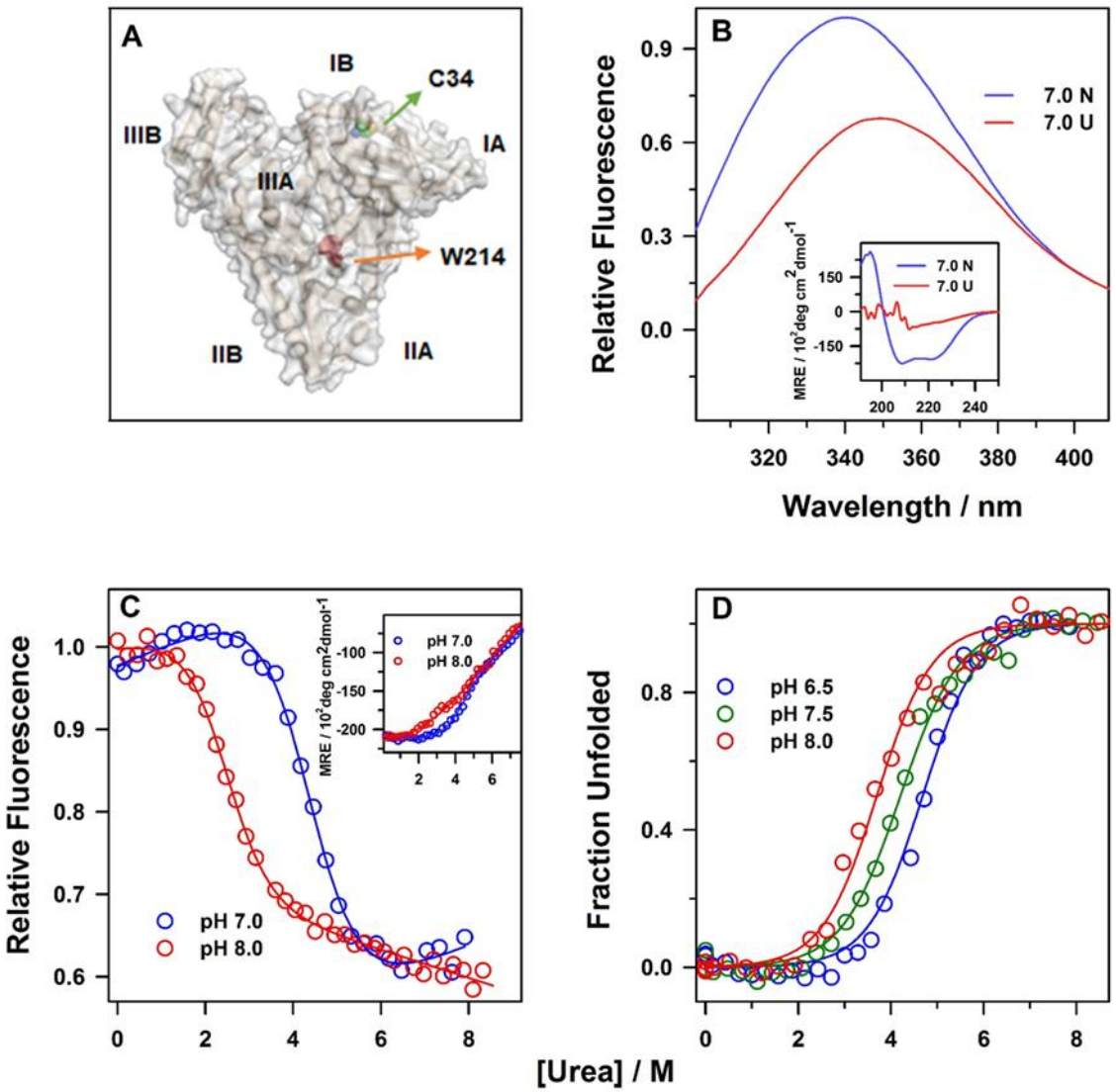


Figure 4.2. *The thermodynamic stability of HSA is pH dependent. (A) Structure of HSA, depicting its six sub-domains along with the locations of free cysteine residue (C34) and single tryptophan residue (W214). This image was drawn from the pdb file 1AO6 using the program PyMOL. (B) Fluorescence emission spectra of W214 in the native (N) and unfolded (U) states of HSA at pH 7.0. The inset of panel B shows the change in the mean residue ellipticity (MRE) of HSA in the N and U states at pH 7.0 in the far-UV region. (C) Urea induced equilibrium unfolding transitions of HSA at pH 7.0 and pH 8.0 as monitored by the changes in the fluorescence signal of W214 at 340 nm and change in the MRE at 222 nm (inset). (D) The fractions of the unfolded protein at pH 6.5, pH 7.5 and pH 8.0 are plotted as a function of the*

[urea]. The solid lines through the data in panel C, inset of panel C and panel D are fits to a two-state, N⇌U model (equation 4.3 and equation 4.5, Section 4.2).

4.3.3 C34-IAEDANS is buried in the protein core

Since pH is a well-known modulator of the structure of proteins and, by extension, their thermodynamic stabilities,[33, 35, 42] it can be argued that the observed dependence of thermodynamic stability of HSA on pH was due to the change in its structural properties. However, it has been previously reported that HSA conserves its N like structure in the pH range of 4.8 to 8.5.[29, 31-33] The similar values of m_{NU} at all the pH (Figure 4.2D), which measures the change in solvent exposed surface area, also indicated that the pH dependence of the stability was not due to a structural change. The other possibility for the destabilization of protein in the pH range of 7.0 to 8.0 could be due to the titration of a buried ionizable amino acid residue inside the core of the protein. Since free cysteine has a pKa ~ 8.2 and upon burial inside a hydrophobic core, its pKa is likely to get modulated, we hypothesized that the titration of a buried cysteine might be leading to the protein destabilization in this pH range. HSA has 35 cysteines, out of which 34 form disulfide linkages in the protein. The sole free cysteine, C34, has its side chain buried in the hydrophobic core of protein (Figure 4.2A).[43] The titration of the protonation-deprotonation equilibrium of cysteine could be blocked in two ways: i) by site-directed mutation of C34 to a non-ionizable residue and ii) test-tube' engineering in which C34 is covalently labeled with a molecule that prohibits its ionization. As we had successfully labeled the C34 and performed multiple experiments with labeled HSA, previously,[32-33, 35, 43] we opted for the second alternative. To block its titration with pH, we labeled C34 with a fluorophore 1,5-IAEDANS (C34-IAEDANS) (Section 4.2). It is important to confirm that labeling does not alter the conformational properties of the protein. We compared the global secondary structural difference between the unlabeled HSA and the labeled HSA (HSA-IAEDANS) by far-UV CD (Figure 4.3A). The MRE for HSA and HSA-IAEDANS at 222 nm were -18,690 deg cm^2 dmol^{-1} and -18,282 deg cm^2 dmol^{-1}, respectively, which implied that the labeling of HSA did not perturb its secondary structural content. We further compared the tertiary structural difference of HSA and HSA-IAEDANS as measured by W214 fluorescence and we observed that λ_{max}^{em} in the N and the U states of HSA did not change upon labeling (Figure 4.2B, Figure 4.3B). In previous global tertiary structural studies, low values of MRE at 261 nm in the N state (-121 ± 3 deg cm^2 dmol^{-1}) were observed which was marginally higher than the MRE value of the U state (-72 ± 10 deg cm^2 dmol^{-1}). [43] This result conveyed that the

global tertiary structure of HSA is flexible and loosely packed. So it is not surprising that the loose packing of the N state of HSA is being able to accommodate 1,5-IAEDANS without disrupting its conformation. In our earlier studies, we have shown that W214 and C34-IAEDANS form a FRET pair.[32-33, 35] Upon excitation of W214 (Donor) at 295 nm, the fluorescence of C34-IAEDANS (Acceptor) increased significantly in the N state with a λ_{max}^{em} of 469 nm due to FRET as compared to the U state (λ_{max}^{em} of 483 nm) (Figure 4.3B). The differences in the fluorescence intensity and λ_{max}^{em} in the N and U states of HSA-IAEDANS implied that the N and the U states have characteristic spectroscopic properties and C34-IAEDANS is buried in the core of the protein in the N state.

4.3.4 Thermodynamic stability of HSA-IAEDANS does not change with pH

In order to investigate the dependence of the thermodynamic stability of HSA-IAEDANS on pH, we performed the urea-induced equilibrium unfolding experiments on HSA-IAEDANS in the pH range of 6.0-9.0. We first monitored the W214 fluorescence as a function of [urea] at different pH and observed that the sigmoidal curves at 340 nm for the N⇌U transition perfectly overlapped on each other (Figure 4.3C). We also checked C34-IAEDANS fluorescence to compare the thermodynamic stability of HSA-IAEDANS (Figure 4.3C, inset). The equilibrium values at 469 nm showed a gradual and non-cooperative dependence on [urea] which limited the correct estimation of the thermodynamic parameters. This observation could be due to the high fluorescence sensitivity of C34-IAEDANS. Nevertheless, the C_m values obtained from W214 and C34-IAEDANS fluorescence measurement studies in the pH range of 6.0-9.0 were analogous. We converted the unfolding curves of HSA-IAEDANS acquired from W214 fluorescence measurements to fractions of unfolded protein (Figure 4.3D) and analyzed them using a two-state N⇌U model. The obtained thermodynamic parameters, both $\Delta G_{NU}^{H_2O}$ and m_{NU} were strikingly similar in the pH range of 6.0-9.0. All these results suggested that the thermodynamic stability of HSA-IAEDANS was pH independent.

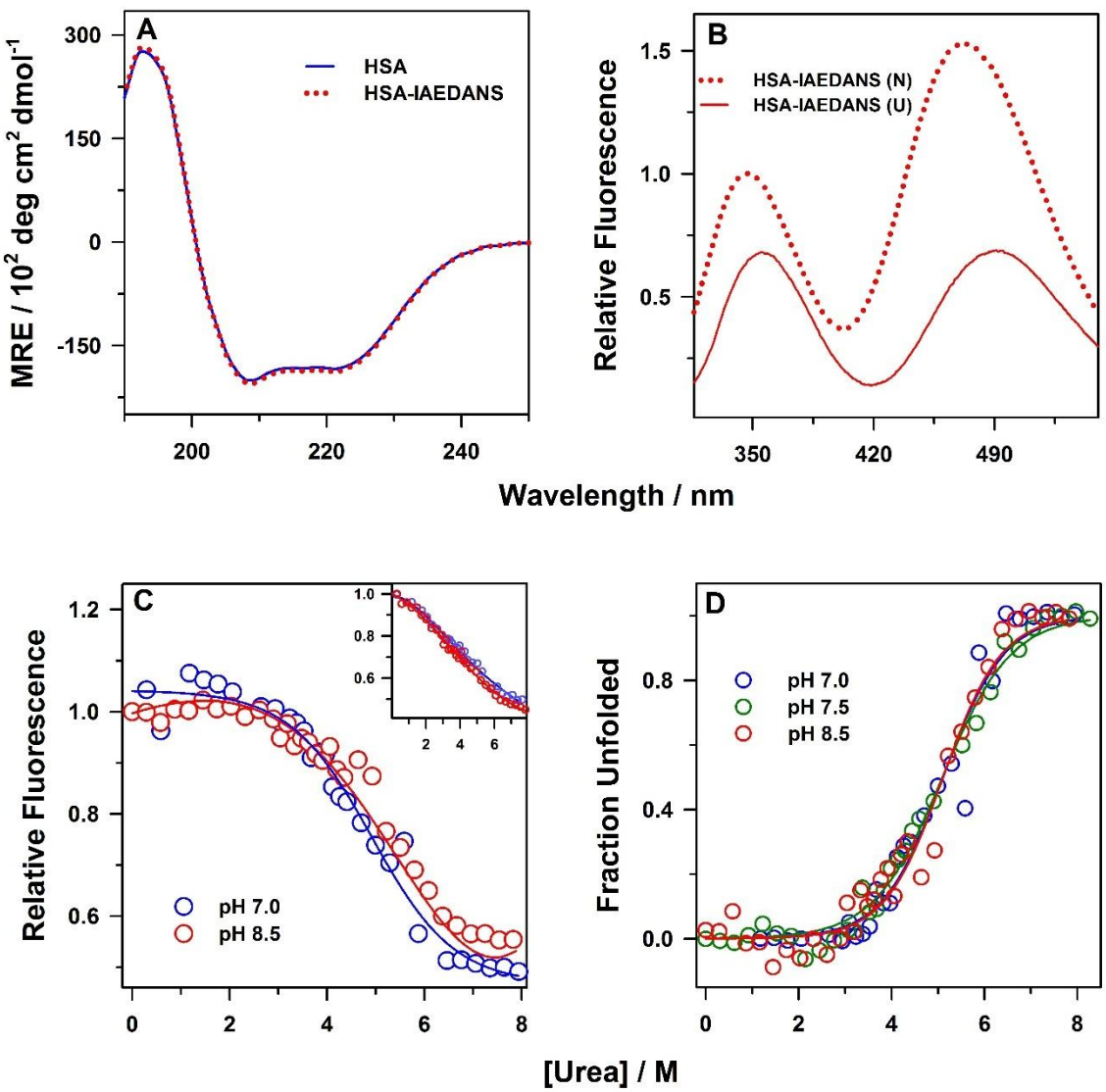


Figure 4.3. *The thermodynamic stability of HSA-IAEDANS is independent of pH. (A) far-UV CD spectra of HSA and HSA-IAEDANS at pH 7.0 as a measure of their global secondary structure content are shown in panel A. (B) Fluorescence emission spectra of W214 and C34-IAEDANS, when excited at 295 nm, in the N and the U states of HSA-IAEDANS at pH 7.0 are shown. (C) Urea induced equilibrium unfolding transitions of HSA-IAEDANS as monitored by the change in W214 fluorescence at 340 nm at pH 7.0 and pH 8.5. The inset shows the pH dependence of the urea-induced equilibrium unfolding transitions as monitored by the change in fluorescence signal of C34-IAEDANS at 469 nm at pH 7.0 and pH 8.5. (D) The fractions of unfolded protein were plotted as a function of [urea] at pH 7.0, pH 7.5 and pH 8.5. The solid lines through the data in panel C, inset of panel C and panel D are fit to a two-state, N⇌U model.*

4.3.5 Titration of C34 was responsible for the destabilization of HSA in the pH range of 6.0-9.0

We compared the change in thermodynamic parameters, $\Delta G_{NU}^{H_2O}$ and m_{NU}, governing the denaturation of HSA and HSA-IAEDANS as a function of pH (Figure 4.4A and Figure 4.4B). Figure 4.4A shows that the dependency of $\Delta G_{NU}^{H_2O}$ on pH in HSA was sigmoidal, while in HSA-IAEDANS, this dependency was abolished. Above pH 6.0, the thermodynamic stability of HSA decreased in a sigmoidal manner and plateaued near pH 8.5 to pH 9.0. The m_{NU} values for HSA and HSA-IAEDANS did not change with pH (Figure 4.4B). All the above results clearly conveyed that the titration of buried C34 was responsible for the pH dependency of $\Delta G_{NU}^{H_2O}$ of HSA. We fitted the pH dependency of $\Delta G_{NU}^{H_2O}$ of HSA (Figure 4.4A) to equation 4.6 to obtain the pKa values of C34 in the N and the U states (Section 4.2). For a cysteine residue buried in the protein core, an increase in its pKa value is expected. For the U state, the anticipated pKa was similar to that of free cysteine in water (8.2). The pKa value obtained for the buried C34 in the N state was 8.7 ± 0.5. Interestingly, in the open U state, the pKa value of C34 was 6.9 ± 0.3.

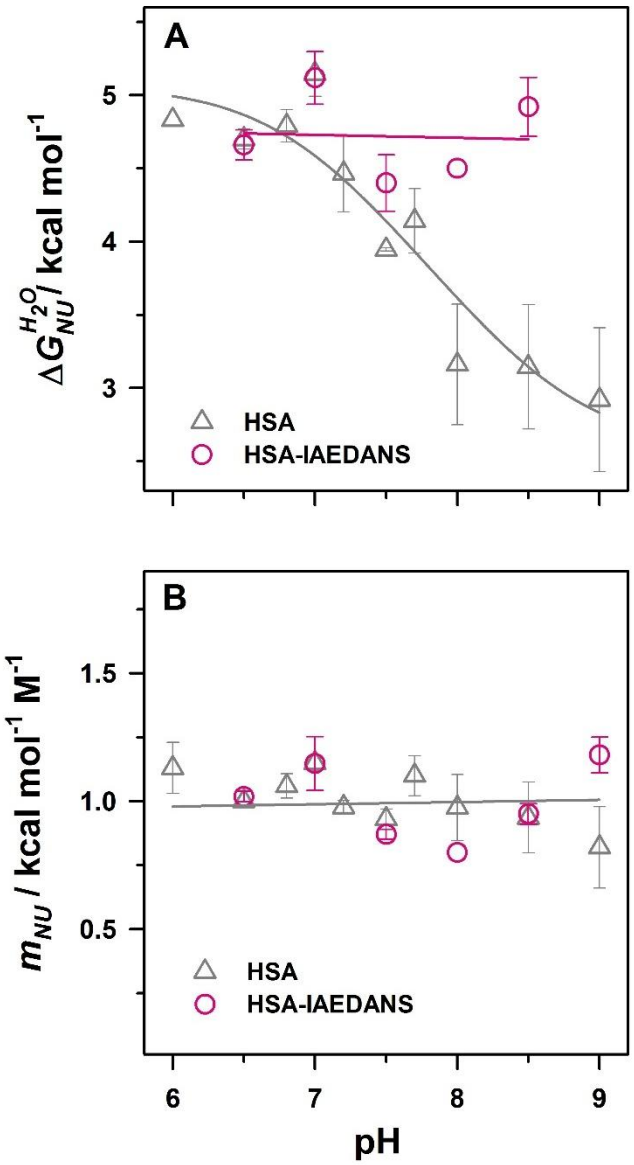


Figure 4.4. *pH dependence of the thermodynamic parameters. (A) Changes in the standard free energy of unfolding ($\Delta G_{NU}^{H_2O}$) of HSA and HSA-IAEDANS as a function of pH. The solid line through the data of HSA is a least-square fit to equation 4.6, while that of HSA-IAEDANS was drawn to guide the eye. (B) Dependence of change in solvent exposed surface area (m_{NU}) during N⇌U transition of HSA and HSA-IAEDANS with respect to pH. Error bars shown in panel A and panel B represent the spread for at least 3 independent sets of experiments.*

The observed decrease in the mean pKa of the C34 in the U state by 1.3 units was quite significant which could be due to the presence of electrostatic interactions near C34 in the U state. However, it is important to confirm that the observed pKa of the thiol group in the N and

the U states were not an artifact of the fitting errors. In order to gauge the reliability of the values of pK_a^N and pK_a^U obtained from the fit, we simulated and compared the graphs of $\Delta G_{NU}^{H_2O}$ as a function of pH for different values of pK_a^N and pK_a^U, using equation 4.6 (Figure 4.5). Figure 4.5A shows that when $pK_a^N = pK_a^U$, no pH dependent change in the values of $\Delta G_{NU}^{H_2O}$ will be observed. Figure 4.5A also shows that when $pK_a^N < pK_a^U$, the value of values of $\Delta G_{NU}^{H_2O}$ will increase with pH and not decrease, as observed experimentally. Hence, for the experimentally observed dependence of $\Delta G_{NU}^{H_2O}$ on pH, pK_a^N must be greater than pK_a^U. Figure 4.5A shows that for the values of pK_a^N = 8.7 and pK_a^U= 6.9, the simulated curve fits the experimental data very well.

We further checked the robustness of the observed values of pK_a^U and pK_a^N by simulating $\Delta G_{NU}^{H_2O}$ *versus* pH plots, keeping one parameter constant and varying the other by about ± 0.5 pH units (Figure 4.5B). The generated $\Delta G_{NU}^{H_2O}$ *versus* pH curves markedly deviated from the experimental data (Figure 4.5B). Collectively these results confirmed that the experimentally obtained values of pK_a^U and pK_a^N of C34 were quite reliable.

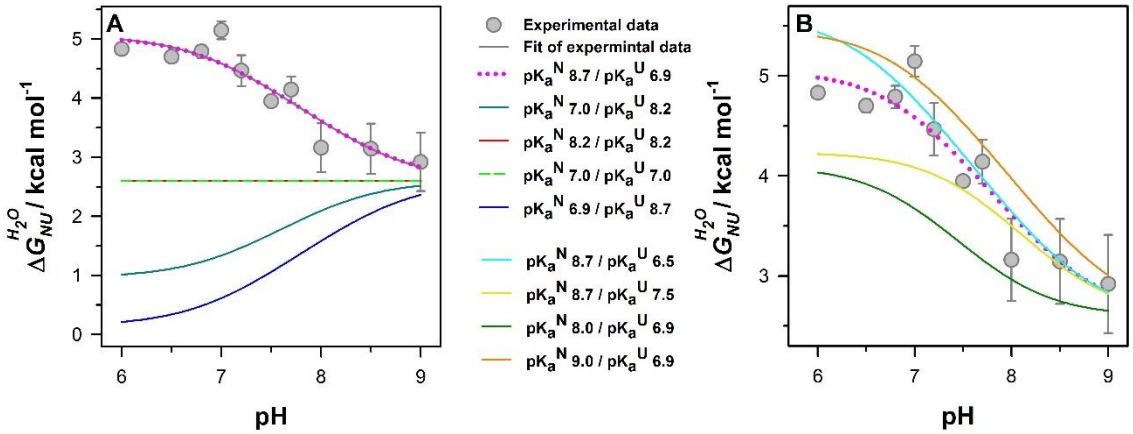


Figure 4.5. *The pKa of C34 in the N state is upshifted, while in the U state, it is downshifted with respect to a free thiol group. (A) The simulated* $\Delta G_{NU}^{H_2O}$ *versus pH curves when i)* pK_a^U *= 6.9 and* pK_a^N *= 8.7 (pink dotted line) and hence are equal to the experimental value (gray filled circles); ii)* $pK_a^U = pK_a^N$ *(red line and green dashed line); and iii)* $pK_a^U > pK_a^N$ *(blue line and dark cyan dashed line). (B) Comparison of simulated* $\Delta G_{NU}^{H_2O}$ *versus pH curves by keeping* pK_a^N *constant while varying* pK_a^U *and vice versa. All the curves were simulated using equation 4.6.*

4.3.6 The U state of HSA has some residual structure

Our results show that the protonation-deprotonation equilibrium of a single ionizable residue, C34, modulates the thermodynamic stability of HSA in the near neutral pH range of 6.0 – 9.0 (Figure 4.4). This result is significant considering the vital role HSA plays in maintaining the pH of the blood plasma, extravascular fluids and ascitic fluid.[30] In the crystal structure of HSA, ~99% of the total surface area of C34 is buried inside the core of domain I of HSA in the N state and is solvated by the side-chains of 12 amino acid residues out of which 2 basic residues (H39 and R144), 1 acidic residue (D38) and 3 polar residues (Q33, Y84 and Y140) directly affect the pKa of C34.[43] The interactions of C34 with H39 and R144 will be stabilizing and will lead to a decrease in its pKa; whereas, D38 will destabilize and increase the pKa of C34. The observed pKa of C34 in the N state of HSA could be a combination of its burial as well as its interactions with adjacent acidic/basic amino acid residues. Such observations have been previously reported for other proteins.[18-20, 44] For HSA, the pKa of C34 in the N state of the protein have been reported to range from 8.0 – 9.0 depending upon the methodologies and experimental conditions used (buffer, ionic strength, temperature etc.).[45-47] The estimated pKa of C34 in the N state of HSA from our study, in its error range (8.7 ± 0.5), is in good agreement with the previously reported pKa values.

Another significant result from the present study is that the mean pKa of the C34 in the U state is ~1.3 units less than the expected pKa of thiol in water (~8.2). We conjectured this could be due to the presence of some polar residues in the primary sequence of HSA near C34, thereby reducing its pKa. Assuming the U state to be completely unfolded and random coil-like structure, we checked for the adjacent amino acid residues of C34 in the sequence. The pentapeptide segment of HSA containing C34 in the middle has other residues as Q32, Q33, P35 and F36. These residues, being neutral in nature, are unlikely to affect the pKa of C34. The only other possibility is that there exists some residual structure in the U state in which some charged residues are interacting with C34, thereby, lowering its pKa. It is important to note that the m_{NU} value of a large protein like HSA is only ~1 kcal mol^{-1} M^{-1} (Figure 4.4B). This result indicates that only a small amount of the buried surface area is exposed to solvent upon unfolding and there exist substantial residual structure in the U state. It appears that the residual structure brings one or more charged residue in the spatial proximity of C34 and hence decreasing its pKa.

HSA has 17 disulfide bonds distributed all over its three domains and they seem to stitch the global structure of HSA. We speculated that these disulfide bonds might be providing the framework for the residual structure in the U state. In order to understand the origin of the residual structure in the U state, we first compared the global tertiary structure of HSA as monitored by near-UV CD spectroscopy in the N and U states at pH 7.0 (Figure 4.6A). The near-UV CD spectrum of HSA shows an absorption band mainly at 255 nm – 270 nm owing to the presence of 31 phenylalanine residues. The MRE value at 261 nm for the N and the U states were -118 deg cm^2 $dmol^{-1}$ and -69 deg cm^2 $dmol^{-1}$, respectively. Interestingly, the MRE value for the U state was only slightly lower than that of the N state. However, upon addition of 1 mM Dithiothreitol (DTT) to the U state of HSA which disrupts the disulfide bonds, the MRE value at 261 nm subsequently decreased to -12 deg cm^2 $dmol^{-1}$, confirming that the presence of disulfide bonds accounts for the residual structure in the U state. We further compared the fluorescence anisotropy values of C34-IAEDANS as a measure of its local dynamics in the N and the U states (Figure 4.6B). The steady-state fluorescence anisotropy values in the N and the U states of C34-IAEDANS were 0.1288 and 0.0636, respectively. Due to its burial in the core, the local dynamics of C34-IAEDANS is highly restricted in the N state, whereas in the U state, the reduced value of anisotropy is due to the higher degree of flexibility available to the side-chains of C34-IAEDANS in comparison to the N state. However, in the presence of DTT, the anisotropy value (0.0296) further decreased by 50 % implying a significant increase in the flexibility and local dynamics of C34-IAEDANS. All these results confirmed that the presence of residual structure in the U state of HSA is a manifestation of the disulfide staples. Due to the residual structure, some local interactions around C34, which may or may not be native-like, are leading to a decrease in its pKa value in the U state.

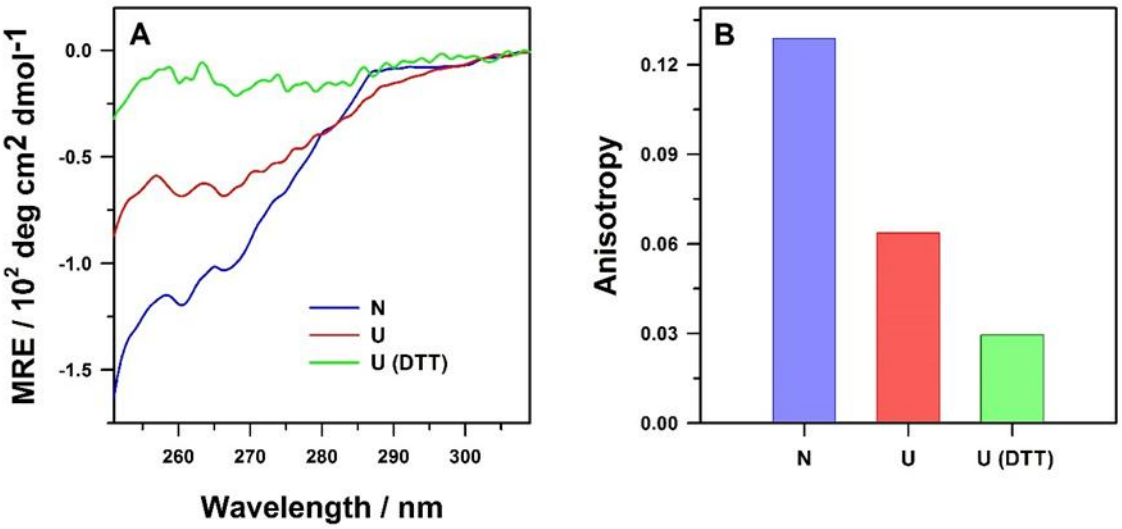


Figure 4.6. *Near-UV CD and anisotropy experiments reveal that the U state of HSA has residual structure. (A) Global tertiary structural characterization of HSA using the near-UV CD in the N state, U state and U state in the presence of 1 mM DTT at pH 7.0 (B) Comparison of fluorescence anisotropy of C34-IAEDANS in the N state, U state and U state in the presence of 1 mM DTT at pH 7.0*

4.3.7 The electrostatic interactions in both the N and the U states are important for the thermodynamic stability of proteins

The thermodynamic stability of proteins is a difference in the free energy of their N and the U states. Therefore, understanding the thermodynamic and structural properties of the U state is also important in addition to that of the N state for decoding the factors that govern the stability of proteins. Generally, the free energy of the U state is considered zero and the difference between the free energies of the N and the U states is taken as the thermodynamic stability of the N state. The initial belief that the unfolded state is a featureless random coil-like structure has been challenged by several recent experimental evidence.[48-54] The unfolded state ensemble has been shown to retain residual secondary and tertiary structure, some of which might be native-like and some might show non-native interactions.[54-57] The extent to which electrostatic interactions stabilize the residual structure in the U state is poorly understood. Due to the hydrophilic nature of the ionizable amino acid residues, it is often believed that the unfolded states of proteins retain minor intramolecular electrostatic interactions. However, multiple reports have demonstrated the presence of both favorable and unfavorable electrostatic interactions in the U states largely affects their energetics.[22-24, 58-61] Nevertheless, the estimation of the pKa of the ionizable residues in the U states has remained

challenging because of their structurally and dynamically wide-ranging nature.[62-63] Our observation that the pH dependent thermodynamic stability of HSA is controlled by the perturbed pKa of a single ionizable group in the N state and the U state and that in the U state the ionizable residue participates in electrostatic interactions due to the residual structure indicate that the electrostatics of the U state is also an important determinant of protein stability. These results, together, will help in a better understanding of the determinants and pathways followed in the folding processes of multi-domain proteins like HSA.

4.4 Conclusion

In this study, we have explored the pH dependence of the thermodynamic stability of the human serum protein, HSA. We used fluorescence spectroscopy to understand the coupling of folding/unfolding of the protein with the ionization of the ionizable residues. We observed that the thermodynamic stability of HSA was pH dependent and changed in a sigmoidal fashion in the near neutral pH range (pH 6.0 – 9.0). In order to understand the cause of this dependency, we labeled a buried ionizable residue C34 with a dye, 1,5-IAEDANS and blocked its titration due to pH. The thermodynamic stability of the labeled protein became pH independent. These results conveyed that the pH dependent changes in the stability of HSA were due to the aberrant titration of the single thiol group. We calculated the pKa of the thiol in the N and U states. In the N state, the mean value of the calculated pKa was ~0.5 units higher than that of the free cysteine in water, whereas, in the U state, there was a significant decrease in the pKa of C34 (by ~1.3 unit). The pKa in the N state could be attributed to the burial of the ionizable thiol group inside the hydrophobic core and its interactions with multiple nearby polar/ionizable residues. However, the perturbation of the pKa of the thiol in the U state indicated some local electrostatic interactions near C34 in the U state. This hypothesis was supported by global tertiary structure measurements by near-UV CD spectroscopy in which the MRE value of the U state of HSA was marginally less than the N state indicating the presence of some tertiary structural elements in the U state. HSA in its native state has 17 disulfide linkages that act as a framework for its structure. We removed the disulfide linkages in the U state by adding 1 mM DTT and observed that the residual structure of the U state was lost upon the breakage of the disulfide bonds. This observation was further supported by steady-state fluorescence anisotropy measurements, in which the anisotropy value of the C34-IAEDANS in the U state in the presence of DTT was considerably small owing to the loss of structure and increase in the dynamics of the C34-IAEDANS. These results together suggest that the electrostatic

interactions regulate the ionization equilibrium of the ionizable residues and therefore affect their pKa values. The knowledge of pKa of these residues in both the N and the U states is essential in understanding the pH dependence of the thermodynamic stability of proteins.

4.5 References

1. Tanford, C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* **1970,** *24*, 1-95.
2. Matthew, J. B. Electrostatic effects in proteins. *Annu Rev Biophys Biophys Chem* **1985,** *14*, 387-417.
3. Allewell, N. M.; Oberoi, H. Electrostatic effects in protein folding, stability, and function. *Methods Enzymol* **1991,** *202*, 3-19.
4. Pace, C. N. Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* **2001,** *40* (2), 310-313.
5. Kumar, S.; Nussinov, R. Close-range electrostatic interactions in proteins. *Chembiochem* **2002,** *3* (7), 604-617.
6. Takano, K.; Scholtz, J. M.; Sacchettini, J. C.; Pace, C. N. The contribution of polar group burial to protein stability is strongly context-dependent. *J Biol Chem* **2003,** *278* (34), 31790-31795.
7. Makhatadze, G. I.; Loladze, V. V.; Ermolenko, D. N.; Chen, X.; Thomas, S. T. Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J Mol Biol* **2003,** *327* (5), 1135-1148.
8. Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. Hydrophobicity of amino acid residues in globular proteins. *Science* **1985,** *229* (4716), 834-838.
9. Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. Interior and surface of monomeric proteins. *J Mol Biol* **1987,** *196* (3), 641-656.
10. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133-7155.
11. Lesser, G. J.; Rose, G. D. Hydrophobicity of amino acid subgroups in proteins. *Proteins* **1990,** *8* (1), 6-13.
12. Lins, L.; Thomas, A.; Brasseur, R. Analysis of accessible surface of residues in proteins. *Protein Sci* **2003,** *12* (7), 1406-1417.

13. Iwata, S.; Ostermeier, C.; Ludwig, B.; Michel, H. Structure at 2.8 A resolution of cytochrome c oxidase from Paracoccus denitrificans. *Nature* **1995,** *376* (6542), 660-669.

14. Luecke, H.; Richter, H. T.; Lanyi, J. K. Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science* **1998,** *280* (5371), 1934-1937.

15. Jiang, Y.; Lee, A.; Chen, J.; Ruta, V.; Cadene, M.; Chait, B. T.; MacKinnon, R. X-ray structure of a voltage-dependent K+ channel. *Nature* **2003,** *423* (6935), 33-41.

16. Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* **2009,** *284* (20), 13285-13289.

17. Tollinger, M.; Crowhurst, K. A.; Kay, L. E.; Forman-Kay, J. D. Site-specific contributions to the pH dependence of protein stability. *Proc Natl Acad Sci U S A* **2003,** *100* (8), 4545-4550.

18. Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Velu, P. D.; Garcia-Moreno, E. B. Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci U S A* **2010,** *107* (37), 16096-16100.

19. Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Garcia-Moreno, B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc Natl Acad Sci U S A* **2011,** *108* (13), 5260-5265.

20. Aghera, N.; Dasgupta, I.; Udgaonkar, J. B. A buried ionizable residue destabilizes the native state and the transition state in the folding of monellin. *Biochemistry* **2012,** *51* (45), 9058-9066.

21. Fitch, C. A.; Whitten, S. T.; Hilser, V. J.; Garcia-Moreno, E. B. Molecular mechanisms of pH-driven conformational transitions of proteins: insights from continuum electrostatics calculations of acid unfolding. *Proteins* **2006,** *63* (1), 113-126.

22. Whitten, S. T.; Garcia-Moreno, E. B. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry* **2000,** *39* (46), 14292-14304.

23. Pace, C. N.; Alston, R. W.; Shaw, K. L. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci* **2000,** *9* (7), 1395-1398.

24. Lindman, S.; Bauer, M. C.; Lund, M.; Diehl, C.; Mulder, F. A.; Akke, M.; Linse, S. pK(a) values for the unfolded state under native conditions explain the pH-dependent stability of PGB1. *Biophys J* **2010,** *99* (10), 3365-3373.

25. Yang, A. S.; Honig, B. On the pH dependence of protein stability. *J Mol Biol* **1993,** *231* (2), 459-474.

26. Makowska, J.; Baginska, K.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A. Acidic-basic properties of three alanine-based peptides containing acidic and basic side chains: comparison between theory and experiment. *Biopolymers* **2008,** *90* (5), 724-732.

27. Schutz, C. N.; Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **2001,** *44*, 400-417.

28. Luisi, D. L.; Snow, C. D.; Lin, J. J.; Hendsch, Z. S.; Tidor, B.; Raleigh, D. P. Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9. *Biochemistry* **2003,** *42* (23), 7050-7060.

29. Carter, D. C.; Ho, J. X. Structure of serum albumin. *Adv Protein Chem* **1994,** *45*, 153-203.

30. Peters, T. U. *All about Albumin: Biochemistry, Genetics, and Medical Applications*. Academic Press: 1996.

31. Dockal, M.; Carter, D. C.; Ruker, F. Conformational transitions of the three recombinant domains of human serum albumin depending on pH. *J Biol Chem* **2000,** *275* (5), 3042-3050.

32. Mishra, P.; Jha, S. K. An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein. *J Phys Chem B* **2017,** *121* (40), 9336-9347.

33. Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* **2016,** *7* (1), 173-179.

34. Olivieri, J. R.; Craievich, A. F. The subdomain structure of human serum albumin in solution under different pH conditions studied by small angle X-ray scattering. *Eur Biophys J* **1995,** *24* (2), 77-84.

35. Acharya, N.; Mishra, P.; Jha, S. K. A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Phys Chem Chem Phys* **2017,** *19* (44), 30207-30216.

36. Steglich, M.; Lombide, R.; López, I. Expression, purification and initial characterization of human serum albumin domain I and its cysteine 34. *Plos One* **2020,** *15* (10), e0240580.

37. Turell, L.; Radi, R.; Alvarez, B. The thiol pool in human plasma: the central contribution of albumin to redox processes. *Free Radic Biol Med* **2013,** *65*, 244-253.

38. Turell, L.; Carballal, S.; Botti, H.; Radi, R.; Alvarez, B. Oxidation of the albumin thiol to sulfenic acid and its implications in the intravascular compartment. *Braz J Med Biol Res* **2009,** *42* (4), 305-311.

39. Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **1986,** *131*, 266-280.

40. Painter, L.; Harding, M. M.; Beeby, P. J. Synthesis and interaction with human serum albumin of the first 3,18-disubstituted derivative of bilirubin. *J Chem Soc Perk T 1* **1998,** (18), 3041-3044.

41. Street, T. O.; Courtemanche, N.; Barrick, D. Protein folding and stability using denaturants. *Methods Cell Biol* **2008,** *84*, 295-325.

42. Khurana, R.; Hate, A. T.; Nath, U.; Udgaonkar, J. B. pH dependence of the stability of barstar to chemical and thermal denaturation. *Protein Sci* **1995,** *4* (6), 1133-1144.

43. Mishra, P.; Jha, S. K. Slow Motion Protein Dance Visualized Using Red-Edge Excitation Shift of a Buried Fluorophore. *J Phys Chem B* **2019,** *123* (6), 1256-1264.

44. Giletto, A.; Pace, C. N. Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1. *Biochemistry* **1999,** *38* (40), 13379-13384.

45. Bocedi, A.; Cattani, G.; Stella, L.; Massoud, R.; Ricci, G. Thiol disulfide exchange reactions in human serum albumin: the apparent paradox of the redox transitions of Cys(34). *FEBS J* **2018,** *285* (17), 3225-3237.

46. Bonanata, J.; Turell, L.; Antmann, L.; Ferrer-Sueta, G.; Botasini, S.; Méndez, E.; Alvarez, B.; Coitiño, E. L. The thiol of human serum albumin: Acidity, microenvironment and mechanistic insights on its oxidation to sulfenic acid. *Free Radic Biol Med* **2017,** *108*, 952-962.

47. Spiga, O.; Summa, D.; Cirri, S.; Bernini, A.; Venditti, V.; De Chiara, M.; Priora, R.; Frosali, S.; Margaritis, A.; Di Giuseppe, D.; Di Simplicio, P.; Niccolai, N. A structurally driven analysis of thiol reactivity in mammalian albumins. *Biopolymers* **2011,** *95* (4), 278-285.

48. Russell, B. S.; Melenkivitz, R.; Bren, K. L. NMR investigation of ferricytochrome c unfolding: detection of an equilibrium unfolding intermediate and residual structure in the denatured state. *Proc Natl Acad Sci U S A* **2000,** *97* (15), 8312-8317.

49. Shortle, D.; Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **2001,** *293* (5529), 487-489.

50. Bu, Z.; Cook, J.; Callaway, D. J. Dynamic regimes and correlated structural dynamics in native and denatured alpha-lactalbumin. *J Mol Biol* **2001,** *312* (4), 865-873.

51. Platt, G. W.; McParland, V. J.; Kalverda, A. P.; Homans, S. W.; Radford, S. E. Dynamics in the unfolded state of beta2-microglobulin studied by NMR. *J Mol Biol* **2005,** *346* (1), 279-294.

52. Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M.; Fersht, A. R. Solution structure of a protein denatured state and folding intermediate. *Nature* **2005,** *437* (7061), 1053-1056.

53. Ratcliff, K.; Marqusee, S. Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic Chlorobium tepidum: comparison with thermophilic and mesophilic homologues. *Biochemistry* **2010,** *49* (25), 5167-5175.

54. Bruun, S. W.; Iesmantavicius, V.; Danielsson, J.; Poulsen, F. M. Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc Natl Acad Sci U S A* **2010,** *107* (30), 13306-13311.

55. Sosnick, T. R.; Trewhella, J. Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* **1992,** *31* (35), 8329-8335.

56. Crowhurst, K. A.; Tollinger, M.; Forman-Kay, J. D. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J Mol Biol* **2002,** *322* (1), 163-178.

57. Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. Long-range interactions within a nonnative protein. *Science* **2002,** *295* (5560), 1719-1722.

58. Guzman-Casado, M.; Parody-Morreale, A.; Robic, S.; Marqusee, S.; Sanchez-Ruiz, J. M. Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of Thermus thermophilus ribonuclease H. *J Mol Biol* **2003,** *329* (4), 731-743.

59. Cho, J. H.; Sato, S.; Raleigh, D. P. Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J Mol Biol* **2004,** *338* (4), 827-837.

60. Bowler, B. E. Thermodynamics of protein denatured states. *Mol Biosyst* **2007,** *3* (2), 88-99.

61. Arbely, E.; Rutherford, T. J.; Neuweiler, H.; Sharpe, T. D.; Ferguson, N.; Fersht, A. R. Carboxyl pK(a) values and acid denaturation of BBL. *J Mol Biol* **2010,** *403* (2), 313-327.

62. Anil, B.; Li, Y.; Cho, J. H.; Raleigh, D. P. The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry* **2006,** *45* (33), 10110-10116.

63. Bradley, J.; O'Meara, F.; Farrell, D.; Nielsen, J. E. Highly perturbed pKa values in the unfolded state of hen egg white lysozyme. *Biophys J* **2012,** *102* (7), 1636-1645.

Chapter 5.

Conclusions and Future Scopes

This chapter reviews the important contributions of the thesis and discusses the future scopes of the studies. Some limitations of the work are also discussed in this chapter.

5.1 Summary

A detailed thermodynamic and structural understanding of the near native states of proteins in the energy landscape has remained a significant challenge. The dynamics among the near native states are assisted by the motion of different structural elements of a protein out of which side-chains of amino acids hold a significant position due to their involvement in various functions such as molecular recognition and dynamic allostery. In this thesis, we have explored the nature, extent and thermodynamic contribution of side-chain packing in the native state ensemble of a multidomain protein, Human Serum Albumin (HSA). The specific aims and conclusions of the study are discussed below.

a) ***Chapter 2:*** In this chapter, we probed the conformational heterogeneity due to side-chain packing in the native state ensemble of HSA at two different native-like conditions. We observed that the native state ensemble of HSA contains an equilibrium intermediate (I) whose inter-domain region bears a resemblance to a DMG. We show that the I state has a native-like secondary structure and intra-domain side-chain packing but the inter-domain region is loosely and alternatively packed. The I state also has a larger inter-domain distance than the native state but has a dry and molten environment at the domain−domain interface. Our results indicate that hydrophobic desolvation and side-chain packing are not concerted processes during protein folding, as believed commonly. Finally, we dissect the energetic contribution of inter-domain side-chain packing interactions and show that they play an important role in protein stability.

b) ***Chapter 3:*** Till date, detection of structural heterogeneity and dynamics due to side-chain packing in the hydrophobic protein core has remained a profound experimental challenge. In this study, we explored the side-chain conformational heterogeneity and dynamics in the core of HSA in its native state ensemble, using red-edge excitation shift experiments. We probed the core of the protein by covalently introducing a fluorophore and observed that that the side-chains of the polar buried residues are heterogeneously packed. We also observed that the side-chains in a subset of the native population solvate the fluorophore on a timescale much slower than the nano second fluorescence

timescale as displayed by the mean population of the native state. Our results conveyed the core of the protein works as a molten dense fluid and the higher entropy conformations of HSA with a dynamic core and alternate side-chain packing arrangements exists in equilibrium with the tightly packed native state which might be necessary for a myriad of entropy-driven functions of this protein.

c) ***Chapter 4:*** Burial of an ionizable amino acid inside the hydrophobic protein core is an energetically unfavorable process that differentially shifts its pKa in the native and the unfolded states of a protein by disturbing its ionization equilibrium. This coupling of the ionization equilibria of the buried ionizable residue with the folding/unfolding cycle of the protein can considerably affect its thermodynamic stability. In this chapter, we studied the coupling of the folding/unfolding cycle with the ionization of a buried ionizable residue in HSA. We observed that the thermodynamic stability of HSA changed as function of pH in the near native pH range (pH 6.0 – 9.0). This pH-dependence of protein stability was arising because of the protonation-deprotonation equilibrium of a buried cysteine residue. We calculated the pKa of the cysteine residue in the native and the unfolded states of the protein. The pKa of the thiol in the native state was upshifted by 0.5 units due to its burial in the protein core. Interestingly, the pKa of the thiol in the unfolded state was downshifted by 1.3 units which could be due to the local residual structure proximal to the thiol group in the unfolded state. Contrary to the initial belief that the unfolded state has a random coil-like structure, our results strongly indicate that the unfolded state can possess some structural elements. Our results also suggest that, in addition to the native state, electrostatic interactions present in the unfolded state are crucial determinants of the thermodynamic stability of proteins.

5.2 Contributions to the Field

The significant contributions to the field of protein folding made by this thesis work are as follows:

a) We have experimentally detected and characterized a native-like intermediate with an alternative arrangement of the side-chain of a residue. This intermediate had the characteristics of a DMG which provided direct evidence that DMG are a subset of native state ensemble.

b) The DMG-like properties can be localized only in the selective regions of the protein structure. In our study, we observed that the inter-domain region had DMG-like characteristics.

c) The equilibrium between native and DMG-like forms, hence, side-chain conformational heterogeneity can be modulated by changes in environmental conditions such as pH.

d) The inter-domain side-chain packing contributes significantly to the thermodynamic stability of the protein which also suggests that van der Waal's interactions, in addition to hydrophobic interactions are the major contributors to the protein stability. Contrary to the common notion, our study shows that hydrophobic desolvation and development of tertiary structure are not concerted steps during protein folding.

e) The core of the proteins have a heterogeneous spatial distribution of side-chains, which is in contrast to the static image projected by the crystal structures.

f) The side-chains in the core fluctuate at a timescale much slower than the fluorescence timescale due to which the core behaves as a dense molten fluid.

g) In addition to the native state, electrostatic interactions present in the unfolded state are important determinants of the thermodynamic stability of proteins.

h) Our results contradict the prevalent notion that unfolded states of proteins are featureless and have random structure and highlight the significance of electrostatic interactions in the unfolded state of a protein.

5.3 Future Directions

The discussed thesis work paves for the following scopes which can be explored in the future.

a) All these studies were performed under in vitro conditions for the simplicity of the system. However, in vivo conditions of the proteins may vary considerably and it would be interesting to compare the results of these studies with that of the studies revisited under in vivo conditions or mimicking the cellular environment such as crowding agents.

b) How chaperones alter the native state ensemble of a protein is not well understood. It will be interesting to explore the molecular mechanism of chaperone action and the subsequent population shift in the near native states, which would provide new evidence for the contribution of side-chain packing in the native state ensemble.

c) Since HSA is known to perform a multitude of functions, study of the site-specific conformational change at the active site during a specific function and its subsequent results in dynamic allostery can be explored in future studies.

ABSTRACT

Name of the Student: Prajna Mishra **Registration No.: 10CC15J26011**
Faculty of Study: Chemical Sciences **Year of Submission: 2021**
CSIR Lab: NCL, Pune **Name of the Supervisor: Dr. Santosh Kumar Jha**

Title of the thesis: Thermodynamic and Spectroscopic Characterization of Near Native States in Protein Folding

---

Proteins perform multiple functions by dynamically interconverting between various near native states in the heterogeneous native state basin of the energy landscape. In order to understand the structure-function paradigm of proteins, a detailed thermodynamic and structural characterization of these near native states becomes essential. The structural heterogeneity among the near native states is due to the dynamics of various structural elements of the protein, of which side-chain packing of amino acids contributes significantly towards the thermodynamic stability and functions of proteins. However, it has remained extremely challenging to understand the nature, extent and the thermodynamic contribution of side-chain packing in the near native states of proteins. In this thesis, we have probed the side-chain conformational heterogeneity in the native state ensemble of a multi-domain protein, Human serum albumin (HSA), in different native-like conditions. We observed that in a subgroup of the native population of HSA, the side-chains at the inter-domain region were packed loosely and alternatively than the native population resembling the characteristics of dry molten globules (DMG). We quantified the contribution of inter-domain side-chain packing to protein stability which was found to be substantial. We further probed the structural heterogeneity and dynamics due to side-chain packing in the core of HSA using red-edge excitation shift experiments. The side-chains of the polar buried residues were observed to be heterogeneously packed and the timescale of their dynamics was much slower than the fluorescence timescale. The burial of a charged side-chain in the core of a protein can considerably affect its thermodynamic stability. In HSA, we observed that its thermodynamic stability was pH-dependent in the near-native pH range (pH 6.0 - pH 9.0). This pH-dependency was attributed to the coupling of the ionization equilibrium of a buried thiol residue with the folding/unfolding cycle of HSA. We estimated the pKa of the thiol in the native state of HSA which was upshifted due to its burial in the core. Interestingly, the pKa of the thiol in the unfolded state was downshifted, which could be due to the local residual structure proximal to the thiol group in the unfolded state. Our results suggest that, in addition to the native state, electrostatic interactions present in the unfolded state are crucial determinants of the protein stability.

List of Publications

1. Nirbhik Acharya, **Prajna Mishra**, and Santosh Kumar Jha, 2016. Evidence for dry molten globule-like domains in the pH-induced equilibrium folding intermediate of a multidomain protein. *The journal of physical chemistry letters*, *7*(1), 173-179. **(equal contribution)**
2. **Prajna Mishra**, and Santosh Kumar Jha, 2017. An alternatively packed dry molten globule-like intermediate in the native state ensemble of a multidomain protein. *The journal of physical chemistry B, 121*(40), 9336–9347.
3. Nirbhik Acharya, **Prajna Mishra**, and Santosh Kumar Jha, 2017. A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Physical chemistry chemical physics*, *19*(44), 30207-30216.
4. **Prajna Mishra**, and Santosh Kumar Jha, 2019. Slow motion protein dance visualized using red-edge excitation shift of a buried fluorophore. *The journal of physical chemistry B, 123*(6), 1256–1264.
5. **Prajna Mishra**, Divya Patni, and Santosh Kumar Jha, 2021. A pH-dependent protein stability switch coupled to the perturbed pKa of a single ionisable residue. *Biophysical chemistry, 274*, 106591.

List of Papers at national/international conferences/seminars

1. Presented poster in “International Symposium on Protein Folding and Dynamics” at NCBS, Bangalore, India. November 08-11, **2016**
2. Presented poster in “Annual Meeting of the Indian Biophysical Society” at IISER, Pune, India. March 09-11, **2018**

Publications

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

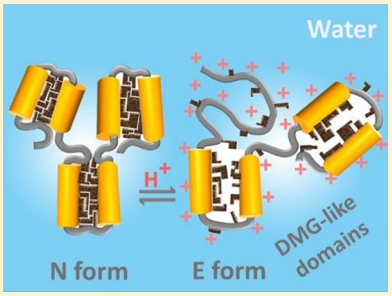
Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein

Nirbhik Acharya, Prajna Mishra, and Santosh Kumar Jha*

Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India

Supporting Information

ABSTRACT: The role of van der Waals (vdW) packing interactions compared to the hydrophobic effect in stabilizing the functional structure of proteins is poorly understood. Here we show, using fluorescence resonance energy transfer, dynamic fluorescence quenching, red-edge excitation shift, and near- and far-UV circular dichroism, that the pH-induced structural perturbation of a multidomain protein leads to the formation of a state in which two out of the three domains have characteristics of dry molten globules, that is, the domains are expanded compared to the native protein with disrupted packing interactions but have dry cores. We quantitatively estimate the energetic contribution of vdW interactions and show that they play an important role in the stability of the native state and cooperativity of its structural transition, in addition to the hydrophobic effect. Our results also indicate that during the pH-induced unfolding, side-chain unlocking and hydrophobic solvation occur in two distinct steps and not in a concerted manner, as commonly believed.



The nature of physicochemical forces that stabilize the functional native structure of proteins is poorly understood.[1−12] It is commonly believed that seclusion of hydrophobic amino acids from water (hydrophobic effect) is the major contributor to the protein stability.[1−4] The hydrophobic amino acids are also tightly packed inside the protein core, similar to the crystals of small organic molecules,[13] in order to maximize the strength of the van der Waals (vdW) interactions. However, the role and energetic contribution of vdW interactions relative to hydrophobic effect in stabilizing protein molecules is not yet clear.[5−8,11,12]

Wet molten globules (WMG) have been observed as equilibrium and kinetic intermediates during the folding of many proteins.[14−20] WMG possess substantial secondary structure and a fluctuating tertiary structure with perturbed side-chain packing and water-solvated hydrophobic core. Theoretical studies of the thermodynamics of protein unfolding predict that a side-chain unlocking step, resulting in the formation of dry molten globules (DMG), precedes hydrophobic solvation during protein unfolding.[21,22] According to these theoretical studies, DMG are expanded forms of the native protein in which tight side-chain packing interactions are ruptured, but the water molecules have not penetrated the hydrophobic core. In contrast to WMG, experimental evidence for DMG have come mainly from kinetic studies in which they have been observed as transient initial intermediates during protein unfolding.[23−29] It is important to identify them under equilibrium conditions where detailed structural characterization using high-resolution probes can be done and thermodynamic contribution of hydrophobic desolvation and side-chain locking in protein stability can be dissected. However, until date DMG have been reported at equilibrium in limited cases and involve only small single domain proteins.[19,30,31]

Various fundamental questions related to DMG remain unanswered.[11,27,32] Are DMG universal protein folding intermediates and form during all kinds of unfolding reactions? Is the core of DMG solid-like as in the native protein or like a molten liquid? What is the protection factor of DMG? When does the hydrophobic solvation occur? Are side-chain packing interactions disrupted uniformly or only in selective regions of the protein structure? In DMG, are all regions of the hydrophobic core of the protein dry or is the dryness only in patches? It becomes more important to determine the answers to these questions for multidomain proteins because for large proteins there exists a possibility that some regions have characteristics of DMG whereas others behave like WMG. However, to the best of our knowledge, DMG-like domains have not been identified in any multidomain protein until date.

In this study, we have used a battery of spectroscopic probes including near- and far-UV circular dichroism (CD), fluorescence resonance energy transfer (FRET), dynamic fluorescence quenching, and red-edge excitation shift (REES) to dissect the pH-induced structural perturbation of a well-studied multidomain protein, human serum albumin (HSA) (Figure 1A), into distinct structural events characterized by changes in secondary and tertiary structure, protein expansion, side-chain unlocking, hydrophobic solvation, and solvation dynamics of the protein matrix. The structure of HSA consists

Received: November 13, 2015
Accepted: December 23, 2015
Published: December 23, 2015

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

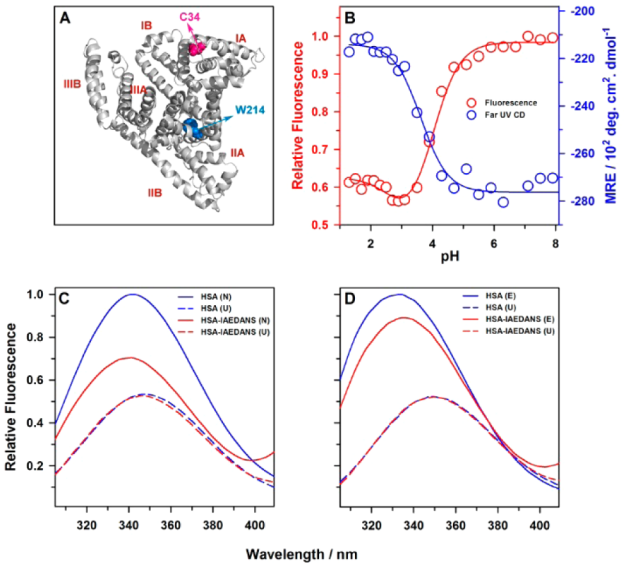


Figure 1. pH dependence of conformation of HSA. (A) Structure of HSA drawn from PDB file 1AO6 using the program PyMOL. (B) Change in fluorescence emission of W214 at 340 nm and the mean residue ellipticity at 222 nm are plotted against pH, according to the left and the right *y* axis, respectively. The blue and red lines through the data are fits to a 2-state and a 3-state pH titration model, respectively (eqs S1 and S2, SI). pH-induced expansion of HSA monitored by FRET at (C) pH 7.0 and (D) pH 2.2. Fluorescence emission spectra of HSA and HSA-IAEDANS are shown in the absence (N form and E form) and the presence of 9 M urea (U form).

of three domains—I, II, and III—and each of them is divided into two subdomains, A and B (Figure 1A). It has been demonstrated earlier that in the low pH forms of HSA, domain III remains fully unfolded, but there are only subtle structural changes in domain I and II.[33] Here, we report that decrease in pH of the solvent leads to the formation of a state, in which domain I and domain II of the protein have characteristics of DMG. We discuss the implications of this observation for protein stability and the mechanism of pH-induced protein denaturation.

We observed that the fluorescence of the sole tryptophan residue, W214, of HSA changes in two distinct sigmoidal steps on decreasing the pH from 7.0 to 1.0 (Figure 1B), indicating that the protein undergoes two structural transitions, coupled to the protonation of two ionizable groups [see the Supporting Information (SI)]. The midpoints of the two transitions are estimated to be at pH 3.7 and pH 2.7. It has been reported that the native form (N) of HSA transforms into a fast migrating form (F) between pH 5.0 and 3.5 (N ⇌ F transition), and the F form converts into an acid-expanded or the extended (E) form below pH 3.5 (F ⇌ E transition).[33−36] Our biphasic pH titration data monitored by W214 fluorescence supports this conclusion. The pH titration of the protein was also monitored by far-UV CD signal at 222 nm (Figure 1B). We observed that there is no apparent change in the secondary structure during F ⇌ E transition and the change in secondary structure from the N to either F or E form occurs in a single sigmoidal step, with the midpoint of transition at pH 3.6.

We explored whether the formation of the E form (at pH 2.2) from N form (at pH 7.0) is accompanied by any expansion in the dimension of the protein using FRET. W214, which is located in a helical segment of domain II, served as the donor (D) fluorophore. There are 35 cysteine residues in HSA, 34 of which form 17 disulfide bridges. The sole free thiol moiety, C34, located at the N-terminal of helix 3 in domain I, was labeled with the fluorescent dye 5-((((2-iodoacetyl)amino)-ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS), which served as the acceptor (A) fluorophore. We observed that C34 can be quantitatively labeled with 1,5-IAEDANS (SI, Figure S1), as has been observed previously for other fluorescence dyes,[37] and that the secondary structure and thermodynamic stability of the unlabeled (HSA) and the 1,5-IAEDANS labeled (HSA-IAEDANS) proteins are similar (SI, Figure S2). The absorbance spectrum of C34-IAEDANS overlaps with the emission spectrum of W214 (SI, Figure S3) (forming a FRET pair), and this pair has been used previously to monitor the change in distances during folding and unfolding of proteins.[38] The fluorescence of W214 is quenched significantly in the N form of HSA-IAEDANS (Figure 1C) because of the spatial proximity with C34-IAEDANS. However, it is not quenched in the urea unfolded protein (U form), both at pH 7.0 and at pH 2.2 (Figure 1C and D), where W214 and C34-IAEDANS are farther than the FRET distance. This observation indicates that C34-IAEDANS quenches the fluorescence of W214 in a distance-dependent manner. The extent of quenching of W214 by C34-IAEDANS in the E form is significantly less than that in the N form (Figure 1D). We used the data on HSA and HSA-IAEDANS, in Figure 1C and D, to quantitate the change in FRET efficiency (E) and D−A distance (R) during N ⇌ E transition, using eq 1[39]

$$E = 1 - \frac{F_{DA}}{F_D} = \left(1 + \frac{R^6}{R_0^6}\right)^{-1} \quad (1)$$

The fluorescence signal of W214 in HSA at the wavelength of the maximum emission was taken as F_D and that in the HSA-IAEDANS was taken as F_{DA}. We experimentally determined the value of Forster's distance, R_0, in the N form to be ∼25.8 Å, and in the E form to be ∼23.8 Å, respectively (SI, Figure S3 and Table S1). The D−A distance in the N form was determined to be ∼29.6 Å and that in the E form to be ∼33.1. This result indicates that the protein is expanded by ∼3.5 Å in the E form compared to the N form, along the axis connecting C34-IAEDANS and W214. Because vdW interactions have a steep dependence on the interatomic distances, the movement of the helix containing C34-IAEDANS in domain I and the helix containing W214 in domain II away from each other by ∼3.5 Å in the E form must be accompanied by severe disruption of the side-chain packing interactions involving these helices. We also observed that the fluorescence of 18 tyrosine residues that are distributed throughout the structure of HSA is quenched dramatically in the N form due to FRET with W214, but the extent of quenching decreases significantly in the E form (as in the U form) (SI, Figure S4), suggesting that in the E form the expansion of the protein might be global in nature.

We next investigated the water solvation of protein core in domain I and domain II. The fluorescence Stokes shift of a fluorophore is a very sensitive measure of the polarity of its surrounding medium. We monitored the changes in the hydrophobicity of domain I by monitoring the wavelength of the maximum fluorescence emission, λ_{max}^{em}, of C34-IAEDANS as a function of pH. In the N form, C34-IAEDANS is buried inside the hydrophobic core and its λ_{max}^{em} is 468 nm, which shifts toward red to 489 nm in the water-solvated U form (Figure 2A). We observed that in the E forms at pH 2.2 and pH 1.0, λ_{max}^{em} of C34-IAEDANS are 470 and 468 nm, respectively,

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

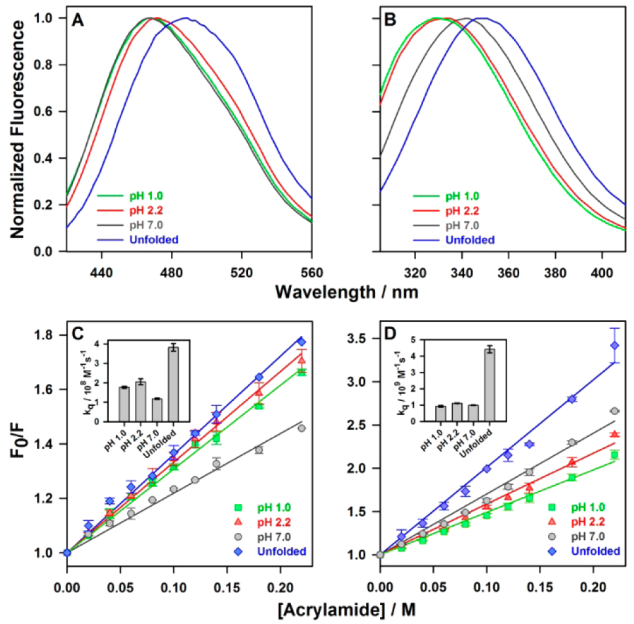


Figure 2. pH dependence of the solvation of hydrophobic core of domains I and II. Comparison of fluorescence spectra at different pH of (A) C34-IAEDANS and (B) W214. For comparison, each spectrum in *A* and *B* is normalized to the value of 1 at its emission maximum. Stern–Volmer plots for quenching of fluorescence of (C) C34-IAEDANS and (D) W214 by acrylamide. The solid lines through the data are fits to the equation $((F_0/F) = 1 + K_{SV}[Q])$. Insets in panel C (C34-IAEDANS) and in panel D (W214) compare the values of the bimolecular quenching constant, k_q (see text).

indicating that inner core of domain I retains its native-like hydrophobicity and is not accessible to water.

In the N form, W214 is partially buried in the hydrophobic core of the domain II and its λ_{max}^{em} is ~342 nm (Figure 2B). In the U form, W214 is fully solvated by water, and we observed that the λ_{max}^{em} is red-shifted to ~348 nm (Figure 2B). In the E forms at pH 2.2 and pH 1.0, λ_{max}^{em} is blue-shifted to ~333 nm, indicating that the inner core of domain II is hydrophobic and not accessible to water. It appears that the loosening of side-chain packing in domain II due to protein expansion allows the movement of W214 side-chain to more hydrophobic environment. It is possible that this movement of W214 brings it in the proximity of disulfide bonds, histidine residues, or carboxyl groups, any of which can quench its fluorescence intensity[39] (Figure 1B).

Because the cores of domain I and domain II are hydrophobic but loosely packed in the E forms, we next employed dynamic fluorescence quenching experiments[39−41] and explored whether transient structural fluctuations allow the buried C34-IAEDANS and W214 side-chains to come in molecular contact with quencher molecules on the periphery of the protein structure. Dynamic fluorescence quenching is described by the Stern–Volmer equation as[39]

$$\frac{F_0}{F} = 1 + K_{SV}[Q] = 1 + k_q\tau_0[Q] \quad (2)$$

We measured the fluorescence intensities of C34-IAEDANS (Figure 2C) and W214 (Figure 2D) in the absence (F_0) and in the presence (F) of different concentrations of acrylamide (Q), a neutral collisional quencher of IAEDANS and tryptophan fluorescence, in the N form, E forms, and the U forms. (F_0/F) was plotted against [acrylamide] (Figure 2C and D) and the slope of the straight line yielded the value of Stern–Volmer constant, K_{SV} (eq 2). The values of K_{SV} under different conditions are listed in Table S2 (SI). We observed that for C34-IAEDANS, the values of K_{SV} in the E forms at pH 2.2 and 1.0, are intermediate between that in the N form and the U form (Figure 2C). For W214, the values of K_{SV} in the E forms at pH 2.2 and 1.0, are slightly lower than that in the N form (Figure 2D). Using the values of K_{SV}, we determined the values of k_q, the bimolecular rate constant for the formation of the molecular contact between the fluorophore and acrylamide in the photoexcited state of the fluorophore ($K_{SV} = k_q\tau_0$). The values of τ_0, the intensity averaged fluorescence lifetime, were measured separately under above conditions (SI, Figure S5) and are listed in Table S2 (SI).

We observed that for C34-IAEDANS, the value of k_q is 3.7×10^8 M^{-1} s^{-1} in the water-exposed U form, which decreases to 1.2×10^8 M^{-1} s^{-1} in the N form (Figure 2C (inset)) due to the burial of C34-IAEDANS in the protein structure. For W214, the value of k_q is 4.4×10^9 M^{-1} s^{-1} in the U form, which decreases to 1.0×10^9 M^{-1} s^{-1} in the N form (Figure 2D (inset)). The values of k_q in the N and U forms are only 3.1-fold and 4.4-fold different, respectively, for C34-IAEDANS and W214. This difference, however, is significant as the errors in the values of k_q are unlikely to be more than ±10%, considering the small errors associated with the determination of the slopes of the straight lines (values of K_{SV}) and in the measurement of the fluorescence lifetimes (values of τ_0). We observed that for C34-IAEDANS, the values of k_q in the E forms at pH 2.2 and pH 1.0 are 1.9×10^8 M^{-1} s^{-1} and 1.8×10^8 M^{-1} s^{-1}, respectively, which are intermediate between that in the N form and the U form (Figure 2C (inset)). For W214, the values of k_q in the E forms at pH 2.2 and pH 1.0 are 1.1×10^9 M^{-1} s^{-1} and 0.9×10^9 M^{-1} s^{-1}, respectively, which are similar to that in the N form (Figure 2D (inset)). Fluorescence quenching for the fluorophores that are buried inside the hydrophobic core of proteins occurs through nanosecond (ns) structural fluctuations that expose the fluorophore to the quencher.[40,41] The above results suggest that in the E forms, the core of domain I exhibits faster structural fluctuations than the native protein, whereas domain II has native-like structural flexibility.

It is not fully understood whether the inner core of DMG is solid-like as in the native proteins or like a molten liquid as expected by disruption of the side-packing interactions.[11,27,32] We explored the solvation dynamics of the protein matrix in domain II using REES experiments[39,42] on W214. In general, λ_{max}^{em} of a fluorophore does not depend on the excitation wavelength. However, we observed that the λ_{max}^{em} of W214 in the E forms at pH 2.2 and pH 1.0, respectively, shifts to ~339 nm and ~338 nm from ~333 nm, when it is excited at 305 nm (at the red-edge of the excitation spectrum) compared to 295 nm (Figure 3A and B). In contrast, the λ_{max}^{em} of W214 remains constant at ~348 nm in the U form (Figure 3A (inset) and B) and changes nominally from ~342 nm to ~344 nm in the N form (Figure 3B), when the excitation wavelength is changed from 295 to 305 nm.

REES is a special case of fluorescence emission that occurs for polar fluorophores when (a) there exist a heterogeneous distribution of solvent dipoles around the fluorophore resulting in a broad distribution of solvent-fluorophore interaction energies and (b) the relaxation of solvent dipoles is slower than the fluorescence lifetime of the fluorophore, as in a highly viscous solvation environment.[39,42] In the U form, no REES is observed because W214 is surrounded by highly dynamic water

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

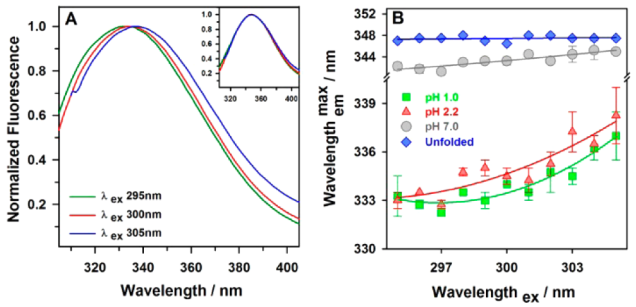


Figure 3. Dependence of wavelength of maximum fluorescence emission (λ_{max}^{em}) of W214 on the wavelength of excitation (λ_{ex}) as a function of pH. Panel A and its inset, respectively, show the representative fluorescence emission scan of W214 at pH 2.2 and in the U form when excited at different λ_{ex}. For comparison, each spectrum in (A) and its inset has been normalized to the value of 1 at its emission maximum. (B) $\lambda_{max}^{em}$is plotted as a function of λ_{ex}. In panel B, the solid lines through the data are drawn to guide the eye.

molecules. In the N form, W214 is located in a partially exposed hydrophobic cavity. The small magnitude (~2 nm) of REES observed in the N form implies that the complex solvation environment around W214 is less dynamic than the U form. In the E forms, W214 is buried in the hydrophobic core of domain II (Figure 2B and D). Hence, the large magnitude (~5−6 nm) of REES observed in the E form indicates that in this dry globular form the solvation environment created by the dipoles of protein matrix is highly heterogeneous and is similar to a molten and viscous liquid.[42]

We compared the global secondary structure and side-chain packing in the N form and the E form using far-UV and near-UV CD spectrum (Figure 4A and 4B). The mean residue ellipticity (MRE) at 222 nm, a measure of the α-helical content in proteins, is -28350 ± 3550 deg cm^2 dmol^{-1} in the N form, -6500 ± 1000 deg cm^2 dmol^{-1} in the U form, and -21400 ± 2000 deg cm^2 dmol^{-1} in the E forms at pH 2.2 and pH 1.0. These results indicate that E forms retain ~68% of the α-helical content of the N form. It has been demonstrated previously that domain III gets fully unfolded in the E form,[33] which apparently corresponds to melting of roughly one-third of the protein structure and contributes to ~32% change in the hydrogen-bonded secondary structure during N ⇌ E transition. The changes in the fluorescence of W214 during pressure-induced denaturation also reveal that domain III of HSA unfolds prior to domain II.[43]

There are 31 phenyl alanine residues packed tightly in the hydrophobic core of the N form and the near-UV CD spectrum appears to be dominated by their chirality. The observed near-UV CD spectrum has fine structures mainly in the 255 to 270 nm region (Figure 4B). The MRE at 261 nm in the N form is -275 ± 25 deg cm^2 dmol^{-1}. The MRE at 261 nm in the E forms at pH 2.2 and pH 1.0 are -200 ± 15 deg cm^2 dmol^{-1} which is very similar to the value of -175 ± 15 deg cm^2 dmol^{-1} in the U form. As expected for a dry molten globular structure, these results indicate that the core packing interactions are severely disrupted in the E form compared to the N form.

In the E forms, domain I and II retain their secondary structure and have dry cores, but vdW packing interactions are severely disrupted. We used these observations to estimate the relative contributions of vdW interactions and hydrophobicity of the protein cores, in domain I and II, in the stability of the N form (see the model energy diagram in Figure 5). We measured

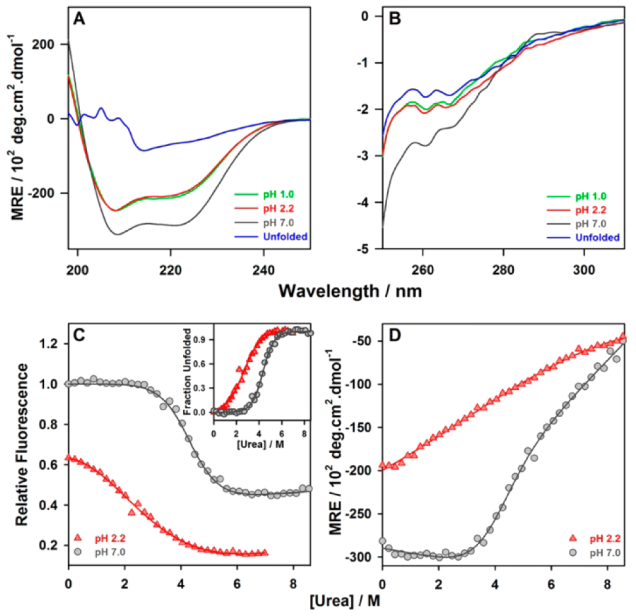


Figure 4. Changes in the secondary (A) and the tertiary structure (B) and the changes in the stability (C, D) of HSA as a function of pH. Urea-induced equilibrium unfolding transitions of the N form and the E form monitored by the change in fluorescence of W214 at 340 nm (C) and by the change in far-UV CD signal at 222 nm (D). In the inset of panel C, fraction of unfolded protein is plotted against [urea]. The solid lines in panel C (and its inset) and through pH 7.0 data in panel D are fits to a two-state model.[44] The solid red line through pH 2.2 data in panel D is drawn to guide the eye.

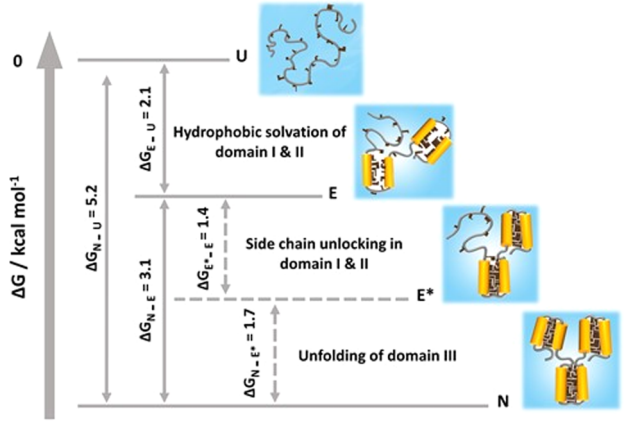


Figure 5. Model energy diagram.

the thermodynamic stabilities of the N form and the E form by the urea-induced equilibrium unfolding transitions at pH 7.0 and pH 2.2 which, respectively, represent N ⇌ U and E ⇌ U transitions, using fluorescence (Figure 4C) and far-UV CD (Figure 4D). For N ⇌ U transition monitored by both fluorescence and far-UV CD, the equilibrium values, respectively at 340 and 222 nm, show a cooperative or sigmoidal dependence on the concentrations of urea. The values of the free energy of unfolding (ΔG_{N-U}) and the slope of the transition (m_{N-U}, which represents the change in the solvent accessible surface area), as extracted from fitting the data to a two-state equation,[44] are 5.2 kcal mol^{-1} and 1.21 kcal mol^{-1} M^{-1} for fluorescence monitored transition and 5.1 kcal mol^{-1} and 1.24 kcal mol^{-1} M^{-1} for far-UV CD monitored transition.

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

The E form appears to be severely destabilized compared to the N form. The E ⇌ U transition monitored by fluorescence appears sigmoidal, and the values of ΔG_{E-U} and m_{E-U}, extracted from the two-state fit, are 2.1 kcal mol^{-1} and 0.82 kcal mol^{-1} M^{-1} (Figure 4C). These values can be used to determine ΔG_{N-E} ($\Delta G_{N-E} = \Delta G_{N-U} - \Delta G_{E-U}$) and m_{N-E} ($m_{N-E} = m_{N-U} - m_{E-U}$). The value of m_{N-E} is 0.39 kcal mol^{-1} M^{-1}. It is important to note that the value of m_{N-E} corresponds to ∼32% of m_{N-U}, indicating that ∼32% of the protein core is accessible to water in the E form due to the unfolding of domain III, which is in a very good agreement with the results of far-UV CD experiments (Figure 4A). Hence, the change in *m*-value during N ⇌ E transition is not due to the water solvation of domain I and domain II. The value of ΔG_{N-E} = 3.1 kcal mol^{-1}. Assuming an intermediate state E* (see Figure 5) in which domain III is completely solvated and unfolded but side-chain packing interactions in domain I and II remain intact, ΔG_{N-E} can be divided into two parts: (a) ΔG_{N-E^*}, due to the unfolding of domain III during N ⇌ E* transition; and (b) ΔG_{E^*-E}, due to the disruption of vdW packing interactions in domain I and II during E*⇌E transition. If we assume that N form loses roughly ∼32% of its stabilization energy (ΔG_{N-U}) due to the unfolding of domain III during N ⇌ E* transition, as suggested by the *m*-value and the far-UV CD measurements (see above), then the value of ΔG_{N-E^*} will be 1.7 kcal mol^{-1} and the value of ΔG_{E^*-E} will be 1.4 kcal mol^{-1}. The total contribution of domain I and II in the stability of the N form will be, then, given by the sum of ΔG_{E^*-E} and ΔG_{E-U}, which is 3.5 kcal mol^{-1}. A comparison of ΔG_{E^*-E} with ΔG_{E-U} (which is mainly due to the hydrophobic solvation of the cores of domain I and II) reveals that vdW packing and hydrophobic desolvation in domain I and II contribute in a 40:60 ratio in the thermodynamic stability of the N form.

We observed that the secondary structure in the E form unfolds in a gradual and noncooperative fashion (Figure 4D). This result suggests that (a) DMG might not be discrete states but an ensemble of loosely packed forms and (b) the cooperativity of the structural transition observed during the folding reaction of proteins is closely associated with the development of the side-chain packing interactions.[45,46] This result also supports the phase diagram of the previous theoretical studies,[21,22] which predicted that the transition of DMG to the fully unfolded state upon entry of water into protein interior is a continuous swelling process.

It is important to note that DMG-like expanded state was originally postulated to be the hypothetical transition state during the unfolding reaction of proteins.[21,22] However, here and in a few other studies, they have been observed as unfolding intermediates,[27,32,47] which probably lie on the native-side of the free-energy barrier. DMG-like intermediates have been observed when the free-energy landscape of proteins have been modulated by denaturants,[23−26,29] pressure,[31,48] and pH.[19] For the small protein HP35, the DMG-like state has been shown to exist in equilibrium with the native state.[30] It has also been observed that the unlocked and conformationally flexible DMG-like state of HP35 is compact in volume than the native state,[49] but the transition state for native to DMG transition is expanded.[49] Recent NMR studies on a mimic of the initial kinetic intermediate of RNase H suggest that the intermediate states with dry cores can probably also form on the unfolded-side of the free-energy barrier.[50] DMG-like states have also been observed in molecular dynamic simulations of folding transitions of model hydrocarbon chains.[51] In total, above studies appear to suggest that DMG-like intermediate states might be a general feature of the free-energy landscape of protein folding as predicted in a recent theoretical study.[52]

In summary, we have shown that the domain I and domain II of HSA resemble a DMG in the low pH state of the protein implying that DMG-like characteristics can be localized in a multidomain protein. We observed that (a) the domains are expanded at low pH compared to the native protein and that the side-chain packing is disrupted, but the hydrophobic core is not solvated by water; (b) the hydrophobic core of domain I exhibits ns structural fluctuation and that of domain II resembles a viscous and molten liquid in the DMG-like state; (c) the secondary structure in the DMG-like state unfolds in a gradual manner; (d) side-chain packing interactions contribute significantly to the protein folding cooperativity; and (e) the vdW interactions and hydrophobic effect in domain I and domain II relatively contribute in a 40:60 ratio toward the stability of the native state, indicating that side-chain packing plays a significantly important role in protein stability. Our observation of DMG-like intermediate during the pH-induced unfolding has an important implication for the mechanism of action of protons in unfolding proteins. According to the traditional mechanism, the electrostatic repulsion created by the protonation of surface residues results in protein swelling, side-chain unlocking and water penetration into the hydrophobic core in a single concerted step.[2,3] Our results, however, indicate that rupture of tight side-chain packing interactions and hydrophobic solvation are two distinct structural events during the pH-induced unfolding of proteins. DMG-like intermediates are usually not very stable under equilibrium conditions. In the case of HSA, it appears that the loss in enthalpy due to disruption of side-chain packing interactions is fairly compensated by the gain in conformational entropy of the protein, providing enough stability to the DMG-like state to be populated at equilibrium. Finally, the results of this study have important implications for signaling and other proteins where the function is dictated by entropy-driven conformational changes.[32,53−55]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.5b02545.

Experimental methods and data analysis; calibration curve using Bradford assay for the quantitation of IAEDANS-HSA (Figure S1); far-UV CD data comparing structure and stability of HSA and HSA-IAEDANS (Figure S2); data on calculation of R_0 (Figure S3); tyrosine−tryptophan FRET data (Figure S4); additional data on fluorescence quenching and data on fluorescence lifetimes (Figure S5); Table S1 containing FRET parameters; and Table S2 containing values of K_{SV}, k_q, and τ_0. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sk.jha@ncl.res.in. Tel.: 91-20-25902588. Fax: 91-20-25902615.

Author Contributions

(N.A. and P.M.) These authors contributed equally to this work.

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by CSIR-National Chemical Laboratory MLP grant number MLP030026. N.A. is a recipient of a Shyama Prasad Mukherjee Fellowship awarded by the Council of Scientific and Industrial Research, India. P.M. is a recipient of a Junior Research Fellowship by University Grants Commission, India. We thank Dr. Amitava Das for allowing us to use the time-correlated single photon counting setup in his laboratory.

■ REFERENCES

(1) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1−63.
(2) Tanford, C. Protein Denaturation. *Adv. Protein Chem.* **1968**, *23*, 121−282.
(3) Tanford, C. Protein Denaturation. C. Theoretical Models for the Mechanism of Denaturation. *Adv. Protein Chem.* **1970**, *24*, 1−95.
(4) Privalov, P. L. Stability of Proteins: Small Globular Proteins. *Adv. Protein Chem.* **1979**, *33*, 167−241.
(5) Sandberg, W. S.; Terwilliger, T. C. Influence of Interior Packing and Hydrophobicity on the Stability of a Protein. *Science* **1989**, *245*, 54−57.
(6) Kellis, J. T., Jr.; Nyberg, K.; Fersht, A. R. Energetics of Complementary Side-Chain Packing in a Protein Hydrophobic Core. *Biochemistry* **1989**, *28*, 4914−4922.
(7) Behe, M. J.; Lattman, E. E.; Rose, G. D. The Protein-Folding Problem: The Native Fold Determines Packing, but Does Packing Determine the Native Fold? *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 4195−4199.
(8) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science* **1992**, *255*, 178−183.
(9) Honig, B.; Yang, A. S. Free Energy Balance in Protein Folding. *Adv. Protein Chem.* **1995**, *46*, 27−58.
(10) Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. Forces Contributing to the Conformational Stability of Proteins. *FASEB J.* **1996**, *10*, 75−83.
(11) Bhattacharyya, S.; Varadarajan, R. Packing in Molten Globules and Native States. *Curr. Opin. Struct. Biol.* **2013**, *23*, 11−21.
(12) Baldwin, R. L. Dynamic Hydration Shell Restores Kauzmann's 1959 Explanation of How the Hydrophobic Factor Drives Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13052−13056.
(13) Richards, F. M. Areas, Volumes, Packing and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.
(14) Ohgushi, M.; Wada, A. 'Molten-Globule State': A Compact Form of Globular Proteins with Mobile Side-Chains. *FEBS Lett.* **1983**, *164*, 21−24.
(15) Ikeguchi, M.; Kuwajima, K.; Mitani, M.; Sugai, S. Evidence for Identity between the Equilibrium Unfolding Intermediate and a Transient Folding Intermediate: A Comparative Study of the Folding Reactions of Alpha-Lactalbumin and Lysozyme. *Biochemistry* **1986**, *25*, 6965−6972.
(16) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. Structural Characterization of a Partly Folded Apomyoglobin Intermediate. *Science* **1990**, *249*, 1544−1548.
(17) Colon, W.; Roder, H. Kinetic Intermediates in the Formation of the Cytochrome c Molten Globule. *Nat. Struct. Biol.* **1996**, *3*, 1019−1025.
(18) Raschke, T. M.; Marqusee, S. The Kinetic Folding Intermediate of Ribonuclease H Resembles the Acid Molten Globule and Partially Unfolded Molecules Detected under Native Conditions. *Nat. Struct. Biol.* **1997**, *4*, 298−304.
(19) Rami, B. R.; Udgaonkar, J. B. Mechanism of Formation of a Productive Molten Globule Form of Barstar. *Biochemistry* **2002**, *41*, 1710−1716.
(20) Prajapati, R. S.; Indu, S.; Varadarajan, R. Identification and Thermodynamic Characterization of Molten Globule States of Periplasmic Binding Proteins. *Biochemistry* **2007**, *46*, 10339−10352.
(21) Shakhnovich, E. I.; Finkelstein, A. V. Theory of Cooperative Transitions in Protein Molecules. I. Why Denaturation of Globular Protein Is a First-Order Phase Transition. *Biopolymers* **1989**, *28*, 1667−1680.
(22) Finkelstein, A. V.; Shakhnovich, E. I. Theory of Cooperative Transitions in Protein Molecules. Ii. Phase Diagram for a Protein Molecule in Solution. *Biopolymers* **1989**, *28*, 1681−1694.
(23) Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. Direct NMR Evidence for an Intermediate Preceding the Rate-Limiting Step in the Unfolding of Ribonuclease A. *Nature* **1995**, *375*, 513−515.
(24) Hoeltzli, S. D.; Frieden, C. Stopped-Flow NMR Spectroscopy: Real-Time Unfolding Studies of 6-^{19}F-Tryptophan-Labeled *Escherichia Coli* Dihydrofolate Reductase. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9318−9322.
(25) Jha, S. K.; Udgaonkar, J. B. Direct Evidence for a Dry Molten Globule Intermediate During the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12289−12294.
(26) Jha, S. K.; Marqusee, S. Kinetic Evidence for a Two-Stage Mechanism of Protein Denaturation by Guanidinium Chloride. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4856−4861.
(27) Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry Molten Globule Intermediates and the Mechanism of Protein Unfolding. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2725−2737.
(28) Dasgupta, A.; Udgaonkar, J. B.; Das, P. Multistage Unfolding of an SH3 Domain: An Initial Urea-Filled Dry Molten Globule Precedes a Wet Molten Globule with Non-Native Structure. *J. Phys. Chem. B* **2014**, *118*, 6380−6392.
(29) Sarkar, S. S.; Udgaonkar, J. B.; Krishnamoorthy, G. Unfolding of a Small Protein Proceeds Via Dry and Wet Globules and a Solvated Transition State. *Biophys. J.* **2013**, *105*, 2392−2402.
(30) Reiner, A.; Henklein, P.; Kiefhaber, T. An Unlocking/Relocking Barrier in Conformational Fluctuations of Villin Headpiece Subdomain. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4955−4960.
(31) Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled Motion in Proteins Revealed by Pressure Perturbation. *J. Am. Chem. Soc.* **2012**, *134*, 8543−8550.
(32) Baldwin, R. L.; Rose, G. D. Molten Globules, Entropy-Driven Conformational Change and Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 4−10.
(33) Dockal, M.; Carter, D. C.; Ruker, F. Conformational Transitions of the Three Recombinant Domains of Human Serum Albumin Depending on pH. *J. Biol. Chem.* **2000**, *275*, 3042−3050.
(34) Carter, D. C.; Ho, J. X. Structure of Serum Albumin. *Adv. Protein Chem.* **1994**, *45*, 153−203.
(35) Era, S.; Sogami, M. ^{1}H-NMR and CD Studies on the Structural Transition of Serum Albumin in the Acidic Region–the N– > F Transition. *J. Pept. Res.* **1998**, *52*, 431−442.
(36) Olivieril, J. R.; Craievich, A. F. The Subdomain Structure of Human Serum Albumin in Solution under Different pH Conditions Studied by Small Angle X-Ray Scattering. *Eur. Biophys. J.* **1995**, *24*, 77−84.
(37) Krishnakumar, S. S.; Panda, D. Spatial Relationship between the Prodan Site, Trp-214, and Cys-34 Residues in Human Serum Albumin and Loss of Structure through Incremental Unfolding. *Biochemistry* **2002**, *41*, 7443−7452.
(38) Huang, F.; Settanni, G.; Fersht, A. R. Fluorescence Resonance Energy Transfer Analysis of the Folding Pathway of Engrailed Homeodomain. *Protein Eng., Des. Sel.* **2008**, *21*, 131−146.
(39) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Springer: Singapore, 2006.
(40) Eftink, M. R.; Ghiron, C. A. Dynamics of a Protein Matrix Revealed by Fluorescence Quenching. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 3290−3294.

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

(41) Strambini, G. B.; Gonnelli, M. Fluorescence Quenching of Buried Trp Residues by Acrylamide Does Not Require Penetration of the Protein Fold. *J. Phys. Chem. B* **2010**, *114*, 1089−1093.
(42) Demchenko, A. P. Site-Selective Red-Edge Effects. *Methods Enzymol.* **2008**, *450*, 59−78.
(43) Tanaka, N.; Nishizawa, H.; Kunugi, S. Structure of Pressure-Induced Denatured State of Human Serum Albumin: A Comparison with the Intermediate in Urea-Induced Denaturation. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1997**, *1338*, 13−20.
(44) Agashe, V. R.; Udgaonkar, J. B. Thermodynamics of Denaturation of Barstar: Evidence for Cold Denaturation and Evaluation of the Interaction with Guanidine Hydrochloride. *Biochemistry* **1995**, *34*, 3286−3299.
(45) Jha, S. K.; Udgaonkar, J. B. Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry. *J. Biol. Chem.* **2007**, *282*, 37479−37491.
(46) Jha, S. K.; Dhar, D.; Krishnamoorthy, G.; Udgaonkar, J. B. Continuous Dissolution of Structure During the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11113−11118.
(47) Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. Urea Denaturation by Stronger Dispersion Interactions with Proteins Than Water Implies a 2-Stage Unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16928−16933.
(48) de Oliveira, G. A.; Silva, J. L. A Hypothesis to Reconcile the Physical and Chemical Unfolding of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2775−2784.
(49) Neumaier, S.; Kiefhaber, T. Redefining the Dry Molten Globule State of Proteins. *J. Mol. Biol.* **2014**, *426*, 2520−2528.
(50) Rosen, L. E.; Connell, K. B.; Marqusee, S. Evidence for Close Side-Chain Packing in an Early Protein Folding Intermediate Previously Assumed to Be a Molten Globule. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 14746−14751.
(51) Mountain, R. D.; Thirumalai, D. Molecular Dynamics Simulations of End-to-End Contact Formation in Hydrocarbon Chains in Water and Aqueous Urea Solution. *J. Am. Chem. Soc.* **2003**, *125*, 1950−1957.
(52) Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein Folding: From Theory to Practice. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22−29.
(53) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational Entropy in Molecular Recognition by Proteins. *Nature* **2007**, *448*, 325−329.
(54) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The Role of Conformational Entropy in Molecular Recognition by Calmodulin. *Nat. Chem. Biol.* **2010**, *6*, 352−358.
(55) Tzeng, S. R.; Kalodimos, C. G. Protein Activity Regulation by Conformational Entropy. *Nature* **2012**, *488*, 236−240.

DOI: 10.1021/acs.jpclett.5b02545
J. Phys. Chem. Lett. 2016, 7, 173−179

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

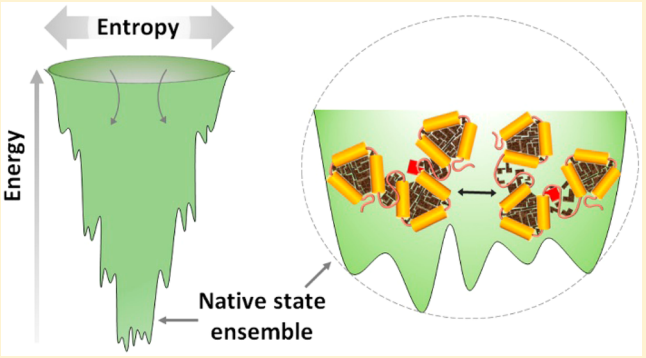
An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein

Prajna Mishra and Santosh Kumar Jha*

Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune, Maharashtra 411008, India

Supporting Information

ABSTRACT: It has been difficult to quantify the degree of side-chain conformational heterogeneity in the native (N) state ensemble of proteins and the relative energetic contributions of the side-chain packing and the hydrophobic effect in protein stability. Here, we show using multiple site-specific spectroscopic probes and tools of thermodynamics that the N state ensemble of a multidomain protein contains an equilibrium intermediate (I) whose interdomain region resembles a dry molten globule. In the I state, a tryptophan residue in the interdomain region is alternatively packed, but its secondary structure and intradomain packing are N-like. The I state also has a larger interdomain distance, but the domain−domain interface is dry and molten. Our results indicate that hydrophobic desolvation and side-chain packing are decoupled during protein folding and that interdomain packing interactions have an important energetic contribution in protein stability. Dynamic interconversion between alternatively packed N-like states could be important for multiple allosteric and ligand binding functions of this protein.



INTRODUCTION

The extent of conformational heterogeneity in the side-chain packing in the native (N) state ensemble of proteins is poorly understood.[1−12] Crystal structures of proteins project a time-averaged snapshot of the N state and show that the hydrophobic side-chain residues are packed tightly in the core and the packing densities are close to crystalline amino acids and solids.[1−4] In contrast, a body of theoretical work[5,7,11−13] suggests that the interior of folded proteins is dynamic in nature. It has been postulated that the low and the high energy conformations of the proteins with alternative side-chain packing arrangements exist at equilibrium in the folded protein ensemble and their populations follow Boltzmann probability distribution.[11,12,14] Fluctuations between these different N-like forms could play an important role in many biological functions like ligand-binding,[15] catalysis,[14] and signaling.[16] However, a high-resolution experimental characterization of these alternatively packed N-like conformations due to side-chain flexibility in the N state ensemble remains a significant challenge.

It has been proposed recently that the N state of proteins could be an ensemble of loosely packed dry molten globular forms.[10,17−19] Dry molten globules (DMG) are hypothesized to be expanded forms of the N state, which retain N-like global secondary and tertiary structure but have loose and molten side-chain packing and dehydrated cores. It has also been suggested using phase diagrams that during unfolding, proteins possibly sample an ensemble of DMG-like states near the N state basin and unfold via a DMG-like transition state.[20,21] Experimental data on DMG-like states are, however, very limited.[17,22−27] Structural and thermodynamic characterization of DMG-like intermediate (I) states can provide important insights on the relative contributions of the van der Waals (vdW) interactions and the hydrophobic effect in protein stability.[17] The conformational entropy of DMG-like states is higher as compared to that of the N states, which could also enable them to undergo dynamic volume fluctuations in the N states required for allosteric functions.[18,28] However, it is not clear how mobile and different are the packing interactions in the core of the DMG as compared to the N state.[10,18] It also remains to be understood whether DMG-like properties can be localized only in the selective regions of the protein structure. DMG have mainly been observed as transient kinetic intermediates during the unfolding of few proteins,[17,22−25,27] and it has been difficult to identify them as a subset of the N state ensemble under equilibrium conditions.[26,29,30] In particular, the detection of the rotation or movement of a protein side-chain residue in a member of the N state ensemble, which would constitute the most direct evidence in the support of heterogeneity of side-chain packing in the N state ensemble, has been challenging to capture for any multidomain protein.

Received: July 17, 2017
Revised: September 7, 2017
Published: September 12, 2017

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

In this study, we have site-specifically probed the energetics and structural composition of the N state ensemble of a multidomain protein, human serum albumin (HSA), under two different N-like environmental conditions, that is, at pH 5 and pH 7, using tools of thermodynamics and a battery of site-specific and high-resolution spectroscopic probes, including fluorescence resonance energy transfer (FRET), dynamic fluorescence quenching, red-edge excitation shift (REES), and near- and far-UV circular dichroism (CD). The application of multiple spectroscopic probes allowed us to dissect the conformational heterogeneity of the N state ensemble in terms of distinct structural events characterized by changes in protein expansion, vdW packing, core hydration, solvation dynamics of the hydrophobic core, and secondary and tertiary structure. We show that at pH 5, but not at pH 7, the N state ensemble is in equilibrium with an expanded DMG-like I state, which has global secondary and tertiary structure similar to that of the N state, but has a dry and molten core with an alternative interdomain side-chain packing arrangement. We use these observations to estimate the energetic contribution of interdomain side-chain packing in protein stability.

MATERIALS AND METHODS

Spectroscopic Methods and Instruments. All of the absorption and fluorescence emission spectra were measured using a PerkinElmer lambda 650 UV/vis spectrophotometer and a PerkinElmer fluorescence spectrometer LS 55, respectively, using a cuvette of path length 1 cm. All of the CD studies were done on a Jasco J-815 CD spectrometer. For far- and near-UV CD, the path lengths of the quartz cells used were 1 mm and 1 cm, respectively. Background signal corrections were done for all of the fluorescence and CD spectra by subtracting the buffer signals. An Abbe refractometer from Rajdhani Scientific Instruments Co. (model: RSR-2) was used to measure the refractive index of solutions wherever required.

Chemicals, Reagents, Buffers, and Experimental Conditions. HSA (99% pure, fatty acid and globulin free) and urea (ultrapure grade) were purchased from Alfa Aesar. 5-((((2-Iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was obtained from Life Technologies. Guanidine hydrochloride (GdmCl), acrylamide, and all other chemicals of highest purity grade were purchased from Sigma. All of the chemicals were used as received without any further purification. HSA concentration was determined by measuring the absorbance at 280 nm and using a molar extinction coefficient[31] of 36 500 M^{-1} cm^{-1}.

For all of the experiments, 20 mM phosphate and 20 mM acetate buffers were used as pH 7 and pH 5 native buffers, respectively. The unfolding buffers were composed of 9 M urea and 20 mM phosphate/acetate for the respective pH. For the I state, the buffer composition was 1.6 M urea and 20 mM acetate at pH 5. The urea and GdmCl concentrations were determined from refractive index measurements.[32] For pH titrations, the universal buffer was composed of 20 mM acetate, 20 mM phosphate, and 20 mM borate, and the pH was adjusted with HCl/NaOH.

Preparation of 1,5-IAEDANS Labeled HSA. We used the standard protocol to label HSA with 1,5-IAEDANS at the sole free cysteine, C34, as reported earlier.[30,33] In brief, the calculated volume of GdmCl unfolded HSA was added to a 20-fold molar excess of 1,5-IAEDANS in the unfolding buffer (6 M GdmCl and 20 mM Tris at pH 8). The mixture was then continuously stirred in the dark at room temperature for 4 h. After the completion of the labeling reaction, the reaction mixture was diluted 10-fold with the refolding buffer (20 mM phosphate at pH 7) and kept overnight at 4 °C. It was then concentrated to 2.5 mL using a centrifugal concentrator (GE healthcare) and passed through a PD-10 desalting column (GE healthcare) to separate the labeled HSA from free dye and GdmCl. The labeled protein concentration was determined by measuring the absorbance at 337 nm and using the molar extinction coefficient (ε_{337})[34,35] of 4500 M^{-1} cm^{-1}. The total protein concentration was determined by measuring the absorbance at 280 nm and using a molar extinction coefficient (ε_{280})[31] of 36 500 M^{-1} cm^{-1}. It should be noted that 1,5-IAEDANS has a molar extinction coefficient of around 900−1100 M^{-1} cm^{-1} at 280 nm,[36,37] which is 35−40-fold less as compared to that of the intrinsic ε_{280} of the protein (ε_{280} = 36 500 M^{-1} cm^{-1}). Hence, the contribution of 1,5-IAEDANS to the total absorbance at 280 nm will be only 2.5−3%, if the protein is completely labeled with 1,5-IAEDANS. For example, for a typical absorbance of 0.2 at 280 nm, at which all of the experiments have been performed, the contribution of 1,5-IAEDANS for the fully labeled protein will be 0.005, which is negligible. The percentage of labeling was calculated by dividing the labeled protein concentration by the total protein concentration and multiplying the ratio by 100. The protein was observed to be >95% labeled. It should be noted that our assumption of the negligible contribution of the absorbance of 1,5-IAEDANS at 280 nm gave the lower limit on the percentage of labeling, and the actual percentage will be slightly (2−3%) higher.

pH-Induced Equilibrium Unfolding Transition Monitored by Fluorescence and CD. For the fluorescence monitored pH titrations, the protein concentration in each sample was 4−5 μM. The samples were kept overnight at different pH values at room temperature. The fluorescence spectra for the protein samples were acquired by exciting the sole tryptophan (W214) at 295 nm and collecting the emission from 310 to 420 nm. Nominal slit widths for excitation and emission were used. Scan speed was 100 nm/min. The far- and near-UV CD spectra were monitored in the wavelength range of 190−250 nm and 250−310 nm, respectively. The protein concentration for far-UV CD experiment was 4−5 μM, whereas, for the near-UV CD experiment, it was ∼20 μM. For each CD spectrum, the data pitch, data integration time, bandwidth, and scan speed were 1 nm, 1 s, 2, and 100 nm/min, respectively. Three scans were averaged for each CD spectrum.

Analysis of the pH Titration. During the pH-induced unfolding transition,[30] HSA transforms from the fast-migrating (F) form to the N form between pH 3.5 and pH 4.8. Between pH 4.8 and pH 8.5, the protein remains in the N form followed by a transition to the basic (B) form beyond pH 8.5.

The far-UV CD monitored pH titration was analyzed using a model in which the structural transition was coupled to a single deprotonation step because the change in CD signal with respect to pH was observed to occur in a single sigmoidal manner. The data were fitted to eq 1:[30]

$$Y_{obs} = \frac{Y_F + Y_N 10^{(pH-pH_{m1})}}{1 + 10^{(pH-pH_{m1})}} \quad (1)$$

where Y_{obs} is the far-UV CD signal at 222 nm at a particular pH; Y_F and Y_N are the signals of the F and the N forms, respectively; and pH_{m1} is the midpoint of the observed titration.

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

Because the change in fluorescence signal with respect to pH occurred in two sigmoidal steps, the data were fitted to eq 2. This equation is based on a model in which the structural transitions from F to N form and from N to B form are each coupled to one deprotonation step:[30]

$$Y_{obs} = \frac{Y_F + Y_N 10^{(pH-pH_{m1})}}{1 + 10^{(pH-pH_{m1})}} + \frac{Y_B + Y_N 10^{(pH_{m2}-pH)}}{1 + 10^{(pH_{m2}-pH)}} \quad (2)$$

where Y_{obs} is the fluorescence signal at 340 nm at a particular pH value; Y_F, Y_N, and Y_B are the signals of the F, N, and B forms, respectively; and pH_{m1} and pH_{m2}, respectively, are the midpoints of the F to N and N to B transitions.

Urea-Induced Equilibrium Unfolding Transitions Monitored by Fluorescence and Far-UV CD. For the urea-induced equilibrium unfolding transitions at pH 7 and pH 5, the protein samples (4−5 μM) were incubated overnight at different concentrations of urea at room temperature. The equilibrium unfolding was monitored by the change in fluorescence of W214 at 328 nm upon excitation at 295 nm. For far-UV CD, it was monitored by the change in CD signal at 222 nm.

Analysis of the Two-State Equilibrium Unfolding Transitions at pH 7. For pH 7, the fluorescence and the far-UV CD monitored equilibrium unfolding transitions were analyzed using a two-state, N⇌U model. The data were fitted to eq 3:[38]

$$y_{obs} = \frac{y_N + y_U e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (3)$$

where Y_{obs} is the observed fluorescence/CD signal; Y_N and Y_U are the signals of the N and the unfolded (U) states, respectively; and ΔG_{NU} is the free energy of unfolding of the N⇌U transition. It is assumed that ΔG_{NU} has a linear dependence on denaturant concentration ([D]) and is given by the following equation:

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \quad (4)$$

where $\Delta G_{NU}^{H_2O}$ and m_{NU} are the standard free energy at 0 M urea and slope of the N⇌U transition, respectively.

The fractions of the native (f_N) and the unfolded (f_U) states at any particular denaturant concentration at pH 7 were determined using eqs 5 and 6, respectively.

$$f_N = \frac{1}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (5)$$

$$f_U = \frac{e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (6)$$

Analysis of the Three-State Equilibrium Unfolding Transitions at pH 5. For pH 5, the fluorescence and the far-UV CD monitored equilibrium unfolding transitions were analyzed using a three-state, N⇌I⇌U model. For the global fit, all of the data were fitted to eq 7:[39,40]

$$y_{obs} = \frac{y_N + (y_N + (y_U - y_N)z_I) e^{\frac{-\Delta G_{NI}}{RT}} + y_U e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (7)$$

where Y_N and Y_U are the signals of the N and the U states, respectively; and ΔG_{NI} and ΔG_{NU} are the free energies of unfolding of N⇌I and N⇌U transitions, respectively. z_I, which measures the optical signal of I (Y_I), relative to Y_N and Y_U, is given by

$$z_I = \frac{y_I - y_N}{y_U - y_N} \quad (8)$$

It is assumed that all of the free energies have a linear dependence on [D] and are given by the following equations:

$$\Delta G_{NI} = \Delta G_{NI}^{H_2O} + m_{NI}[D] \quad (9)$$

$$\Delta G_{IU} = \Delta G_{IU}^{H_2O} + m_{IU}[D] \quad (10)$$

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \quad (11)$$

where $\Delta G_{NI}^{H_2O}$, $\Delta G_{IU}^{H_2O}$, and $\Delta G_{NU}^{H_2O}$ are the standard free energies of N⇌I, I⇌U, and N⇌U transitions, respectively, and m_{NI}, m_{IU}, and m_{NU} are the respective slopes associated with the above-mentioned transitions. ΔG_{NU} and m_{NU} can also be expressed as $\Delta G_{NU} = \Delta G_{NI} + \Delta G_{IU}$ and $m_{NU} = m_{NI} + m_{IU}$, respectively.

The fractional change in signal, f_{app}, was determined from Y_{obs} (eq 7) by[40]

$$f_{app} = \frac{y_{obs} - (y_N + m_N[D])}{(y_U + m_U[D]) - (y_N + m_N[D])} \quad (12)$$

In the above equation, y_N and y_U are assumed to have a linear dependence on [D], which are given by the slopes m_N and m_U, respectively. The data obtained upon the transformation using eq 12 were fitted to eq 13.

$$f_{app} = \frac{z_I \times e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (13)$$

The fractions of the native (f_N), the intermediate (f_I), and the unfolded (f_U) states at pH 5, as a function of the [D], were calculated using the following equations:

$$f_N = \frac{1}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (14)$$

$$f_I = \frac{e^{\frac{-\Delta G_{NI}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (15)$$

$$f_U = \frac{e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (16)$$

All of the thermodynamic parameters derived from the global two-state fit for the data at pH 7 and three-state fit for the data at pH 5 are mentioned in Table 1.

FRET between W214 and C34-IAEDANS and Determination of Forster's Distance. We determined the intermolecular distance between W214 and C34-IAEDANS using the FRET methodology as described earlier.[30] For FRET measurements, all of the fluorescence spectra of the labeled and unlabeled protein in the N, the I, and the U states at pH 5 were collected in an identical manner with the excitation wavelength of 295 nm.

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

Table 1. Comparison of Thermodynamic Parameters for Urea-Induced Equilibrium Unfolding of HSA at pH 5 and pH 7 by Fluorescence Intensity (flu), CD, and λ_{max}^{em}

parameters[a]	flu/CD[b] (pH 7)	flu/CD[b] (pH 5)	λ_{max}^{em} [c] (pH 5)
$\Delta G_{NI}^{H_2O}$ (kcal mol^{-1})		2.1 ± 0.3	2.1 ± 1.2
$\Delta G_{IU}^{H_2O}$ (kcal mol^{-1})		3.2 ± 0.8	3.2 ± 1.4
$\Delta G_{NU}^{H_2O}$ (kcal mol^{-1})	5.4 ± 0.3	5.3 ± 0.7	5.3 ± 0.7
m_{NI} (kcal mol^{-1} M^{-1})		−2.3 ± 0.4	−2.6 ± 3.0
m_{IU} (kcal mol^{-1} M^{-1})		−1.0 ± 0.5	−1.0 ± 4.1
m_{NU} (kcal mol^{-1} M^{-1})	−1.2 ± 0.1	−3.3 ± 0.3	−3.6 ± 2.9
Z_I[d]		0.4 ± 0.2	0.1 ± 0.8

[a]All of the thermodynamic parameters for pH 5 and pH 7 were determined, respectively, from the three-state (N⇌I⇌U) and the two-state (N⇌U) analysis of urea-induced equilibrium unfolding experiments. A global analysis was performed as discussed in the Materials and Methods. All of the parameters shown here can be used to generate all of the fits shown in Figures 2C, inset, and 3A, inset. [b]For pH 5, the error bars shown were estimated from three independent measurements, whereas, for pH 7, the errors shown resulted from five independent measurements. [c]Although λ_{max}^{em} is an intensive property and not directly proportional to the amount of the molecule, we fitted the pH 5 data in Figure 2B using the three-state equilibrium unfolding model, and constraining the value of $\Delta G_{NU}^{H_2O}$. Very surprisingly, we observed that the thermodynamic parameters are quite similar to the equilibrium unfolding curve monitored by fluorescence and far-UV CD. [d]Z_I is a measure of the optical signal of I relative to the optical signals of the N and the U states.

The efficiency of energy transfer, E, was determined according to eq 17:[41]

$$E = 1 - \frac{F_{DA}}{F_D} \tag{17}$$

where F_{DA} and F_D are the donor fluorescence intensities in the presence and the absence of the acceptor, respectively.

The donor−acceptor distance, R, is related to the efficiency of energy transfer, E, by[41]

$$R = R_0\left[\frac{1-E}{E}\right]^{1/6} \tag{18}$$

where R_0 is the Forster's distance at which the energy transfer is 50%. The R_0 was determined by eq 19:[25,30,41]

$$R_0 = 0.211[Q_D J\kappa^2 n^{-4}]^{1/6} \tag{19}$$

In the above equation, Q_D is the quantum yield of the donor fluorescence, J is the overlap integral, κ^2 is the orientation factor, and n is the refractive index of the medium.

Q_D at pH 7 was taken as 0.31.[30] Because Q_D is directly proportional to the area under the fluorescence emission spectrum,[41] we used the value for the Q_D at pH 7 and the ratio of the area under the fluorescence emission spectrum at pH 5 to the area under the fluorescence emission spectrum at pH 7, and determined the value of Q_D at pH 5 to be 0.30. Similarly, the value of Q_D for the I state was determined to be 0.27.

The overlap integral, J, expresses the spectral overlap between the donor emission and the acceptor absorption, and is given by[25,41]

$$J = \frac{\int F(\lambda)\varepsilon(\lambda)\lambda^4\, d\lambda}{F(\lambda)\, d\lambda} \tag{20}$$

where $F(\lambda)$ is the fluorescence emission spectra of W214 and $\varepsilon(\lambda)$ is the absorption spectra of C34-IAEDANS. The refractive index, n, of the medium at pH 5 was determined to be 1.3315 ± 0.0005, whereas, for the I state, it was 1.3451 ± 0.0005. The value of the orientation factor, κ^2, was taken to be 2/3, assuming that the donor and the acceptor are randomly oriented with respect to each other. The values of R_0, J, and other FRET parameters are listed in Table S1.

Acrylamide Quenching Experiments. The solvent accessibility assay was performed by using acrylamide, a neutral quencher of W214 fluorescence. All of the protein samples under different conditions were incubated overnight at room temperature. Calculated volumes from a freshly prepared 2 M acrylamide stock solution were added to the protein samples just before collecting the spectra. The protein was excited at 295 nm, and the emission spectra were collected from 310 to 420 nm.

We used the Stern−Volmer equation[41] to analyze the quenching of fluorescence intensity at the emission maxima, that is:

$$\frac{F_0}{F} = 1 + K_{SV}[Q] \tag{21}$$

where F_0 and F are the fluorescence intensities at the emission maxima at 0 M and any particular acrylamide concentration [Q], respectively; and K_{SV} is the Stern−Volmer constant, which is given by[41]

$$K_{SV} = \tau_0 k_q \tag{22}$$

In the above equation, τ_0 and k_q are the intensity averaged fluorescence lifetime and bimolecular quenching constants, respectively. We determined the values of τ_0 for W214 in the N and the U states at pH 7 and pH 5 from the fluorescence lifetime data reported in the previous studies.[42,43] The fluorescence lifetime decays of W214 fit to a sum of three exponentials. Using the three components of the fluorescence lifetimes and their relative amplitudes, we calculated the values of τ_0 using the following equation:[30,41]

$$\tau_0 = \frac{\sum_i \alpha_i \tau_i^2}{\sum_i \alpha_i \tau_i} \tag{23}$$

where α_i is the amplitude of the ith fluorescence lifetime τ_i such that $\sum_i \alpha_i = 1$ and i = 1−3. The calculated values of τ_0 are listed in Table S2.

For the I state, τ_0 was estimated as per the following procedure. Using the fluorescence lifetime data reported in the previous studies,[42,43] we first calculated the amplitude averaged fluorescence lifetime,[41] τ_m, using the following equation:[30,41]

$$\tau_m = \sum \alpha_i \tau_i \tag{24}$$

where α_i is the amplitude of the ith fluorescence lifetime τ_i such that $\sum_i \alpha_i = 1$ and i = 1−3. Because quantum yield, Q_D, is directly proportional to τ_m[41] at any particular condition, we used the known values of Q_D and τ_m for the N state at pH 7 (0.31 and 6.2 ns, respectively) and the value of Q_D for the I state (0.27) and calculated the value of τ_m for the I state to be 5.4 ns. In the previously reported studies,[42,43] we observed that the ratio of τ_0 with respect to τ_m for all of the conditions is 1.11. Because of this, we assumed that for the I state also, this ratio will hold good, and we calculated the value of τ_0 for the I state to be 6.0 ns.

The values of k_q were calculated from the values of K_{SV} and τ_0, using eq 22. For the U states at pH 7 and pH 5, the values of

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

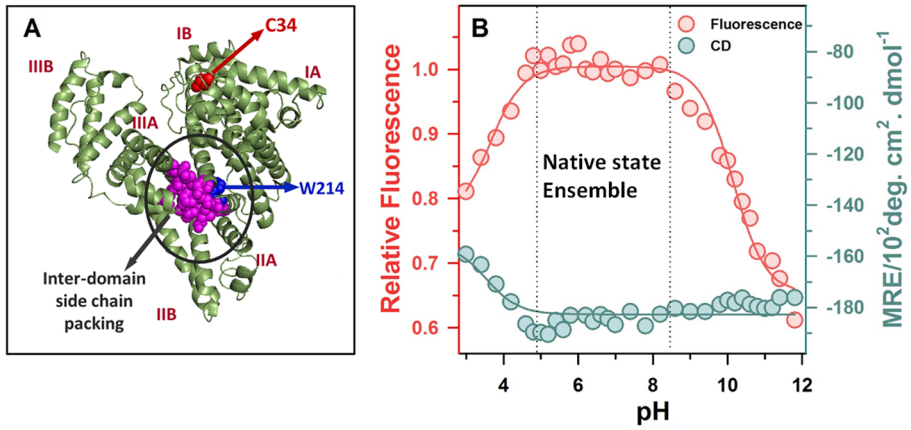


Figure 1. HSA maintains N-like conformation between pH 4.8 and pH 8.5. (A) The location of W214 is shown, which is involved in the interdomain side-chain packing in the three domain structure of the protein (drawn from PDB file 1AO6 using the program PyMOL). The location of C34 in domain I is also shown. (B) Mean residue ellipticity (MRE) of the protein at 222 nm and fluorescence emission of W214 at 340 nm are shown as a function of pH according to the right and the left *y*-axis, respectively. The blue and the red lines through the data represent a fit to a two-state (eq 1) and a three-state (eq 2) pH titration model, respectively.

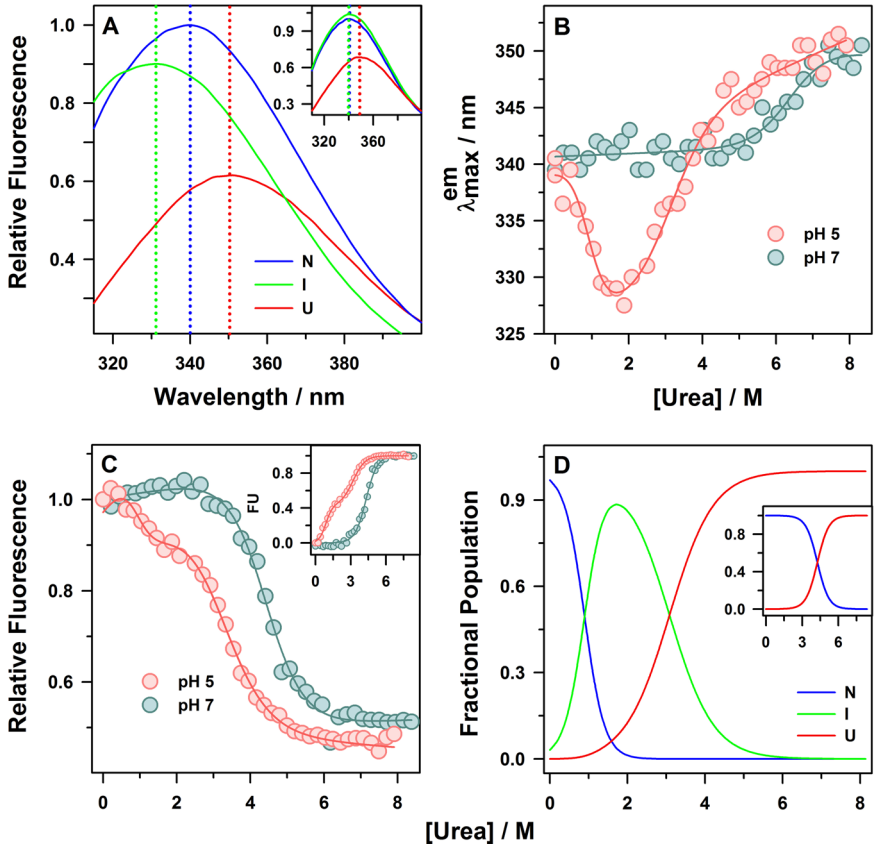


Figure 2. Urea-induced structural transition of HSA at pH 5 and pH 7. (A) Panel A and its inset, respectively, show the fluorescence emission spectra of the protein at pH 5 and pH 7 in 0 M (blue solid line), 1.6 M (green solid line), and 9 M (red solid line) urea. The dotted lines represent the wavelength of maximum fluorescence in the respective conditions. (B) Changes in the wavelength of maximum fluorescence emission (λ^{em}_{max}) are plotted as a function of [urea]. (C) Equilibrium unfolding transition monitored by the change in W214 fluorescence at 328 nm. The inset shows the apparent fraction of the unfolded protein (FU) as a function of [urea]. (D) The changes in the population of the N, the I, and the U states at pH 5 and the N and the U states at pH 7 (inset) are plotted as a function of [urea] (eqs 9−11, 5, and 6). The blue and the red lines through the data in panels B and C and in the inset of panel C are fits to a two-state and a three-state model, respectively. The axes of the insets in panels A, C, and D, wherever not labeled, are the same as that of the main panel.

k_q were corrected for the viscosity of 9 M urea as described previously.[30] All of the values of k_q are mentioned in Table S2.

REES Experiments. The fluorescence emission spectra for the REES experiments were collected after excitation at different wavelengths ranging from 295 to 305 nm. The protein concentrations used were 8−15 μM. The spectra were recorded with a scan speed of 80 nm/min and averaged over three scans. A slit width of 5−8 nm was used for excitation. The

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

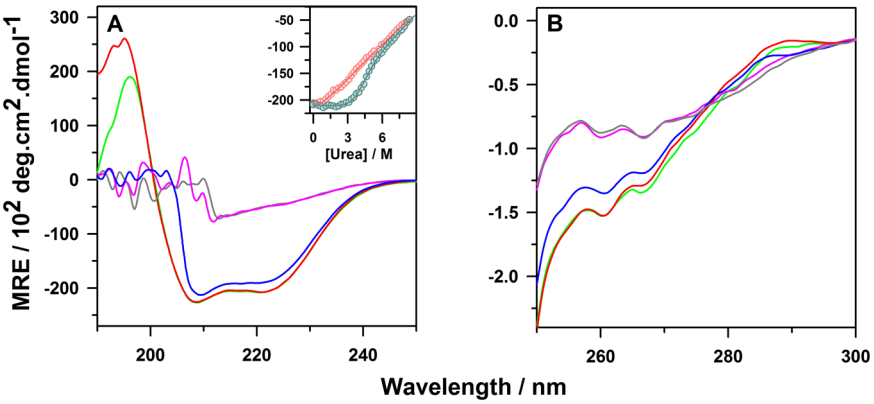


Figure 3. Global structural analysis of HSA at pH 5 and pH 7. (A) Changes in MRE as a function of wavelength in the peptide region (secondary structure) and (B) in the aromatic region (tertiary structure). In both of the panels, the signals of the N state at pH 5 and pH 7 are shown by the green line and the dark red line, respectively; the signals of the U state at pH 5 and pH 7 are shown by the gray line and the pink line, respectively, and the signal of the I state is shown by the dark blue line. The inset in panel A shows the changes in MRE at 222 nm as a function of [urea] at pH 5 (red ○) and pH 7 (cyan ○). The blue and the red lines through the data in the inset of panel A are fits to a two-state and a three-state model, respectively. The *y*-axis of the inset in panel A has the same label as that of the main panel.

error bars for the values of the emission maximum were estimated from two independent measurements.

RESULTS AND DISCUSSION

pH Dependence of the Spectroscopic Properties of the N State Ensemble of HSA. The three-dimensional structure of HSA consists of 28 helices packed and distributed over its three domains (Figure 1A). The three domains of the protein are held together by a tightly packed interdomain cluster of hydrophobic atoms of residues L198, S202, F206, A210, W214, A217, V343, V344, L347, N458, L481, V482, and R484. The sole tryptophan residue of the protein, W214, is located in a helical segment of domain II (Figure 1A). During pH titration of the protein, we observed that the fluorescence of W214 remains constant between pH 4.8 to pH 8.5, but decreases in a sigmoidal manner below pH 4.8 and above pH 8.5 (Figure 1B). We also monitored the changes in secondary structure of the protein during pH titration by measuring the changes in the far-UV CD signal at 222 nm (Figure 1B). We observed that the secondary structure of the protein decreases marginally below pH 4.8, but remains almost constant from pH 4.8 to pH 12. These results indicate that HSA undergoes an acid-induced structural transition below pH 4.8 and a base-induced structural transition above pH 8.5, as reported in previous studies.[44,45] The acid-induced structural transition is accompanied by changes in both the tertiary and the secondary structure of the protein, but the base-induced transition is only accompanied by changes in tertiary structure. Most importantly, these results suggest that HSA maintains N-like secondary and tertiary structure between pH 4.8 and 8.5, which constitute its N state ensemble.

Thermodynamic Characterization of the N State Ensemble under Different N-like Conditions. We measured the structural stability of the N state ensemble at two different N-like environmental conditions, pH 5 and pH 7, by urea denaturation experiments (Figure 2). The fluorescence Stokes shift of the tryptophan residue is very sensitive to the polarity of its surrounding medium. W214 is partially buried in the hydrophobic core at the interface of domains I and II in the N state (Figure 1A). Accordingly, we observed that, at both pH 5 and pH 7, the mean value of the wavelength of maximum fluorescence emission (λ_{max}^{em}) in the N state is 340 nm (Figure 2A and inset). The λ_{max}^{em} shifts to red to 350 nm in the U state in 9 M urea, at both pH 5 and pH 7, indicating the complete hydration of W214 upon unfolding (Figure 2A, inset and B). At pH 7, the λ_{max}^{em} of W214 changes in an apparently two-state fashion as a function of [urea], indicating that W214 broadly experiences only N-like and U-like solvation during unfolding (Figure 2A, inset and B). However, at pH 5, we observed that λ_{max}^{em} of W214 changes in two distinct sigmoidal steps during urea-induced unfolding (Figure 2A and B). At pH 5, the λ_{max}^{em} first shifts to blue to 330 nm and plateaus around 1.6 M urea, which later shifts to red to 350 nm upon complete unfolding in 9 M urea. These results indicate that urea-induced unfolding at pH 5 occurs via an I state in which tertiary structural changes at the interface of domain I and II allow the movement of W214 from a partially exposed to a highly hydrophobic environment.

Because λ_{max}^{em} is thermodynamically an intensive property, we monitored the changes in fluorescence intensity of W214 as a function of [urea] to determine the changes in free energies and populations of the N, the I, and the U states during N⇌I⇌U transition. We observed that fluorescence intensity of W214 during urea-induced equilibrium unfolding transition changes in a single sigmoidal step at pH 7, but in two distinct sigmoidal steps at pH 5 (Figure 2C). We analyzed our data according to a two-state model at pH 7 and a three-state model at pH 5 (Figure 2C, inset). The thermodynamic parameters derived from the global analysis of the data are listed in Table 1. We observed that the standard free energy of unfolding during the N⇌U transition, $\Delta G_{NU}^{H_2O}$, at pH 5 is 5.3 ± 0.7 kcal mol^{-1} and at pH 7 is 5.4 ± 0.3 kcal mol^{-1}, respectively. The $\Delta G_{NU}^{H_2O}$ values at both pH's are similar (within error), indicating that the structural stability of the N state is the same at the two N-like conditions. The value of $\Delta G_{NI}^{H_2O}$, the standard free energy of unfolding during the N⇌I transition, is 2.1 ± 0.3 kcal mol^{-1}, indicating that 35−40% of the change in the total free energy of unfolding occurs during the partial unfolding of the N state to the I state.

We determined the changes in the populations of the N, the I, and the U states as a function of [urea] by using the thermodynamic parameters for N⇌I, I⇌U, and N⇌U transitions (Figure 2D). We found that the I state is maximally populated to the extent of 85 ± 4% at ∼1.6 M urea, where the population of the N state is 8 ± 6% and the U state is 7 ± 3%,

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

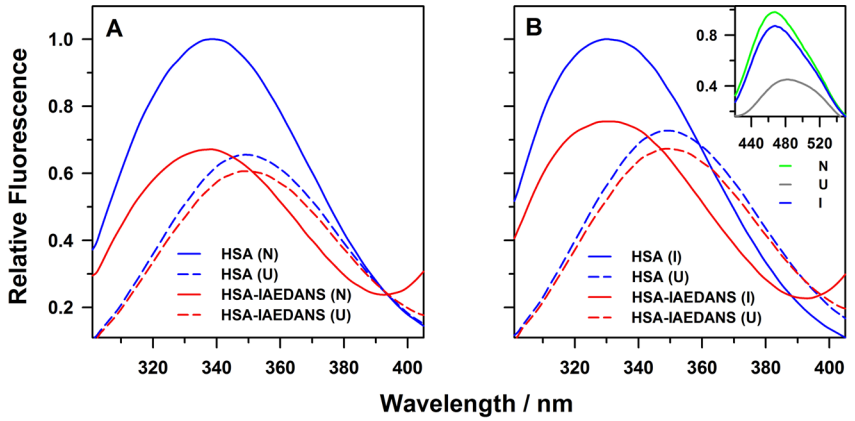


Figure 4. Interdomain region in the I state is expanded. The distance between W214 and C34-IAEDANS was monitored by FRET. Fluorescence emission spectra of W214 in HSA and HSA-IAEDANS at pH 5 are shown in panel (A) for the N and the U states and in panel (B) for the I and the U states. The inset in panel (B) shows the fluorescence emission spectra of C34-IAEDANS in the N, the I, and the U states upon excitation at 295 nm. The axes labels of the inset in panel B are the same as that of the main panel.

respectively. No intermediate state was observed to be populated at pH 7 (Figure 2D, inset). Very interestingly, we observed that around 5% of the I state is populated under N-like conditions at pH 5, that is, at 0 M urea. This result indicates that the N state ensemble at pH 5 is structurally and energetically heterogeneous.

Comparison of the Global Structure of the N and the I States. We maximally populated the I state at 1.6 M urea (pH 5) and characterized its structure in terms of secondary and tertiary structural content, side-chain packing, core hydration, and solvation dynamics of the protein side-chains in its core. First, we compared the global secondary structure and stability of the I state and the N state ensemble by far-UV CD spectroscopy. We observed that the far-UV CD spectra of the N state and the U state at pH 5 are identical to those of the N state and U state at pH 7 (Figure 3A). The mean residue ellipticity (MRE) at 222 nm is a measure of the secondary structural content in proteins. We observed that the mean values of MRE at 222 nm for the I, the N, and the U states are −18 850, −20 650, and −4950 deg cm^2 dmol^{-1}, respectively. The change in MRE during the N⇌I transition is only 10% of the total change in MRE between the N and the U states. Because 7 ± 3% of unfolded molecules are also present at 1.6 M urea at pH 5 (Figure 2D), this result indicates that the secondary structure of the I state is similar to that of the N state, and that the protein−protein hydrogen bonds (H-bonds) are not replaced by protein−water H-bonds.

We also measured the stability of the secondary structure of the N state ensemble at pH 5 and pH 7. We observed that the stability of the N state ensemble at pH 7 as measured by far-UV CD is similar to that of the fluorescence (Figure 3A, inset, Figure 2C, and Table 1). For both fluorescence and CD measured unfolding transitions, the midpoint of transition is 4.5 M [urea]. At pH 5, we observed a three-state urea-induced equilibrium unfolding transition (Figure 3A, inset). However, the change in MRE between the N and the I states is very small, and unfolding of the secondary structure in the I state appears to be gradual and noncooperative. Nevertheless, the equilibrium unfolding curve can be satisfactorily fitted to a three-state model using parameters similar to those of the fluorescence monitored equilibrium unfolding curve (Table 1).

We next compared the global tertiary structure of the I state and the N state ensemble by near-UV CD spectroscopy (Figure 3B). The near-UV CD spectrum of HSA in the N state has absorption bands mainly in the 255− 270 nm region (Figure 3B), indicating that it primarily reports on the chirality created by the asymmetric packing of its 31 phenylalanine residues in the hydrophobic core. The near-UV CD spectra of the N state and the U state at pH 5 are very similar to those of the N state and the U state at pH 7. We observed that MRE values at 261 nm for the I, the N, and the U states are −135 ± 15, −150 ± 15, and −85 ± 15 deg cm^2 dmol^{-1}, respectively, and around 10−20% of the total change in MRE occurs during the N⇌I transition. Accounting for the decrease in MRE due to the presence of 7 ± 3% of unfolded molecules, these results suggest that the global tertiary structures in the I state and the N state are only marginally different. This is a surprising result in light of the large blue shift of the λ_{max}^{em} of W214 in the I state (Figure 2A and B), which indicates that the side-chain packing interactions involving W214 must be broken in the I state. In the crystal structure of HSA, we observed that 30 of the 31 phenylalanine residues are mainly distributed in the core of the three domains, and only one phenylalanine residue, F206, is spatially located in the interdomain region near W214 and participates in the interdomain side-chain packing (Figure S1). Hence, the observation that the λ_{max}^{em} of W214 is highly blue-shifted in the I state but its near-UV CD spectrum is similar to the N state suggests that the intradomain side-chain packing is intact in the I state but the interdomain side-chain packing is disrupted.

Interdomain Expansion in the I State as Revealed by Site-Specific FRET Measurements. The disruption of interdomain side-chain packing must lead to the movement of the protein domains away from each other. To probe this further, we measured the interdomain distances between domains I and II in the N state and the I state using FRET. We used W214, which is located in the middle of a helix in domain II, as the donor (D) fluorophore. HSA has a single free cysteine residue, C34, which is located at the beginning of helix 3 in domain I (HSA has 35 cysteine residues in total, 34 of which form disulfide bonds; Figure S2). We labeled the thiol moiety of C34 with the fluorescent dye 1,5-IAEDANS, and used it as the acceptor (A) fluorophore. We have shown previously that C34 can be quantitatively labeled with 1,5-IAEDANS[30] and that W214 and C34- IAEDANS form a FRET pair.[30] The emission spectrum of W214 significantly overlaps

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

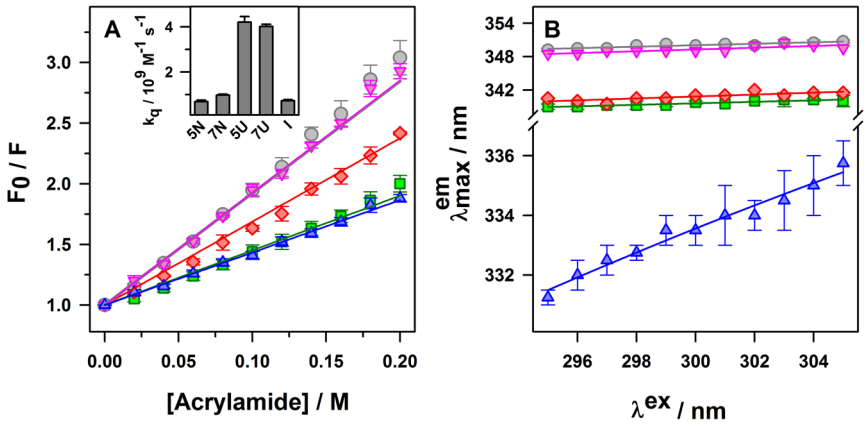


Figure 5. Interdomain region in the I state has a hydrophobic and molten interior. (A) Stern−Volmer plots for quenching of fluorescence of W214 by acrylamide in the N (green ■), the I (dark blue ▲), and the U (gray ●) states at pH 5, and the N (red ◆) and the U (pink ▼) states at pH 7. The solid lines through the data are the least-squares fits to eq 21. The values of bimolecular quenching constant, k_q, are compared in the inset. (B) The wavelength of maximum fluorescence emission (λ_{max}^{em}) of W214 is dependent on the wavelength of excitation (λ^{ex}) in the I state (dark blue ▲) but is largely independent in the N (green ■) and the U (gray ●) states at pH 5 and the N (red ◆) and the U (pink ▼) states at pH 7. In panel B, the solid lines through the data are drawn to guide the eyes.

with the absorbance spectrum of C34-IAEDANS (Figure S3). The secondary structure and thermodynamic stability of the unlabeled (HSA) and the 1,5-IAEDANS labeled (HSA-IAEDANS) proteins are also similar,[30] and hence the data on HSA and HSA-IAEDANS can be compared directly. C34-IAEDANS quenches the fluorescence of W214 to a larger extent in the N state as compared to the U state (Figure 4A). This result is expected because C34-IAEDANS and W214 are located nearer to each other in the N state but far away in the U state. We observed that the extent of quenching of W214 fluorescence by C34-IAEDANS in the I state is significantly less than that in the N state (Figure 4A and B).

We used the data on the fluorescence intensity of the donor in the absence and the presence of the acceptor to quantitate the D−A distance in the N state and the I state using eq 18. The values of Forster's distance is determined experimentally and found to be 25.7 Å in the N state and 25.0 Å in the I state (Figure S3 and Table S1). We observed that the D−A distance in the N state is 29.3 ± 0.1 Å and that in the I state is 30.3 ± 0.1 Å. The D−A distance in the I state is only 1 Å larger than that in the N state. This difference, however, is significant because of the following four reasons: (i) If a D−A pair is nearer to each other, the energy transfer by FRET results in the decrease of the fluorescence emission of the donor fluorophore (quenching) but increase in the fluorescence emission of the acceptor fluorophore. We observed that the fluorescence of C34-IAEDANS in the N state is higher than that in the I state upon excitation at 295 nm (Figure 4B, inset). The change in fluorescence intensity between the N and the I states is ∼33% of the total change in the fluorescence intensity between the N and the U state. These results indicate that the extent of energy transfer is more in the N state than the I state and least in the U state, and hence the D−A distance in the N state is less than the I state and the largest in the U state. (ii) The measured D−A distances in the N and the I states are very near to R_0 and hence lie in the most sensitive region of the FRET measurements (between 0.5 R_0 and 1.5 R_0). (iii) We determined the D−A distances in the N and I states by 3−5 separate measurements and observed that the standard error in D−A distances is only 0.1 Å (Table S1). (iv) Because most of the secondary and tertiary structures of the protein are intact in the I state, a large change in interatomic distances is not even expected. Hence, the results of the FRET experiments imply that domain I and domain II have moved away from each other by 1 Å in the I state as compared to the N state, along the axis connecting W214 and C34-IAEDANS. Because the strength of the vdW interactions varies steeply with interatomic distances, these results indicate that the interdomain packing interactions between domain I and II must be loose in the I state. The loosening of side-chain packing appears sufficient to allow the movement of W214 side-chain to a more hydrophobic region of the domain−domain interface in the I state. It should be kept in mind, however, that the loosely packed N-like intermediate states could also be more compact in size than the N states, as has been reported recently for HP35.[19]

Dynamic Fluorescence Quenching Experiments Reveal that the Interdomain Region is Dry. We observed that the disruption of side-chain packing between domain I and II is not accompanied by hydration of the intervening hydrophobic core. The λ_{max}^{em} of W214 in the I state is 330 nm, indicating that its immediate surrounding is hydrophobic in nature (Figure 2A and B). We used dynamic fluorescence quenching experiments[41,46,47] to understand whether loose structure around W214 allows it to undergo structural fluctuations and come in molecular contact with the quencher molecules present in water. Acrylamide is a neutral collisional quencher of tryptophan fluorescence.[41] We measured the fluorescence intensity of W214 in the absence and in the presence of different concentrations of acrylamide in the N, the I, and the U states (Figure 5A) and used the Stern−Volmer equation[41] to determine the values of the Stern−Volmer constant, K_{SV}, and the bimolecular rate constant, k_q, for the formation of the molecular contact between W214 and acrylamide in the photoexcited state of W214. The values of K_{SV} in the U states at pH 5 and pH 7 are 9.22 and 9.26 M^{-1}, respectively. The values of K_{SV} in the N states at pH 5 and pH 7 are 4.51 and 6.88 M^{-1}, respectively. The value of K_{SV} in the I state is 5.02 M^{-1}, which is very similar to the value observed in the N state ensemble at pH 5.

We used these values of K_{SV} to determine the values of k_q. We observed that the values of k_q in the U state at pH 5 and pH 7, where W214 is completely exposed to water, are 4.20×10^9

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

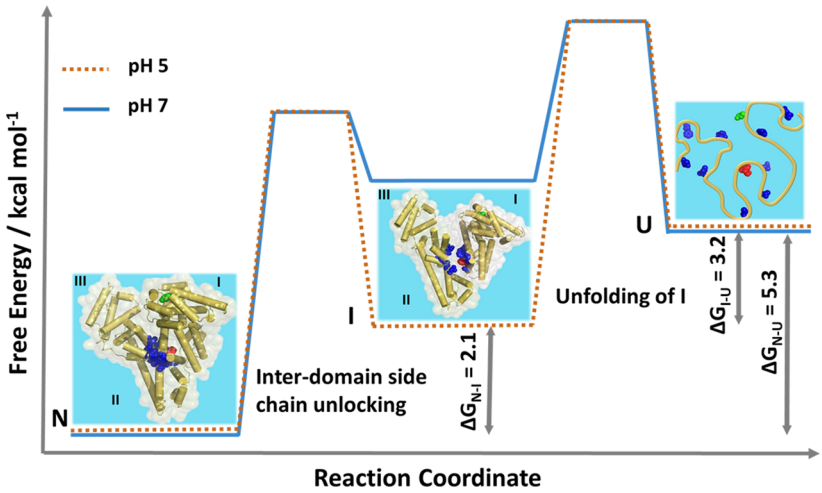


Figure 6. Model energy diagram. The schematic figure shows the relative energies of the N, the I, and the U states. In the I state, the protein is expanded and the side-chain packing is disrupted in the interdomain region, allowing W214 to move to a hydrophobic cavity. These events occur without the hydration of the intervening hydrophobic region. Structural models of the N and the I states are drawn from the PDB file 1AO6 using the program PyMOL. In the N and the I states, yellow cylinders represent the α-helical secondary structure of the protein. In the U state, the yellow curved line shows the protein backbone. The white boundary around the protein in the N and the I states shows the solvent accessible surface area of the protein. The light blue color represents water molecules. The side-chain residues involved in interdomain side-chain packing are shown in dark blue color. The side-chains of W214 and C34 are shown in red and green spheres, respectively.

and 4.02×10^9 M^{-1} s^{-1}, respectively (Figure 5A, inset). The values of k_q decrease to 0.69×10^9 and 0.98×10^9 M^{-1} s^{-1} in the N states at pH 5 and pH 7, respectively, due to the protection provided by the protein structure around W214. The mean value of k_q is ~5-fold smaller in the N state as compared to the U state. However, the errors associated with the determination of the values of K_{SV} and τ_0 are very small (Table S2). Hence, the errors in the values of k_q are unlikely to be more than ±10%, and ~5-fold dynamic range in the value of k_q is quite significant. We observed that the value of k_q in the I state is 0.73×10^9 M^{-1} s^{-1}, which is similar to that in the N state (Figure 5A, inset). These results indicate that in the I state and the N state, W214 is protected against water-solvation to the similar extent. Hence, the loose side-chain packing in the interdomain region in the I state does not expose hydrophobic side-chains to water molecules, and the interdomain region appears to be like a dry globule.

Solvation Dynamics of the Side-Chains in the Interdomain Region is Heterogeneous. It has been predicted that the hydrophobic core of dry globules might resemble a molten liquid due to the disruption of the side-chain packing interactions.[10,17,18] However, it has been very difficult to probe the degree of mobility of the side-chains in the core of the dry globules. We investigated the solvation dynamics of the protein matrix around W214 in the I state using REES experiments.[41,48] In general, the fluorescence emission spectrum of a fluorophore is independent of the excitation wavelength (λ^{ex}) because of the faster relaxation (as compared to the nanosecond fluorescence time scale) of the solvent molecules surrounding it. However, when the distribution of solvent molecules around the fluorophore is heterogeneous (leading to a broad distribution of solute−solvent interaction energies) and the dynamics of solvent molecules is much slower than the fluorescence time scale (like in a viscous solvent), the fluorescence emission spectrum becomes dependent on the λ^{ex}, when excited around the red edge of the absorption spectrum.[41,48] We observed that the λ^{em}_{max} of W214 depends very strongly on the λ^{ex} in the I state but shows no dependence in the N state and the U state, respectively (Figure S4 and Figure 5B). The λ^{em}_{max} of W214 in the I state shifts to 336 nm from 330 nm, when it is excited at 305 nm as compared to 295 nm (Figure 5B). In contrast, the λ^{em}_{max} of W214 remains constant at 340 and 350 nm in the N state and the U state, respectively (Figure 5B, Figure S4), when the λ^{ex} is changed from 295 to 305 nm. These results indicate that water molecules surrounding W214 in the U state and the complex solvation environment around it in the N state are highly dynamic. In the I state, W214 is buried in the hydrophobic core at the interface of domains I and II (Figure 2A and B). Hence, the large magnitude (6 nm) of REES observed in the dry globular I state indicates that the solvation environment created by the dipoles of protein matrix and side-chains around W214 in the interdomain region (i) is highly heterogeneous, (ii) relaxes much slower than the nanosecond fluorescence time scale of W214 (most probably in the tens of μs), and (iii) is similar to a molten and viscous liquid.[48]

The above results suggest that the I state is like a DMG in which the intradomain packing remains similar to the N state but interdomain packing is loose. It should be noted that the three-dimensional structure of HSA contains a network of 34 disulfide bonds that are distributed all over the protein structure except in the interdomain region (Figure S2). The intradomain region in all three domains appears to be stapled by disulfide bonds. It is highly likely that these disulfide stitches stabilize the structure in the intradomain region as compared to the interdomain region in both the N and the I states and contribute in restricting the intradomain side-chain dynamics.

Interdomain Packing Contributes Significantly to Protein Stability. The DMG-like I state has N-like secondary structure and core hydrophobicity, but vdW packing in the interdomain region is loose. Hence, the difference in thermodynamic stability of the I state as compared to the N state can reveal the contribution of interdomain packing interactions in the stability of this protein (Figure 6). As discussed above in Figure 2C and Table 1, $\Delta G^{H_2O}_{NU}$ is 5.3 kcal mol^{-1} and $\Delta G^{H_2O}_{NI}$ is 2.1 kcal mol^{-1}, indicating that a very large

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

fraction (35−40%) of the total change in free energy occurs during the N⇌I transition. One possible explanation for this large change in free energy is that some side-chain packing interactions are loose in parts of the protein other than the interface of domains I and II. For example, our structural model (Figure 6) shows that the movement of domains I and II away from each other could also disrupt some packing interactions between domains I and III. However, the decrease in stability of the I state is not due to the disruption of the intradomain packing interactions as indicated by the results of the CD experiments (Figure 3A and B). Hence, our thermodynamic measurements indicate that interdomain side-chain packing interactions significantly contribute to the stability of the native protein and that small changes in the packing interactions could significantly destabilize the native protein. Because the DMG-like I state is 2.1 kcal mol^{-1} less stable than the N state (Figure 6), it appears that the enthalpic cost of the disruption of interdomain packing interactions during the N⇌I transition is only partially compensated by the increase in conformational entropy of the side-chains.

N⇌I Transition Represents Early Stages of Chemical Denaturation. A three-state fit of the equilibrium unfolding transition at pH 5 also yielded the *m*-values (slope of the linear denaturant dependence of the free energies) for the N⇌I and N⇌U transitions (Figure 2C and Table 1). We observed that the values of m_{NU} and m_{NI} are 3.3 ± 0.3 and 2.3 ± 0.4 kcal mol^{-1} M^{-1}, respectively. The *m*-value changes during structural transitions of proteins are usually interpreted as the change in the solvent-accessible surface area during the transition. The value of m_{NI} corresponds to 65−70% of the total change in *m*-value during the complete unfolding of the protein (m_{NU}). Hence, these results suggest that 65−70% of the protein core is accessible to water in the DMG-like I state. This result is surprising in view of the results of dynamic fluorescence quenching (Figure 5A) and far-UV CD (Figure 3A) experiments, which suggest that in the I state, W214 has N-like solvent accessibility and secondary structure and that the protein−protein H-bonds are not replaced by protein−water H-bonds. It is also important to note that during the N⇌I transition W214 moves to a more hydrophobic (and not hydrated) environment as compared to the N state (Figure 2A and B). One possible origin of the large change in *m*-values during the N⇌I transition could be the direct interaction of the urea molecules with the protein surface during the initial stages of denaturation. It has been shown, both by computer simulations[49−51] and by experiments,[52−54] that denaturants like urea and guanidine hydrochloride unfold proteins by a two-step direct interaction mechanism. In the first step, they preferentially interact with the protein surface as compared to water in the first solvation cell and loosen the protein structure to populate DMG-like structures. Water molecules enter the protein core in the second step to completely unfold the protein. It has been shown for a few proteins, including RNase H[52] and a variant of CTPR protein,[55] that chemical denaturants modulate the equilibrium between N and DMG and that significant changes in the *m*-value occur during the N⇌DMG transition. Taken together, these results suggest that the N⇌I transition observed in this study represents the early stages of chemical denaturation and that the large changes in the *m*-value arise due to the direct interaction of urea molecules with the native HSA, shifting the equilibrium toward the formation of the DMG-like I state. Hence, one should be careful in the interpretation of *m*-value changes when equilibrium intermediates are present on the folding pathways of proteins.

pH-Dependent Modulation of the N State Heterogeneity. We observed that the I state is only populated at pH 5 but not at pH 7. This is interesting because the structure and the thermodynamic stabilities of the N and the U states at pH 5 and pH 7 are almost identical (Figures 1 and 2, Table 1). Although we observe that the equilibrium unfolding transition at pH 7 is apparently two-state, it is possible that the energy of the I state is higher than that of the U state at pH 7, and hence it is not populated (Figure 6). The decrease in pH to 5 stabilizes the I state considerably to be populated at equilibrium but does not affect the energy of the N state and the U state (Figure 6). This proposal is supported by observations on many proteins including barstar,[39] RNase H,[56] and SH3 domain of PI3 kinase[57] where it has been shown that small changes in environmental conditions like single-site mutation and changes in ionic strength can populate high energy intermediate states and convert a two-state folder into a three-state or multistate folder. Hence, our observation suggests that side-chain conformational heterogeneity of the N state ensemble can be modulated by changing the external environmental conditions.

CONCLUSION

In summary, we probed the conformational heterogeneity of the side-chain packing in the native state ensemble of HSA, a multidomain protein, under two different N-like conditions, pH 5 and pH 7, where the protein has similar structure and thermodynamic stability. We observed that the N state ensemble at pH 5 contains a DMG-like I state, which is not populated at pH 7. The I state gets populated to 85 ± 4% at 1.6 M urea. Far- and near-UV CD experiments show that the I state has N-like secondary and tertiary structure. Fluorescence measurements on W214, which is involved in the interdomain side-chain packing and spatially located at the interface of the three domains of the protein, show that interdomain side-chain packing is loose in the I state, and W214 has moved to a more hydrophobic environment as compared to the N state. FRET experiments show that the I state has a larger interdomain distance between domain I and domain II, as compared to the N state. Dynamic fluorescence quenching experiments suggest that the interdomain region in the I state is as dry as the native protein. REES experiments on W214 show that the solvation environment created by the protein matrix and the side-chains around it is like a molten and viscous liquid. A comparison of the thermodynamic stability of the I state and the N state reveals that the interdomain side-chain packing contributes significantly (35−40%) to the stability of the native protein. Our results indicate that the desolvation of the hydrophobic core and development of side-chain packing are two distinct structural events during protein folding and are not coupled processes as commonly believed. Our results support the hypothesis that the N states of proteins might be an ensemble of loosely and alternatively packed DMG-like states,[10,17,18] which allow enough dynamics and volume fluctuations in the N states required for allosteric functions.[28,58] In the case of HSA, this mechanism of dynamic allostery might have physiological and functional importance. The main function of HSA in human serum is to bind and transport ligands of different shapes, sizes, and chemical nature (hormones, fatty acids, bilirubin, etc.). It is possible that the side-chain conformational heterogeneity in the N state ensemble of HSA facilitates fluctuations and interconversion between different DMG-like

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

alternatively packed N states required for binding to a multitude of its ligand partners. In this context, our result that the heterogeneity of the N state ensemble is dependent upon pH suggests that the equilibrium between the N and the DMG forms of HSA can be modulated by the changes in the physiological conditions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b07032.

Structure of HSA showing all of the Phe residues and all of the residues involved in the interdomain packing (Figure S1); structure of HSA showing all of the Cys residues and disulfide bonds (Figure S2); overlap integral data in the N and the I states (Figure S3); REES data of W214 in the I and the U states (Figure S4); Table S1 containing FRET parameters; and Table S2 containing values of K_{SV}, k_q, and τ_0 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: 91-20-25902588. Fax: 91-20-25902615. E-mail: sk.jha@ncl.res.in.

ORCID

Santosh Kumar Jha: 0000-0003-1339-7409

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by an DST-SERB early career research award (project no. ECR/2015/000027) to S.K.J. and CSIR-National Chemical Laboratory (network grant no. CSC 0134 and MLP grant no. 030026). P.M. is a recipient of a Senior Research Fellowship by University Grants Commission, India.

■ ABBREVIATIONS

N, native; DMG, dry molten globules; I, intermediate; vdW, van der Waals; HSA, human serum albumin; FRET, fluorescence resonance energy transfer; REES, red-edge excitation shift; CD, circular dichroism; 1,5-IAEDANS, 5-((((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid; GdmCl, guanidine hydrochloride; W214, tryptophan; F, fast-migrating; B, basic; U, unfolded; MRE, mean residue ellipticity; λ_{max}^{em}, wavelength of maximum fluorescence emission; FU, fraction of the unfolded protein; H-bonds, hydrogen bonds; λ^{ex}, excitation wavelength; D, donor; A, acceptor; HSA-IAEDANS, 1,5-IAEDANS labeled protein

■ REFERENCES

(1) Richards, F. M. The Interpretation of Protein Structures: Total Volume, Group Volume Distributions and Packing Density. *J. Mol. Biol.* **1974**, *82*, 1−14.

(2) Chothia, C. Structural Invariants in Protein Folding. *Nature* **1975**, *254*, 304−308.

(3) Chothia, C.; Janin, J. Principles of Protein-Protein Recognition. *Nature* **1975**, *256*, 705−708.

(4) Richards, F. M. Areas, Volumes, Packing and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.

(5) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585−590.

(6) Wuthrich, K.; Wagner, G. Internal Motion in Globular Proteins. *Trends Biochem. Sci.* **1978**, *3*, 227−230.

(7) Zhou, Y.; Vitkup, D.; Karplus, M. Native Proteins are Surface-Molten Solids: Application of the Lindemann Criterion for the Solid Versus Liquid State. *J. Mol. Biol.* **1999**, *285*, 1371−1375.

(8) Jha, S. K.; Udgaonkar, J. B. Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry. *J. Biol. Chem.* **2007**, *282*, 37479−37491.

(9) Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; Garcia-Moreno, B. E.; Royer, C. A. Cavities Determine the Pressure Unfolding of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6945−6950.

(10) Bhattacharyya, S.; Varadarajan, R. Packing in Molten Globules and Native States. *Curr. Opin. Struct. Biol.* **2013**, *23*, 11−21.

(11) Bowman, G. R.; Geissler, P. L. Extensive Conformational Heterogeneity within Protein Cores. *J. Phys. Chem. B* **2014**, *118*, 6417−6423.

(12) DuBay, K. H.; Bowman, G. R.; Geissler, P. L. Fluctuations within Folded Proteins: Implications for Thermodynamic and Allosteric Regulation. *Acc. Chem. Res.* **2015**, *48*, 1098−1105.

(13) Elber, R.; Karplus, M. Multiple Conformational States of Proteins: A Molecular Dynamics Analysis of Myoglobin. *Science* **1987**, *235*, 318−321.

(14) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964−972.

(15) Tzeng, S. R.; Kalodimos, C. G. Protein Activity Regulation by Conformational Entropy. *Nature* **2012**, *488*, 236−240.

(16) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The Role of Conformational Entropy in Molecular Recognition by Calmodulin. *Nat. Chem. Biol.* **2010**, *6*, 352−358.

(17) Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry Molten Globule Intermediates and the Mechanism of Protein Unfolding. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2725−2737.

(18) Baldwin, R. L.; Rose, G. D. Molten Globules, Entropy-Driven Conformational Change and Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 4−10.

(19) Neumaier, S.; Kiefhaber, T. Redefining the Dry Molten Globule State of Proteins. *J. Mol. Biol.* **2014**, *426*, 2520−2528.

(20) Shakhnovich, E. I.; Finkelstein, A. V. Theory of Cooperative Transitions in Protein Molecules. I. Why Denaturation of Globular Protein is a First-Order Phase Transition. *Biopolymers* **1989**, *28*, 1667−1680.

(21) Finkelstein, A. V.; Shakhnovich, E. I. Theory of Cooperative Transitions in Protein Molecules. II. Phase Diagram for a Protein Molecule in Solution. *Biopolymers* **1989**, *28*, 1681−1694.

(22) Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. Direct NMR Evidence for an Intermediate Preceding the Rate-Limiting Step in the Unfolding of Ribonuclease A. *Nature* **1995**, *375*, 513−515.

(23) Hoeltzli, S. D.; Frieden, C. Stopped-Flow NMR Spectroscopy: Real-Time Unfolding Studies of 6−19F-Tryptophan-Labeled Escherichia Coli Dihydrofolate Reductase. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9318−9322.

(24) Jha, S. K.; Dhar, D.; Krishnamoorthy, G.; Udgaonkar, J. B. Continuous Dissolution of Structure During the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11113−11118.

(25) Jha, S. K.; Udgaonkar, J. B. Direct Evidence for a Dry Molten Globule Intermediate During the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12289−12294.

(26) Reiner, A.; Henklein, P.; Kiefhaber, T. An Unlocking/Relocking Barrier in Conformational Fluctuations of Villin Headpiece Subdomain. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4955−4960.

(27) Sarkar, S. S.; Udgaonkar, J. B.; Krishnamoorthy, G. Unfolding of a Small Protein Proceeds Via Dry and Wet Globules and a Solvated Transition State. *Biophys. J.* **2013**, *105*, 2392−2402.

(28) Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J. Am. Chem. Soc.* **2017**, *139*, 3599−3602.

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

(29) Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled Motion in Proteins Revealed by Pressure Perturbation. *J. Am. Chem. Soc.* **2012**, *134*, 8543−8550.
(30) Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J. Phys. Chem. Lett.* **2016**, *7*, 173−179.
(31) Painter, L.; Harding, M. M.; Beeby, P. J. Synthesis and Interaction with Human Serum Albumin of the First 3,18-Disubstituted Derivative of Bilirubin. *J. Chem. Soc., Perkin Trans. 1* **1998**, 3041−3044.
(32) Pace, C. N. Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol.* **1986**, *131*, 266−280.
(33) Jha, A.; Udgaonkar, J. B.; Krishnamoorthy, G. Characterization of the Heterogeneity and Specificity of Interpolypeptide Interactions in Amyloid Protofibrils by Measurement of Site-Specific Fluorescence Anisotropy Decay Kinetics. *J. Mol. Biol.* **2009**, *393*, 735−752.
(34) Lillo, M. P.; Beechem, J. M.; Szpikowska, B. K.; Sherman, M. A.; Mas, M. T. Design and Characterization of a Multisite Fluorescence Energy-Transfer System for Protein Folding Studies: A Steady-State and Time-Resolved Study of Yeast Phosphoglycerate Kinase. *Biochemistry* **1997**, *36*, 11261−11272.
(35) Saxena, A. M.; Udgaonkar, J. B.; Krishnamoorthy, G. Characterization of Intra-Molecular Distances and Site-Specific Dynamics in Chemically Unfolded Barstar: Evidence for Denaturant-Dependent Non-Random Structure. *J. Mol. Biol.* **2006**, *359*, 174−189.
(36) Hudson, E. N.; Weber, G. Synthesis and Characterization of Two Fluorescent Sulfhydryl Reagents. *Biochemistry* **1973**, *12*, 4154−4161.
(37) Atanasiu, C.; Su, T. J.; Sturrock, S. S.; Dryden, D. T. Interaction of the ocr Gene 0.3 Protein of Bacteriophage T7 with EcoKI Restriction/Modification Enzyme. *Nucleic Acids Res.* **2002**, *30*, 3936−3944.
(38) Street, T. O.; Courtemanche, N.; Barrick, D. Protein Folding and Stability Using Denaturants. *Methods Cell Biol.* **2008**, *84*, 295−325.
(39) Nath, U.; Udgaonkar, J. B. Perturbation of a Tertiary Hydrogen Bond in Barstar by Mutagenesis of the Sole His Residue to Gln Leads to Accumulation of at Least One Equilibrium Folding Intermediate. *Biochemistry* **1995**, *34*, 1702−1713.
(40) Wani, A. H.; Udgaonkar, J. B. Revealing a Concealed Intermediate That Forms after the Rate-Limiting Step of Refolding of the SH3 Domain of PI3 Kinase. *J. Mol. Biol.* **2009**, *387*, 348−362.
(41) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Springer: Singapore, 2006.
(42) Amiri, M.; Jankeje, K.; Albani, J. R. Origin of Fluorescence Lifetimes in Human Serum Albumin. Studies on Native and Denatured Protein. *J. Fluoresc.* **2010**, *20*, 651−656.
(43) Amiri, M.; Jankeje, K.; Albani, J. R. Characterization of Human Serum Albumin Forms with pH. Fluorescence Lifetime Studies. *J. Pharm. Biomed. Anal.* **2010**, *51*, 1097−1102.
(44) Carter, D. C.; Ho, J. X. Structure of Serum Albumin. *Adv. Protein Chem.* **1994**, *45*, 153−203.
(45) Dockal, M.; Carter, D. C.; Ruker, F. Conformational Transitions of the Three Recombinant Domains of Human Serum Albumin Depending on pH. *J. Biol. Chem.* **2000**, *275*, 3042−3050.
(46) Eftink, M. R.; Ghiron, C. A. Dynamics of a Protein Matrix Revealed by Fluorescence Quenching. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 3290−3294.
(47) Strambini, G. B.; Gonnelli, M. Fluorescence Quenching of Buried Trp Residues by Acrylamide does Not Require Penetration of the Protein Fold. *J. Phys. Chem. B* **2010**, *114*, 1089−1093.
(48) Demchenko, A. P. Site-Selective Red-Edge Effects. *Methods Enzymol.* **2008**, *450*, 59−78.
(49) Mountain, R. D.; Thirumalai, D. Molecular Dynamics Simulations of End-to-End Contact Formation in Hydrocarbon Chains in Water and Aqueous Urea Solution. *J. Am. Chem. Soc.* **2003**, *125*, 1950−1957.
(50) Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. Urea Denaturation by Stronger Dispersion Interactions with Proteins than Water Implies a 2-Stage Unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16928−16933.
(51) Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein Folding: From Theory to Practice. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22−29.
(52) Jha, S. K.; Marqusee, S. Kinetic Evidence for a Two-Stage Mechanism of Protein Denaturation by Guanidinium Chloride. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4856−4861.
(53) Dasgupta, A.; Udgaonkar, J. B.; Das, P. Multistage Unfolding of an SH3 Domain: An Initial Urea-Filled Dry Molten Globule Precedes a Wet Molten Globule with Non-Native Structure. *J. Phys. Chem. B* **2014**, *118*, 6380−6392.
(54) de Oliveira, G. A.; Silva, J. L. A Hypothesis to Reconcile the Physical and Chemical Unfolding of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2775−E2784.
(55) Cohen, S. S.; Riven, I.; Cortajarena, A. L.; De Rosa, L.; D'Andrea, L. D.; Regan, L.; Haran, G. Probing the Molecular Origin of Native-State Flexibility in Repeat Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 10367−10373.
(56) Spudich, G. M.; Miller, E. J.; Marqusee, S. Destabilization of the *Escherichia Coli* RNase H Kinetic Intermediate: Switching between a Two-State and Three-State Folding Mechanism. *J. Mol. Biol.* **2004**, *335*, 609−618.
(57) Dasgupta, A.; Udgaonkar, J. B. Four-State Folding of a SH3 Domain: Salt-Induced Modulation of the Stabilities of the Intermediates and Native State. *Biochemistry* **2012**, *51*, 4723−4734.
(58) Cooper, A. Thermodynamic Fluctuations in Protein Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 2740−2741.

DOI: 10.1021/acs.jpcb.7b07032
J. Phys. Chem. B 2017, 121, 9336−9347

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2017, **19**, 30207

A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein†

Nirbhik Acharya, Prajna Mishra and Santosh Kumar Jha *

Received 27th September 2017,
Accepted 27th October 2017

DOI: 10.1039/c7cp06614g

rsc.li/pccp

The nature of the initial structural events during the base-induced unfolding of the native (N) state of proteins is poorly understood. Combining site-specific fluorescence resonance energy transfer, size exclusion chromatography, dynamic fluorescence quenching, red-edge excitation shift and circular dichroism spectroscopy, we show here that an early intermediate during the base-induced unfolding of a multidomain protein, *i.e.*, the B form, has features of a dry molten globule. We show that the N $\rightleftharpoons$ B transition involves protein expansion and loosening of packing of inter-domain helices near domains I and II without the disruption of intra-domain packing or any change in hydration of the inter-domain region which resembles a molten hydrocarbon. Surprisingly, the disruption of inter-domain packing accounts for 40–45% of the total change in free energy of complete unfolding. Our results show that the disruption of van der Waals packing can be decoupled in different regions of a protein and could occur prior to hydrophobic solvation during base-induced unfolding, challenging the existing notion.

Introduction

Basic pH leads to partial or complete unfolding of proteins and is an important factor in protein function and many industrial and biotechnological applications of proteins.[1–9] However, we do not understand the molecular mechanism and sequence of structural events that initiate the base-induced unfolding of proteins. Two different models have been proposed to describe how proteins begin to unfold. The traditional model envisages protein unfolding as a cooperative process where the dissolution of the native (N) structure occurs concomitantly with the penetration of water molecules in the hydrophobic core.[10–13] The alternative model, based on the theoretical phase diagram of proteins, postulates that unfolding begins with the expansion of the protein into a dry molten globule (DMG) state in which the van der Waals (vdW) packing interactions are broken but the hydrophobic core is dehydrated and the N-like main-chain hydrogen bonding and overall tertiary structure are retained.[14–19] According to this model, core hydration and global structural dissolution occur in the second and final step of protein unfolding. However, experimental evidence for a DMG-like intermediate state during the base-induced unfolding of proteins has been rare.[20]

Human serum albumin (HSA) is the most abundant protein in human blood plasma and performs multiple functions including ligand transport[21] and maintenance of blood pH.[22] It is a well-characterized three domain (I, II and III) protein and has been known to undergo a physiologically important base-induced structural transition, from the N to the basic (B) form, for the last five decades.[21,23,24] The B form of serum albumins impacts their ligand binding properties and has physiological importance in the transport of many medicinal drugs in blood.[3,24,25] Surprisingly, the details of the inner molecular structure of the B form of HSA and the mechanism of its formation are very poorly understood. One major reason is that the B form is so similar in structure to the N form that the structural differences are silent to most of the global spectroscopic probes of the protein structure. Various fundamental questions remain unanswered. For example, we do not understand (i) the nature of the structural difference between the B form and the N form; (ii) whether the B form resembles a DMG-like state; (iii) whether the structural loosening and core solvation are coupled during N $\rightleftharpoons$ B transition; (iv) how different the side-chain packing interactions are in the core of the B form compared to the N form; and (v) how the small structural differences between the N and the B form are important for function.

In this study, we have used a battery of site-specific probes to investigate the base-induced partial unfolding of HSA. The simultaneous application of multiple spectroscopic probes, including fluorescence resonance energy transfer (FRET), size exclusion chromatography, dynamic fluorescence quenching, red-edge excitation shift (REES) and near- and far-UV circular

Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Pune 411008, India. E-mail: sk.jha@ncl.res.in

† Electronic supplementary information (ESI) available: Data on calculation of R_0; standard calibration curves for determination of apparent molecular weight and hydrodynamic radius; fluorescence lifetime decays; fluorescence scans for REES; crystal structure of HSA showing the location of Phe residues and disulfide bonds; tables containing FRET parameters and values of K_{sv}, k_q, and τ_0. See DOI: 10.1039/c7cp06614g

dichroism (CD) allowed us to site-specifically unveil the nature of structural events that initiate the formation of the B form and its structure in terms of changes in inter-atomic distances, side-chain packing, protein expansion, core hydration, core flexibility, and global secondary and tertiary structure. Previous studies have suggested that in the B form of HSA, domain III remains fully structured, but there are subtle tertiary structural changes in domains I and II.[26] Here, we report that in the B form, domains I and II of HSA have DMG-like characteristics near the inter-domain interface. Our results shed important light on the mechanism of base-induced protein unfolding and suggest that the vdW packing interactions might be an important factor in protein stability.

Experimental

Reagents, chemicals, buffers and experimental conditions

Human serum albumin (99% pure, fatty acid free, globulin free) was procured from Alfa Aesar and used without further purification. The concentration of HSA was determined spectroscopically by measuring the absorbance at 280 nm, using a molar extinction coefficient of 36 500 M^{-1} cm^{-1}.[27] Urea (ACS grade, $\geq$99% pure) was obtained from Alfa Aesar and 5-((((2-iodoacetyl)amino)ethyl)-amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) from Life Technologies. Acrylamide, guanidine hydrochloride (GdmCl) and other chemicals and reagents were purchased from Sigma.

For the pH titration study, a universal buffer (20 mM sodium acetate, 20 mM sodium phosphate and 20 mM sodium borate) was used in the pH range of 5–12. The pH was adjusted with the addition of HCl or NaOH. For all other experiments, buffers for the native (N) form at pH 7 and for the basic (B) form at pH 11 were composed of 20 mM sodium phosphate. The unfolding buffers at pH 7 and pH 11 were composed of 20 mM sodium phosphate and 9 M urea. All the buffers were filtered through 0.2 μm filters before use. All experiments were performed at room temperature unless otherwise specified. The concentrations of urea and GdmCl solutions were determined by measuring the refractive index.[28]

Spectroscopic instruments

All the absorption spectra were acquired on a Perkin Elmer Lambda 650 UV/Vis spectrometer using a quartz cell of 1 cm path length. All the fluorescence spectra were collected on a Perkin Elmer fluorescence spectrometer LS 55 using a fluorescence quartz cell of 1 cm path length. The collection of all the circular dichroism (CD) spectra was performed on a Jasco J-815 CD spectrometer. CD compatible quartz cells of 0.1 cm and 1 cm path length were used for far- and near-UV CD, respectively. The fluorescence lifetime decays were measured on a time-correlated single photon counting (TCSPC) setup, Deltaflex, from Horiba Scientific. The refractive index of buffers, wherever required, was measured on an Abbe refractometer from Rajdhani Scientific Instruments.

Preparation of IAEDANS labeled HSA

HSA was labeled at the free cysteine residue, C34, with 1,5-IAEDANS as described previously.[29] The extent of labeling was calculated by estimating the concentration of total protein (HSA + HSA-IAEDANS) and the labeled protein (HSA-IAEDANS) spectroscopically. We determined the concentration of HSA-IAEDANS from the absorbance of C34-IAEDANS at 337 nm using a molar extinction coefficient of 4500 M^{-1} cm^{-1}.[29] The total protein concentration was determined by measuring the absorbance of the protein at 280 nm, using the extinction coefficient of 36 500 M^{-1} cm^{-1}. The extinction coefficient of IAEDANS at 280 nm is $\sim$1000 M^{-1} cm^{-1}.[30] Since the extinction coefficient of IAEDANS dye at 280 nm is nearly 35 times less than the extinction coefficient of HSA, the contribution of the dye to the 280 nm absorbance will be negligible ($\sim$3%) if the protein is fully labeled. The measured extent of labeling of the labeled protein was >95%. It is important to note that our assumption is conservative and gave the lower limit of the amount of labeling and the actual percentage will be 2–3% larger than this.

pH titration monitored by fluorescence and CD

The protein samples were incubated at different pH values and equilibrated by keeping overnight at room temperature. For the unlabeled protein, W214 was excited using an excitation wavelength of 295 nm and the emission spectra were collected from 310 to 420 nm. For HSA-IAEDANS, the protein samples were excited at 337 nm and the emission spectra were collected from 350 to 550 nm. All the fluorescence spectra were recorded with an excitation slit of 5–7 nm and a scan speed of 100 nm min^{-1}. All the far-UV CD and the near-UV CD spectra were respectively measured in the 190–250 nm and 250–300 nm wavelength range. All the CD spectra were collected with a bandwidth of 2 nm and a scan speed of 100 nm min^{-1}. The protein concentration used for the fluorescence and the far-UV CD experiments was 4–5 μM and for the near-UV CD experiments, it was 20 μM.

For fluorescence monitored pH titration, the change in the fluorescence signal at 340 nm was observed to occur in a single sigmoidal step. The data were fitted to a model, in which the structural transition from the N form to the B form is coupled to a single deprotonation step, given by a transformed Henderson–Hasselbalch equation,[31]

$$Y_{obs} = \frac{Y_P + Y_D 10^{(pH-pH_m)}}{1 + 10^{(pH-pH_m)}} \quad (1)$$

where Y_{obs} denotes the observed fluorescence signal for a given pH value, Y_P and Y_D are the signals of the protonated and deprotonated forms and pH_m is the midpoint of the transition.

FRET between W214 and C34-IAEDANS and the analysis of FRET data

For FRET measurements, W214 in HSA and HSA-IAEDANS was excited at 295 nm and all the emission spectra were collected identically. The FRET distance (R) between W214 and C34-IAEDANS was calculated from the FRET efficiency (E) and Forster's distance (R_0) as described in the Results and discussion section. The Forster's distance (R_0) between the FRET pair in each condition was calculated using the equation,[32]

$$R_0 = 0.211\left[Q_D J \kappa^2 n^{-4}\right]^{\frac{1}{6}} \quad (2)$$

where Q_D denotes the quantum yield of the donor, J represents the overlap integral between the donor's emission spectrum and acceptor's excitation spectrum, κ^2 is the orientation factor of the fluorophores and n represents the refractive index of the medium.[32] We experimentally determined the values of Q_D as described previously.[29] The values of Q_D in the N form and the B form are 0.31 and 0.23, respectively. The overlap integral J was calculated by

$$J = \frac{\int F(\lambda)\varepsilon(\lambda)\lambda^4 \mathrm{d}\lambda}{\int F(\lambda)\mathrm{d}\lambda}\left(\mathrm{M}^{-1}\ \mathrm{cm}^{-1}\ \mathrm{nm}^4\right) \quad (3)$$

In the above equation, $F(\lambda)$ and $\varepsilon(\lambda)$ represent the fluorescence emission spectra of W214 and the absorbance spectra of C34-IAEDANS (Fig. S1 and Table S1, ESI†). The orientation of the donor and the acceptor fluorophore was assumed to be random and hence the value of κ^2 was taken as 2/3. The observed value of n for the N form at pH 7 and the B form at pH 11 was 1.332. The calculated values of Forster's distance and the FRET distance between W214 and C34-IAEDANS, along with other FRET parameters are given in Table S1 (ESI†).

Size exclusion chromatography

All the size exclusion chromatography (SEC) experiments were performed on a GE AKTA Pure FPLC system. For SEC in the N form and the B form, HSA was dissolved in pH 7 and in pH 11 buffer, respectively. SEC was performed by loading 100 μM protein onto a GE Superdex 75 10/300 GL high performance gel filtration column, pre-equilibrated at the respective pH. Due to the high concentration of protein used for SEC, we observed 5–12% of the dimer in the resultant chromatograms. For a better comparison between the N form and the B form, the SEC chromatograms were fitted to a sum of two Gaussians and monomer peaks were compared.

For the determination of the apparent molecular weight of the N form and the B form, we constructed a calibration curve between the partition coefficient (K_{av}) of five different standard biomolecules (vitamin B12, aprotinin, ribonuclease A, ovalbumin and bovine serum albumin) and their known molecular weights (M_w) (Fig. S2A, ESI†), as described in the GE size exclusion chromatography handbook provided by the manufacturer. The values of K_{av} were calculated using the following equation:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad (4)$$

where V_e is the elution volume for the protein, V_0 is the column void volume and V_t is the total bed volume. The values of V_0 and V_t were taken from the column specification sheet. The V_e values of the standard biomolecules were taken from the manufacturer provided manual of GE Superdex 75 10/300 GL high performance gel filtration column. We then calculated the K_{av} values for the N form and the B form of HSA from their corresponding experimentally measured V_e values and determined the apparent molecular weight of the N form and the B form from the calibration curve.

Similarly, for the determination of the hydrodynamic radius of the N form and the B form, we constructed a calibration curve between K_{av} values of three different standard proteins (ribonuclease A, ovalbumin and bovine serum albumin) and their previously reported hydrodynamic radii[33] (Fig. S2B, ESI†). Using this calibration curve and the above calculated K_{av} values of the N form and the B form, we determined the hydrodynamic radius of the N form and the B form of HSA.

Acrylamide quenching experiments

HSA and HSA-IAEDANS were incubated at pH 7 and pH 11 and their respective unfolding buffers. Appropriate volumes of acrylamide were added just before collecting the fluorescence emission spectra from a freshly prepared 2 M acrylamide stock solution. For HSA, W214 was excited at 295 nm and the emission spectra were collected from 310 to 420 nm. For HSA-IAEDANS, C34-IAEDANS was excited at 337 nm and the emission spectra were collected from 350 to 550 nm.

The values of Stern–Volmer constant (K_{sv}) and the intensity averaged fluorescence lifetime (τ_0) were used to determine the bimolecular quenching constant (k_q)under each condition. For a 9 M urea sample, a correction for viscosity was made, as described previously.[29] The values of K_{sv}, τ_0 and k_q under different conditions are listed in Table S2 (ESI†).

Fluorescence lifetimes

The fluorescence lifetime decays of C34-IAEDANS under different conditions were measured using an excitation laser source of 370 nm (IRF is 196 ps) and collecting the emission decay at 480 nm (Fig. S3, ESI†). The deconvolution and fitting of fluorescence decay into three exponentials were performed on DAS6 analysis software (Horiba Scientific). The intensity averaged fluorescence lifetimes (τ_0) (Table S2, ESI†) and mean fluorescence lifetime (τ_m) were calculated using the following equations:

$$\tau_0 = \frac{\sum_i \alpha_i \tau_i^2}{\sum_i \alpha_i \tau_i} \quad (5)$$

$$\tau_m = \sum \alpha_i \tau_i \quad (6)$$

where α_i is the fractional amplitude of the respective fluorescence lifetime $\tau_i\left(\sum_i \alpha_i = 1\right)$. The values of τ_0 for W214 have been determined from the fluorescence lifetime measurements reported in previous studies.[34,35] These values are listed in Table S2 (ESI†).

REES experiments

For HSA-IAEDANS, the fluorescence emission spectra of C34-IAEDANS were collected by exciting the fluorophore at different wavelengths ranging from 337 nm to 407 nm. For HSA, the fluorescence emission spectra were recorded by exciting W214 at different wavelengths ranging from 295 nm to 305 nm. For both HSA and HSA-IAEDANS, the concentration of protein used was 6–12 μM. All the fluorescence emission spectra were recorded at a scan speed of 50 nm min^{-1}, averaged for three accumulations and subtracted for background buffer intensities.

Urea-induced equilibrium unfolding transitions

For urea-induced equilibrium unfolding transitions, the protein samples were incubated for 3 hours at pH 7 and pH 11 at different concentrations of urea. The concentration of protein used for the fluorescence and far-UV CD experiments was 4 μM. The fluorescence spectra of W214 were obtained by exciting W214 at 295 nm and collecting the emission spectra from 310 nm to 420 nm. The far-UV CD spectra of the same samples were measured in the wavelength range of 200–250 nm. The fluorescence and far-UV CD monitored equilibrium unfolding transitions were analyzed using a two-state Z ⇌ U model[36] (Z is N at pH 7 and B at pH 11) and the data were fitted to the equation,

$$y_{obs} = \frac{y_Z + y_U e^{\frac{-\Delta G_{Z-U} + m_{Z-U}[D]}{RT}}}{1 + e^{\frac{-\Delta G_{Z-U} + m_{Z-U}[D]}{RT}}} \quad (7)$$

where y_{obs} is the observed fluorescence/CD signal; y_Z and y_U are the signals of the native and the unfolded states, respectively; G_{Z-U} and m_{Z-U} are the standard free energy and the slope of the transition, respectively.

Results and discussion

Base-induced structural transition

We monitored the change in the tertiary and the secondary structure of the protein during the base-induced structural transition using intrinsic tryptophan fluorescence and far-UV CD. HSA has a single tryptophan residue, W214, which is located in the first helix of domain II and participates in the inter-domain side-chain packing interactions that hold the three domains together (Fig. 1A). We observed that the fluorescence intensity of W214, reporting upon the local tertiary environment, changes in a sigmoidal fashion upon increasing the pH from 5 to 12 (Fig. 1B), indicating that protein undergoes a base-induced structural transition coupled to the deprotonation of a single ionizable group. The midpoint of the transition is at pH 10.2. It has been reported that the N form of HSA transforms to the B form above pH 8 (N ⇌ B transition)[26] and our sigmoidal pH transition data monitored by W214 fluorescence support this conclusion. It is important to rule out that the quenching of W214 fluorescence at a higher pH is not due to the deprotonation of the indole group or the solvent induced change in its fluorescence quantum yield. In order to do this, we compared the pH-dependent change in W214 fluorescence to the pH-dependent change in the fluorescence of an N-terminal and C-terminal blocked analogue of L-tryptophan, *N*-acetyl-L-tryptophanamide (NATA) (Fig. 1B). We observed that the change in the fluorescence intensity of NATA upon increasing the pH from 7 to 11 is only 10–15% of the total change in W214 fluorescence in this pH range (Fig. 1B). This comparison indicates that the observed change in W214 fluorescence upon increasing the pH is primarily due to the partial unfolding and disruption of the tertiary structure near W214. This result is in accordance with the results of a previous study on a few globular proteins, including serum albumin, which showed that the pH dependent changes in tryptophan fluorescence in the pH range of 7–12 are a result of structural changes near the fluorophore.[37] Very interestingly, we observed that the secondary structure content of the protein, monitored by far-UV CD, does not change and remains almost constant during the N ⇌ B transition (Fig. 1B). Hence, the B form appears to be a state of HSA in which the secondary structure is similar to the N form but the tertiary structure near W214 is disrupted. The experiments presented below compare the structural properties of the N form (HSA in pH 7 buffer) and the B form (HSA in pH 11 buffer).

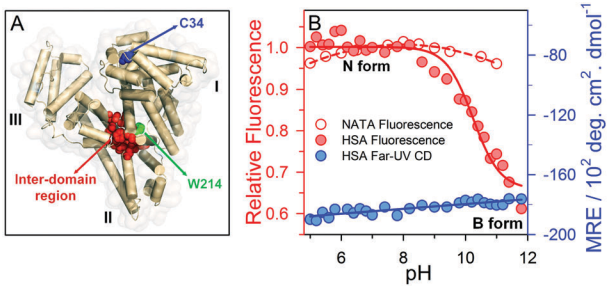


Fig. 1 The base-induced N ⇌ B structural transition is accompanied by a change in the tertiary structure without alteration in the secondary structure. (A) Structure of HSA, highlighting the free cysteine residue C34 (blue) in domain I, the tryptophan residue W214 (green) in domain II and the inter-domain region (red) drawn from PDB file 1AO6 using the program PyMOL. (B) The change in the fluorescence emission of W214 and NATA at 340 nm (left *y*-axis) and the change in mean residual ellipticity at 222 nm (right *y*-axis) are plotted against pH. The solid red and blue lines through the data are fits to a two-state pH transition model (eqn (1)) and a straight line, respectively. The dashed red line through the data is drawn to guide the eyes.

Structural expansion in the B form

Because W214 is located in the inter-domain region at the interface of domain I and domain II, the disruption of the tertiary structure near it will likely lead to the movement of domain I and domain II away from each other. To probe this further at the site-specific level, we measured the inter-domain distances between domains I and II in the N form and the B form using FRET methodology. For FRET measurements, we utilized W214 in domain II as the donor fluorophore. HSA has a single free cysteine, C34, located at the N-terminal of helix 3 in domain I (HSA contains 35 cysteine residues but 34 of them are involved in 17 disulfide bonds). We labeled C34 with the fluorescent dye 5-((((2-iodoacetyl)amino)-ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS), which served as the acceptor of W214 fluorescence. We have reported earlier that C34 of HSA can be efficiently and quantitatively labeled with 1,5-IAEDANS and that the structure and thermodynamic stability of the 1,5-IAEDANS labeled protein (HSA-IAEDANS) are very similar to the unlabeled protein (HSA).[29] Hence, the data on HSA and HSA-IAEDANS can be compared directly as required for the FRET measurements. The fluorescence emission spectrum of W214 overlaps with the absorbance spectrum of C34-IAEDANS (Fig. S1, ESI†), and they form an efficient FRET pair.[29] We observed that in the N form of HSA-IAEDANS, where W214 and C34-IAEDANS are nearer to each other, the fluorescence of W214 is quenched significantly by C34-IAEDANS (Fig. 2A). In contrast, the level of quenching

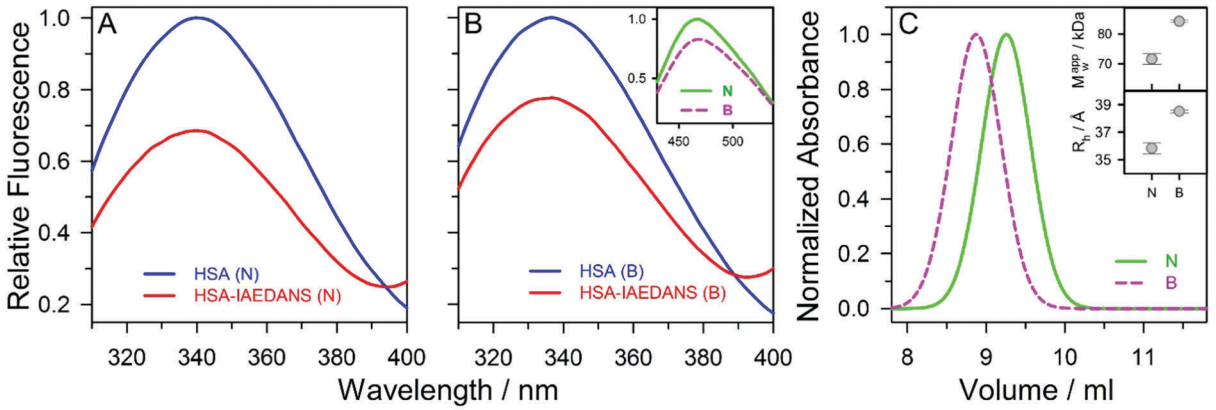


Fig. 2 The B form has a larger inter-domain distance and is globally expanded compared to the N form. The intra-molecular distance between W214 and C34-IAEDANS is monitored by FRET in the N form (A) and in the B form (B). The change in the fluorescence spectra of W214 in HSA and HSA-IAEDANS are shown. The inset in panel B compares the change in the fluorescence emission spectrum of C34-IAEDANS in the N form and the B form after excitation at 295 nm. (C) Size exclusion chromatography of the N form (green solid line) and the B form (pink dashed line). The top and the bottom insets in panel C, respectively, compare the apparent molecular weight (M_w^{app}) and the hydrodynamic radius (R_h) of the N form and the B form. The spread in the values of apparent molecular weight and hydrodynamic radius were estimated from two independent measurements. (N form: HSA at pH 7; B form: HSA at pH 11).

markedly decreases in the B form compared to that in the N form (Fig. 2B), indicating that the donor–acceptor (D–A) distance is larger in the B form compared to the N form. We quantified the FRET efficiency (E) and the D–A distance (R) in the N form and the B form using eqn (8) and (9),[32]

$$E = 1 - \frac{F_{DA}}{F_D} \quad (8)$$

$$R = R_0\left(\frac{1-E}{E}\right)^{\frac{1}{6}} \quad (9)$$

In eqn (8), F_D and F_{DA} represent the fluorescence signal of the donor in the absence and the presence of the acceptor, respectively. In eqn (9), R_0 is the Forster's distance and we have experimentally determined the mean values of R_0 (Fig. S1 and Table S1, ESI†) in the N form and the B form to be 25.8 Å and 24.5 Å, respectively. We observed that the D–A distance in the N form is 29.6 ± 0.1 Å and in the B form is 31.0 ± 0.4 Å. The mean D–A distance in the B form is only 1.4 Å larger than that in the N form. However, this is a significant difference because of the following reasons: (i) the decreased FRET, and the increased D–A distance, in the B form compared to the N form does not only result in the decrease in the extent of quenching of the donor (W214) fluorescence (Fig. 2A and B), but also leads to a significant decrease in the acceptor (C34-IAEDANS) fluorescence (Fig. 2B (inset)). These results indicate that the extent of energy transfer from the donor to the acceptor is less in the B form compared to the N form, due to the relatively large D–A distance in the B form; (ii) the measured FRET distances in the N form and the B form are very near to the experimentally determined R_0 where the extent of energy transfer is most sensitive to the changes in D–A distances (between 0.5 R_0 to 1.5 R_0); (iii) we have determined the standard error in the measurement of D–A distances from at least three separate experiments and the standard error is limited to 0.1 Å in the N form and 0.4 Å in the B form. It is important to note that the errors in the measurement of D–A distances from FRET are typically small due to the sixth root dependence of R on $\left(\frac{1-E}{E}\right)$ (eqn (9)); moreover, R also depends on R_0 (eqn (9)) which itself has a sixth root dependence on other FRET parameters (eqn (2)); (iv) most importantly, we observed that the B form also has a larger hydrodynamic radius (and hence a larger hydrodynamic volume) than the N form as determined by size exclusion chromatography (Fig. 2C, see below). Hence, these results indicate that the D–A distance between domain I and domain II in the B form is 1.4 Å larger compared to the N form along the axis connecting W214 and C34-IAEDANS. The strength of vdW packing interactions between atoms depends steeply on inter-atomic distances. Hence, the larger inter-atomic distances between the helices of domains I and II (domain I_{Helix3}–domain II_{Helix1}) in the B form indicate that inter-domain helices are loosely packed and that some of the inter-domain packing interactions between domains I and II are severely disrupted, compared to the N form.

In order to investigate how the disruption of inter-domain packing interactions and the movement of domain I and domain II away from each other in the B form affect the overall dimensions of the protein, we measured the apparent molecular weight and the hydrodynamic radius of the N form and the B form using analytical size exclusion chromatography on a GE Superdex 75 10/300 GL high performance gel filtration column (Fig. 2C). We observed that the mean elution volume for the N form is 9.25 ± 0.05 ml which decreases to 8.86 ± 0.01 ml for the B form, indicating that the B form has a larger hydrodynamic volume and apparent molecular weight than the N form. Based on the known elution volumes and molecular weights of five different standard biomolecules, we constructed a calibration curve (Fig. S2A, ESI†) to determine the apparent molecular weight of the N form and the B form (Fig. 2C (top inset)) using

the data on their elution volume (Fig. 2C). The apparent molecular weight of the N form is 71.6 ± 1.8 kDa which increases considerably to 84.3 ± 0.4 kDa in the B form (Fig. 2C (top inset)). For the determination of the hydrodynamic radius, we constructed a calibration curve based on the known hydrodynamic radius of three different standard proteins (Fig. S2B, ESI†), and determined the hydrodynamic radius of the N form to be 35.8 ± 0.4 Å and the B form to be 38.5 ± 0.1 Å (Fig. 2C (bottom inset)). The B form has 15–20% larger apparent molecular weight and around 7–10% larger hydrodynamic radius than the N form. Hence, these results indicate that the disruption of inter-domain packing interactions in the B form results in the larger hydrodynamic volume and overall expansion of the protein in the B form compared to the N form.

B form retains N-like hydrophobicity

We observed that the expansion of the protein is not accompanied by any change in the hydration of the protein core in domains I and II. We investigated the core hydration by monitoring the change in the fluorescence Stokes shift of C34-IAEDANS and W214 in domains I and II, respectively, which is a very sensitive indicator of the polarity of the medium surrounding the fluorophore. We compared the Stokes shift of C34-IAEDANS and W214 in the N and the B form to that in the urea unfolded U form (9 M urea, pH 7). The presence of 9 M urea in pH 7 buffer completely unfolds the protein and populates the U form in which C34-IAEDANS and W214 are completely exposed to water. The mean wavelength of maximum fluorescence emission (λ_{max}^{em}) of C34-IAEDANS in the U form is 480 nm (Fig. 3A). In the N form, λ_{max}^{em} of C34-IAEDANS is blue-shifted to 465 nm (Fig. 3A), indicating that it is buried in the hydrophobic core of domain I. We observed that the value of λ_{max}^{em} in the B form is 467 nm, which is very similar to its value in the N form. This result strongly indicates that the core of domain I remains inaccessible to water and has N-like hydrophobicity.

The λ_{max}^{em} of W214 in the U form is 348 nm (Fig. 3B). However, in the N form and the B form, the λ_{max}^{em} of W214 is blue-shifted to 340 nm and 338 nm, respectively (Fig. 3B). These results indicate that in the B form, the solvation environment in the inter-domain region near W214 remains N-like or is slightly more hydrophobic than in the N form.

N-like structural fluctuations and solvent accessibility of the B form

We employed dynamic fluorescence quenching experiments to investigate structural fluctuations and solvent accessibility of domains I and II in the N form, the B form and the U form. In dynamic fluorescence quenching, the fluorescence of a buried fluorophore is quenched by a quencher present in the solvent due to the nanosecond fluctuations in the protein structure[32,38] and is described by the Stern–Volmer equation as:

$$\frac{F_0}{F} = 1 + K_{sv}[Q] \qquad (10)$$

In eqn (10), F_0 and F are the fluorescence intensities of the fluorophore in the absence and the presence of the quencher (Q), respectively, and K_{sv} is the Stern–Volmer constant. We used

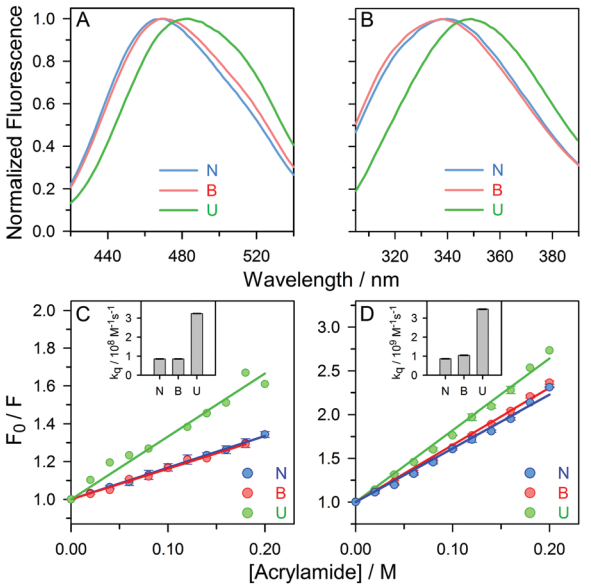


Fig. 3 The B form retains N-like hydrophobicity in domains I and II. The fluorescence spectra of (A) C34-IAEDANS and (B) W214. The wavelength of maximum emission (λ_{max}^{em}) of each spectrum was normalized to 1 for comparison. Stern–Volmer plots for dynamic quenching of (C) C34-IAEDANS and (D) W214 fluorescence. The solid lines through the data are the least-squares fits to eqn (10). The error bars shown in panels C and D represent the spread in the values of fluorescence at the emission maximum estimated from two or more independent measurements. The values of the bimolecular quenching constant, k_q, are compared in the insets of panels C and D. The errors in the values of k_q are less than 5%. (N form: HSA at pH 7; B form: HSA at pH 11; U form: HSA in 9 M urea at pH 7).

acrylamide as the neutral quencher of C34-IAEDANS and W214 fluorescence, as described previously.[32] We observed that the mean value of K_{sv} for C34-IAEDANS in the U form is 3.3 M^{-1} (Fig. 3C). However, the mean value of K_{sv} decreases to 1.6 M^{-1} in both the N form and the B form (Fig. 3C). The observed mean value of K_{sv} for W214 in the U form is 8.2 M^{-1}, which decreases to 6.1 M^{-1} and 6.5 M^{-1} in the N form and the B form, respectively (Fig. 3D). It is interesting to note that the mean values of K_{sv} remain similar in the N form and the B form for both C34-IAEDANS and W214.

The bimolecular quenching constant (k_q) is a measure of the accessibility of the fluorophore to the quencher in its photo-excited state. We measured the values of k_q using the above values of K_{sv} and the experimentally determined values of intensity averaged fluorescence lifetime (τ_0) (Fig. S3 and Table S2, ESI†) ($K_{sv} = k_q\tau_0$). We observed that the mean value of k_q for C34-IAEDANS in the U form is 3.2×10^8 M^{-1} s^{-1}, which reduces significantly to 0.8×10^8 M^{-1} s^{-1} in both the N form and the B form (Fig. 3C (inset)). For W214, the mean values of k_q in the N form (0.9×10^9 M^{-1} s^{-1}) and the B form (1.0×10^9 M^{-1} s^{-1}) are similar and significantly lower than the value of k_q in the U form (3.4×10^9 M^{-1} s^{-1}) (Fig. 3D (inset)). All the values of K_{sv}, τ_0 and k_q are listed in Table S2 (ESI†). It is important to note that the errors in the estimation of k_q values are less than 5%. Hence, the similarities in the values of k_q in the N form and the B form indicate that in the B form, domain I and domain II exhibit N-like structural fluctuations and solvent accessibilities.

B form retains N-like solvation dynamics

We investigated the solvation dynamics of the core of domains I and II around C34-IAEDANS and W214, respectively, by REES experiments.[32,39] In general, the λ_{max}^{em} of a fluorophore is independent of the excitation wavelength (λ_{ex}). However, in the case of REES, the λ_{max}^{em} of a polar fluorophore shifts to a higher wavelength (red shift) upon exciting the fluorophore towards the red edge of its excitation spectrum. This phenomenon occurs in cases where the distribution of solvent dipoles around the fluorophore is heterogeneous and solvation dynamics is slower than the fluorescence time scale.[32,39] In this way, REES probes the dynamics of the immediate surroundings of the fluorophore.

We observed that in both the N form and the B form, the λ_{max}^{em} of W214 and C34-IAEDANS shifts towards the red, when excited at the red edge of the excitation spectrum of the respective fluorophore (Fig. 4 and Fig. S4, ESI†). For C34-IAEDANS, we observed a significant amount of shift in λ_{max}^{em}, from 470 nm to 475 nm, in the N form and the B form, when excited from 337 nm to 407 nm (Fig. 4A). For W214, when excited from 295 nm to 305 nm, we observed a slight shift in λ_{max}^{em}, from 340 nm to 342 nm, in the N form (Fig. 4B). In the B form, the λ_{max}^{em} shifts from 338 nm to 340 nm (Fig. 4B). However, for both the fluorophores, the λ_{max}^{em} remains constant (483 nm for C34-IAEDANS, 348 nm for W214) in the U form (Fig. 4A and B), when excited at different λ_{ex} of the respective fluorophore. Both the fluorophores show no REES in the U form, which is expected due to the complete solvent exposure of the fluorophores and water molecules around the fluorophore are highly dynamic. In the N form and the B form, C34-IAEDANS shows large REES of 5 nm. Because C34-IAEDANS is buried in domain I, it must be measuring the solvation dynamics of the side-chain packing inside the protein core. It appears that both in the N form and the B form, the protein core is heterogeneous and the side-chains solvating C34-IAEDANS exhibit solvation dynamics on a much slower time scale than the nanosecond time scale of fluorescence. For W214, the magnitude of REES is small (2 nm) but similar in both the N form and the B form. However, it is important to note that from comparison of λ_{max}^{em}, W214 in the B form appears to be in a slightly more hydrophobic environment than the N form. Hence, the results for W214 indicate that the solvent molecules around W214 have restricted movements and they move slower than the fluorescence time scale in both the N form and the B form.

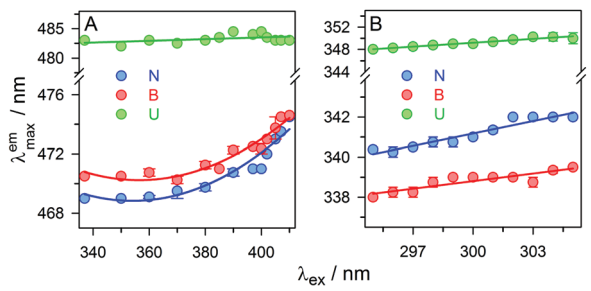


Fig. 4 Effect of the change in excitation wavelength (λ_{ex}) on the wavelength of maximum emission (λ_{max}^{em}). The λ_{max}^{em} of (A) C34-IAEDANS and (B) W214 is plotted as a function of respective λ_{ex}, where the solid lines are drawn to guide the eyes. The error bars in panels A and B represent the spread in the values of emission maximum estimated from two independent measurements. (N form: HSA at pH 7; B form: HSA at pH 11; U form: HSA in 9 M urea at pH 7).

Comparison of the global structure of the N form and the B form

We compared the global secondary and tertiary structure of the N form, the B form and the U form using far- and near-UV CD spectra (Fig. 5A and B). The observed mean residual ellipticity (MRE) at 222 nm, a measure of the global secondary structure, is $-19\,365 \pm 440$ deg cm^2 dmol^{-1} in the N form and $-18\,805 \pm 560$ deg cm^2 dmol^{-1} in the B form (Fig. 5A). The MRE values decrease markedly to -5093 ± 225 deg cm^2 dmol^{-1} in the U form (Fig. 5A). The similar values of MRE in the N form and the B form indicate that the global secondary structure in the B form remains similar to that in the N form.

The intensity of the near-UV CD signal is a measure of the global tertiary structure of proteins. We observed that the near-UV CD spectra in the N form and the B form show absorption mainly in the 255–270 nm wavelength range (Fig. 5B). The MRE at 260 nm, which is the wavelength of maximum absorbance, in the N form and the B form is -177 ± 8 deg cm^2 dmol^{-1} and -175 ± 6 deg cm^2 dmol^{-1}, respectively (Fig. 5B). The MRE at 260 nm in the U form is -92 ± 18 deg cm^2 dmol^{-1} (Fig. 5B). The similarity in MRE in the N form and the B form indicates that the global tertiary structure in the B form is similar to that

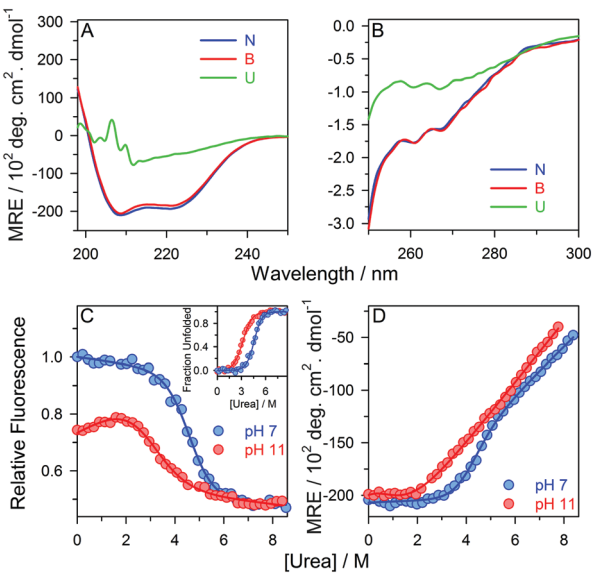


Fig. 5 The global secondary structure (A) and the tertiary structure (B) of the B form are similar to those of the N form but their thermodynamic stability is different. The equilibrium unfolding transition of HSA is monitored by the change in fluorescence of W214 (C) and the far-UV CD signal at 222 nm (D). The fraction of unfolded protein is plotted against [urea] in the inset of panel C. The solid lines through the data in panel C and pH 7 data in panel D are fits to a two-state model (eqn (7)), while that for pH 11 in panel D is drawn to guide the eyes. (N form: HSA at pH 7; B form: HSA at pH 11; U form: HSA in 9 M urea at pH 7).

in the N form. This is a surprising result in view of the fluorescence, size exclusion chromatography and the FRET experiments (Fig. 1 and 2), which indicate that structurally the B form and the N form are different; specifically domains I and II move away from each other in the B form which leads to the disruption of the tertiary structure in the inter-domain region near W214 and protein expansion. It is important to note that near-UV CD experiments in the case of HSA report primarily on the asymmetric burial of its 31 phenylalanine (Phe) residues. A careful examination of the crystal structure of HSA reveals that 30 out of its 31 Phe residues are distributed in the intra-domain region and only one Phe residue, F206, is located near the cluster of side-chains involved in the inter-domain side-chain packing (Fig. S5A, ESI†). It is also interesting to note that the intra-domain structure of HSA (in all the 3 domains) is stapled by a network of 17 disulfide bonds (Fig. S5B, ESI†), but not the inter-domain region. Hence, it appears that in the B form the intra-domain packing is intact resulting in a similar near-UV CD spectrum to that in the N form, but packing at the interface of domains I and II is broken, giving rise to a change in fluorescence, and an increase in the inter-domain distance and protein volume. Previous studies, using nuclear magnetic resonance, ligand binding and monoclonal antibody binding properties, also support our conclusion that the structural fluctuation in the B form is particularly limited to domains I and II.[24,40,41]

Comparison of thermodynamic stabilities of the N form and the B form

The base-induced B form resembles a DMG in domains I and II near the inter-domain region. We estimated the energetic contribution of vdW packing interactions near the inter-domain region of domains I and II in the stability of the N form by measuring the thermodynamic stabilities of the N form and the B form by the urea-induced equilibrium unfolding transitions using fluorescence (Fig. 5C) and far-UV CD (Fig. 5D). The equilibrium unfolding curves at pH 7 represent N $\rightleftharpoons$ U transition, while at pH 11 they represent B $\rightleftharpoons$ U transition. Here, it is important to note that the U form (unfolded in 9 M urea) at pH 7 and pH 11 is identical in terms of all the spectroscopic properties measured in this study, including Stokes shift, hydration of side-chains, solvation dynamics and secondary and tertiary structural content (Fig. S6, ESI†). All the urea-induced unfolding curves show sigmoidal transitions except far-UV CD monitored unfolding transition, which appears to be gradual in nature (Fig. 5D). This result indicates that the unfolding of the secondary structure in the DMG could be non-cooperative and gradual in nature.[42] We analysed the fluorescence monitored data at pH 7 and pH 11 and CD monitored data at pH 7 using a two-state model (eqn (7)) to obtain the values of ΔG (free energy of unfolding) and m (slope of the transition which represents a change in the solvent accessible surface area). For N $\rightleftharpoons$ U transition, the values of ΔG_{N-U} and m_{N-U} are 5.35 $\pm$ 0.17 kcal mol^{-1} and 1.17 $\pm$ 0.03 kcal mol^{-1} M^{-1}, respectively. For B $\rightleftharpoons$ U transition, the values of ΔG_{B-U} and m_{B-U} are 3.0 $\pm$ 0.06 kcal mol^{-1} and 1.0 $\pm$ 0.04 kcal mol^{-1} M^{-1}. The value of m_{N-B} ($m_{N-U} - m_{B-U}$) is very small (0.13 $\pm$ 0.05 kcal mol^{-1} M^{-1}).

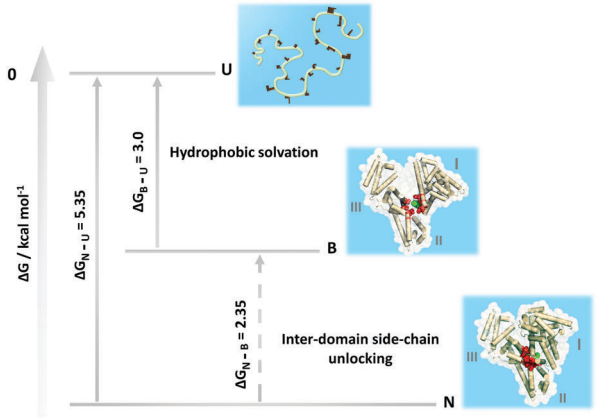


Fig. 6 Model energy diagram showing the relative energies of the N form, the B form and the U form. In the structural models of the N form and the B form, the cylinders represent α-helical secondary structures, red spheres represent the side-chains of the residues involved in the packing of the inter-domain interface, the green sphere represents the side-chain of W214 and the white surface around the protein shows the solvent accessible surface area of the protein. The model of the U form shows only the peptide backbone (white thread) and the representative side-chains (black color) of the unfolded protein. The blue color represents water molecules around the protein. In the B form, the inter-domain packing is disrupted without hydrophobic solvation. The free energies of unfolding of the N form (ΔG_{N-U}) and the hydrophobic solvation (ΔG_{B-U}) were determined from equilibrium unfolding transitions of the N form (N $\rightleftharpoons$ U transition) and the B form (B $\rightleftharpoons$ U transition), respectively, as described in Fig. 5C and the Results and discussion section. The free energy of inter-domain side-chain unlocking (ΔG_{N-B}) was calculated from $\Delta G_{N-B} = \Delta G_{N-U} - \Delta G_{B-U}$.

This result suggests that there is very little or no change in the solvent accessible surface area during N $\rightleftharpoons$ B transition. This result is expected if the B form is like a DMG and its core is dry. The value of ΔG_{N-B} ($\Delta G_{N-U} - \Delta G_{B-U}$) is 2.35 $\pm$ 0.18 kcal mol^{-1}. A comparison of ΔG_{N-U} and ΔG_{N-B} indicates that 40–45% of the total change in free energy occurs during N $\rightleftharpoons$ B transition. Hence, there is a large difference in the thermodynamic stability of the N form and the B form. One possibility for this large difference in stability is that the movement of domains I and II also disrupts some of the packing interactions between domains I and III as shown in our structural model (Fig. 6). However, the decrease in stability is not due to the disruption of the intra-domain secondary structure and side-chain packing as shown by CD experiments. Hence, our results indicate that vdW interactions play a very important role in the stability of the N form, but not the DMG. It is important to note that for a few proteins,[18,43,44] it has been shown that even a single cavity creating point mutation can drastically decrease the stability of the N form.

It has been postulated that vdW packing interactions develop during the last step, *i.e.*, the DMG $\rightleftharpoons$ N transition, of protein folding.[16–19,45] This hypothesis has been supported by experimental observations of DMG-like intermediate states on the native-side of the free energy barrier of a few proteins.[46–53] However, recent findings on a few other proteins, including alpha-lactalbumin,[54] a nuclear coactivator binding domain[55] and RNase H,[56] appear to suggest that N-like packing interactions could also develop early during the folding of an unfolded

polypeptide. Here, our observations that unfolding of the N form of HSA begins with the disruption of some of the inter-domain packing interactions and that intra-domain packing dissolves at a later stage indicate that vdW interactions develop in multiple stages during the folding of proteins. Our results of the REES experiments (Fig. 4) also suggest that proteins in the the N state themselves might not be as tightly packed as previously thought[18,57] but might be an ensemble of loosely packed forms.

Conclusions

In summary, we have shown that the B form of HSA resembles an expanded DMG-like near native state. Fluorescence experiments show that the fluorescence of W214 in the B form has markedly decreased compared to the N form, suggesting the disruption of the tertiary structure near W214, at the interface of domains I and II. This is further confirmed by FRET and size exclusion chromatography measurements, which show that the B form is more expanded than the N form. The expansion in the B form is a result of the loosening of side-chain packing interactions in domains I and II near the inter-domain region. Dynamic fluorescence quenching assays on C34-IAEDANS and W214 suggest that the expansion in the B form is not accompanied by the hydration of the protein core. REES experiments on C34-IAEDANS and W214 show that in both the N form and the B form, the solvation dynamics near respective fluorophores in domains I and II is slower than the fluorescence time scale. However, the results of the far- and near-UV CD experiments indicate that the secondary and the tertiary structures in the B form are similar to the N form. These results suggest that in the B form, domains I and II of the protein have DMG-like characteristics near the inter-domain region. The thermodynamic stability data indicate that the B form has lost 40–45% of the total stability of the N form. Because the N form and the B form appear to be different mainly in the extent of side-chain packing at the inter-domain interface, these results suggest that the vdW packing interactions play an important role in protein stability. Our results strongly indicate that the base-induced unfolding of proteins begins with the disruption of vdW interactions and that structural loosening and core solvation are two distinct steps. Since the B form has physiological importance in binding to many different ligands and drugs,[24] it is possible that the loosening of side-chain packing and the increase in conformational entropy enable the protein to undergo native state volume fluctuations in order to perform these multiple functions.[17,58]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the members of our laboratory for discussion. This work was funded by DST-SERB early career research award (project # ECR/2015/000027) to S. K. J. and CSIR-National Chemical Laboratory (Network grant # CSC 0134 and MLP grant # 030026). N. A. is a recipient of a Shyama Prasad Mukherjee Fellowship awarded by the Council of Scientific and Industrial Research, India. P. M. is a recipient of a Senior Research Fellowship by University Grants Commission, India.

References

1 M. H. Ross, J. O. Ely and J. G. Archer, *J. Biol. Chem.*, 1951, **192**, 561–568.
2 J. R. Whitaker, *Chemical Deterioration of Proteins*, American Chemical Society, 1980, ch. 7, vol. 123, pp. 145–163.
3 S. Wanwimolruk and D. J. Birkett, *Biochim. Biophys. Acta*, 1982, **709**, 247–255.
4 D. E. Schwass and J. W. Finley, *J. Agric. Food Chem.*, 1984, **32**, 1377–1382.
5 Y. Goto and A. L. Fink, *Biochemistry*, 1989, **28**, 945–952.
6 T. Konno, Y. O. Kamatari, N. Tanaka, H. Kamikubo, C. M. Dobson and K. Nagayama, *Biochemistry*, 2000, **39**, 4182–4190.
7 J. Jiang, J. Chen and Y. L. Xiong, *J. Agric. Food Chem.*, 2009, **57**, 7576–7583.
8 K. Talley and E. Alexov, *Proteins*, 2010, **78**, 2699–2706.
9 Y. Y. Shan, M. H. Ma, X. Huang, Y. J. Guo, G. F. Jin and Y. G. Jin, *J. Food Sci.*, 2012, **77**, C740–C745.
10 W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1–63.
11 C. Tanford, *Adv. Protein Chem.*, 1968, **23**, 121–282.
12 C. Tanford, *Adv. Protein Chem.*, 1970, **24**, 1–95.
13 P. L. Privalov, *Adv. Protein Chem.*, 1979, **33**, 167–241.
14 E. I. Shakhnovich and A. V. Finkelstein, *Biopolymers*, 1989, **28**, 1667–1680.
15 A. V. Finkelstein and E. I. Shakhnovich, *Biopolymers*, 1989, **28**, 1681–1694.
16 R. L. Baldwin, C. Frieden and G. D. Rose, *Proteins*, 2010, **78**, 2725–2737.
17 R. L. Baldwin and G. D. Rose, *Curr. Opin. Struct. Biol.*, 2013, **23**, 4–10.
18 S. Bhattacharyya and R. Varadarajan, *Curr. Opin. Struct. Biol.*, 2013, **23**, 11–21.
19 D. Thirumalai, Z. Liu, E. P. O'Brien and G. Reddy, *Curr. Opin. Struct. Biol.*, 2013, **23**, 22–29.
20 B. R. Rami and J. B. Udgaonkar, *Biochemistry*, 2002, **41**, 1710–1716.
21 D. C. Carter and J. X. Ho, *Adv. Protein Chem.*, 1994, **45**, 153–203.
22 J. Figge, T. H. Rossing and V. Fencl, *J. Lab. Clin. Med.*, 1991, **117**, 453–467.
23 W. J. Leonard, K. Kantvija and J. F. Foster, *J. Biol. Chem.*, 1963, **238**, 1984–1988.
24 T. Peters Jr, *All About Albumin: Biochemistry, Genetics, and Medical Applications*, Academic Press, San Diego, 1st edn, 1995.
25 J. H. Droge, J. Wilting and L. H. Janssen, *Biochem. Pharmacol.*, 1982, **31**, 3781–3786.
26 M. Dockal, D. C. Carter and F. Ruker, *J. Biol. Chem.*, 2000, **275**, 3042–3050.

27 L. Painter, M. M. Harding and P. J. Beeby, *J. Chem. Soc., Perkin Trans. 1*, 1998, 3041–3044.
28 C. N. Pace, *Methods Enzymol.*, 1986, **131**, 266–280.
29 N. Acharya, P. Mishra and S. K. Jha, *J. Phys. Chem. Lett.*, 2016, **7**, 173–179.
30 C. Atanasiu, T. J. Su, S. S. Sturrock and D. T. Dryden, *Nucleic Acids Res.*, 2002, **30**, 3936–3944.
31 R. Khurana, A. T. Hate, U. Nath and J. B. Udgaonkar, *Protein Sci.*, 1995, **4**, 1133–1144.
32 J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer, Singapore, 2006.
33 G. B. Irvine, *Current Protocols in Cell Biology*, John Wiley & Sons, Inc., 2001, ch. 5, pp. 5.5.1–5.5.16.
34 M. Amiri, K. Jankeje and J. R. Albani, *J. Fluoresc.*, 2010, **20**, 651–656.
35 M. Amiri, K. Jankeje and J. R. Albani, *J. Pharm. Biomed. Anal.*, 2010, **51**, 1097–1102.
36 V. R. Agashe and J. B. Udgaonkar, *Biochemistry*, 1995, **34**, 3286–3299.
37 R. F. Steiner and H. Edelhoch, *Nature*, 1961, **192**, 873–874.
38 M. R. Eftink and C. A. Ghiron, *Proc. Natl. Acad. Sci. U. S. A.*, 1975, **72**, 3290–3294.
39 A. P. Demchenko, *Methods Enzymol.*, 2008, **450**, 59–78.
40 C. Lapresle, *Anal. Biochem.*, 1988, **174**, 308–312.
41 O. J. M. Bos, J. F. A. Labro, M. J. E. Fischer, J. Wilting and L. H. M. Janssen, *J. Biol. Chem.*, 1989, **264**, 953–959.
42 S. K. Jha, D. Dhar, G. Krishnamoorthy and J. B. Udgaonkar, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 11113–11118.
43 W. S. Sandberg and T. C. Terwilliger, *Science*, 1989, **245**, 54–57.
44 A. E. Eriksson, W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin and B. W. Matthews, *Science*, 1992, **255**, 178–183.
45 S. K. Jha and J. B. Udgaonkar, *J. Biol. Chem.*, 2007, **282**, 37479–37491.
46 T. Kiefhaber, A. M. Labhardt and R. L. Baldwin, *Nature*, 1995, **375**, 513–515.
47 S. D. Hoeltzli and C. Frieden, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 9318–9322.
48 S. K. Jha and J. B. Udgaonkar, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 12289–12294.
49 A. Reiner, P. Henklein and T. Kiefhaber, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 4955–4960.
50 S. K. Jha and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 4856–4861.
51 A. Dasgupta, J. B. Udgaonkar and P. Das, *J. Phys. Chem. B*, 2014, **118**, 6380–6392.
52 S. Neumaier and T. Kiefhaber, *J. Mol. Biol.*, 2014, **426**, 2520–2528.
53 P. Mishra and S. K. Jha, *J. Phys. Chem. B*, 2017, **121**, 9336–9347.
54 S. Ramboarina and C. Redfield, *J. Am. Chem. Soc.*, 2008, **130**, 15318–15326.
55 M. Kjaergaard, F. M. Poulsen and K. Teilum, *Biophys. J.*, 2012, **102**, 1627–1635.
56 L. E. Rosen, K. B. Connell and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14746–14751.
57 F. M. Richards, *Annu. Rev. Biophys. Bioeng.*, 1977, **6**, 151–176.
58 A. B. Law, P. J. Sapienza, J. Zhang, X. Zuo and C. M. Petit, *J. Am. Chem. Soc.*, 2017, **139**, 3599–3602.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 1256−1264

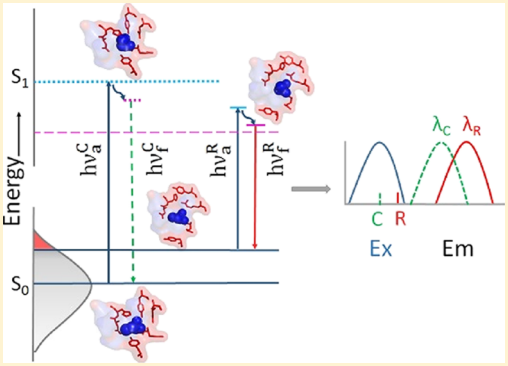
pubs.acs.org/JPCB

Slow Motion Protein Dance Visualized Using Red-Edge Excitation Shift of a Buried Fluorophore

Prajna Mishra and Santosh Kumar Jha*

Physical and Materials Chemistry Division, Academy of Scientific and Innovative Research (AcSIR), CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India

ABSTRACT: It has been extremely challenging to detect protein structures with a dynamic core, such as dry molten globules, that remain in equilibrium with the tightly packed native (N) state and that are important for a myriad of entropy-driven protein functions. Here, we detect the higher entropy conformations of a human serum protein, using red-edge excitation shift experiments. We covalently introduced a fluorophore inside the protein core and observed that in a subset of native population, the side chains of the polar and buried residues have different spatial arrangements than the mean population and that they solvate the fluorophore on a timescale much slower than the nanosecond timescale of fluorescence. Our results provide direct evidence for the dense fluidity of protein core and show that alternate side-chain packing arrangements exist in the core that might be important for multiple binding functions of this protein.



INTRODUCTION

The extent and degree of conformational heterogeneity in side-chain packing inside the hydrophobic interior of proteins are poorly understood. A body of theoretical work predicts a spectrum of possibilities for the dynamic properties of the protein core, including: (i) proteins could be like aperiodic solids which lack the periodic regularity of crystals;[1] (ii) proteins may behave like surface-molten solids with liquid-like surface, whereas interior could be solid-like;[2] and (iii) the side chains in the core of proteins could undergo substantial structural fluctuations and the core could be like a dense fluid.[3,4] In contrast, crystal structures of proteins project a static image of the native (N) state, where the interior of proteins is tightly packed with packing densities as high as 0.7−0.8,[5−7] which is very similar to the solid crystals of small organic compounds. Recent studies propose that the N state ensemble contains loosely packed forms termed dry molten globules (DMG)[8−11] in which tight side-chain packing is ruptured but the core remains dehydrated. DMG-like states have expanded and dynamic cores with high conformational entropy and are important for entropy-driven protein function such as dynamic allostery.[12,13] Experimentally, DMG-like states, however, have been majorly observed as kinetic intermediates,[14−16] and it has remained elusive to detect and characterize them under equilibrium conditions in the N state ensemble.[8,11,17] Thermodynamically, it is possible for proteins to undergo large fluctuations around the mean values of their physical properties such as internal energy, enthalpy, and molecular volume to give rise to DMG-like protein structures because of their small size.[18] However, it has remained extremely challenging to understand the precise molecular and temporal nature of fluctuations and how small and rapid movements of protein side chains in the hydrophobic core may combine to give an overall change in the mean configuration.

There is a scarcity of experimental and spectroscopic methods which can yield dynamic information (on microsecond to nanosecond timescale) on side-chain packing in the native state ensemble of proteins. NMR methods such as dynamic order parameter of methyl side chains and relaxation dispersion experiments have been employed for a few proteins[19] including calmodulin[20,21] and cyclophilin A,[22,23] and side-chain dynamics in picosecond to microsecond timescale have been measured. However, NMR methods are limited to small, soluble proteins and cannot be applied to large, multidomain, and aggregation-prone proteins because of the requirement of millimolar concentrations for NMR studies. Thiol−disulfide exchange is an elegant alternative method that has provided temporal information on the side-chain packing of proteins such as apmyoglobin,[24] barstar,[25] monellin,[26] and RNAse H.[27] A major disadvantage of this method, however, is that undesirable side reactions such as disulfide scrambling and air oxidation of thiols to disulfide compete with the thiol−disulfide exchange reaction.

Fluorescence-based methods are highly sensitive (require micromolar concentrations of proteins) and are methods of choice for large multidomain proteins. Red-edge excitation shift (REES) studies of a fluorophore buried in the core of a protein can potentially give dynamic information on dipoles created by polar and polarizable side chains of the hydrophobic core, if they solvate the fluorophore on nanosecond and slower timescale.[28,29] The phenomenon of REES is marked by a red

Received: November 17, 2018
Revised: January 7, 2019
Published: January 14, 2019

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

shift in the wavelength of the maximum fluorescence emission (λ_{max}^{em}), when excited at the red edge of the excitation spectrum of the fluorophore[28,30,31] (Figure 1). In general, the λ_{max}^{em} for a

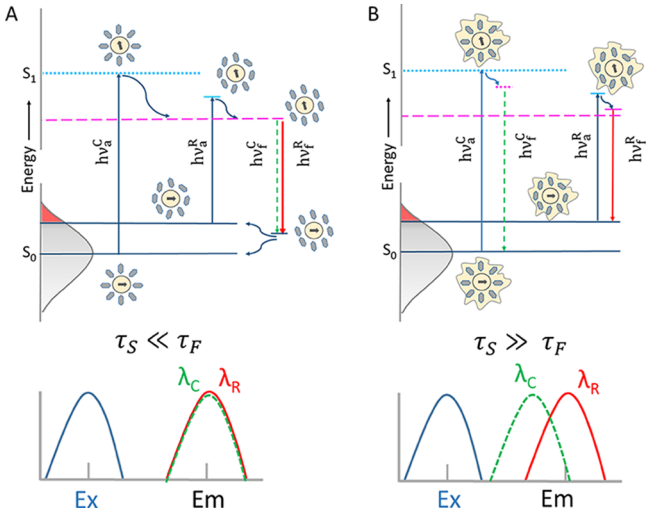


Figure 1. Schematic description of REES. Comparative effects of the central and the red-edge excitations (top) on the emission spectra (bottom) of a fluorophore are shown. (A) Timescale of the solvent relaxation, $\tau_s \ll$ the fluorescence timescale, τ_f, (B) $\tau_s \gg \tau_f$. In both the panels, the inner yellow circle and the thick gray arrow inside it represent the fluorophore and its dipole moment vector, respectively. The gray hexagons and their directions denote the orientations of the solvent molecules. The shaded gray distribution with the red tail represents the distribution of the solute−solvent interaction energy in the ground singlet state (S_0). In the excited singlet state (S_1), the dotted and solid cyan lines show the Franck−Condon (F) states for the center and the red-edge excitations, respectively. The dashed pink lines indicate the complete solvent-relaxed (R) states. The incomplete solvent-relaxed states (R′) for the center and the red-edge excitations are presented by the dotted and the solid pink lines, respectively. The vertical upward blue arrows represent the excitation energies (absorbance), and the vertically downward dashed green and solid red arrows represent the fluorescence emission energies. In the bottom panels, the green dashed lines and the red solid lines, respectively, represent the emission spectra when excited at the center (λ_C) and the red edge (λ_R) of the absorbance spectrum (the blue solid lines). In the panel (B) (top), the crooked line around the gray hexagons denotes a restricted solvation environment, as inside the core of a protein.

fluorophore is independent of the wavelength of excitation (λ^{ex}).[31] However, in the cases where the environment surrounding the fluorophore is heterogeneous and motionally restricted, this principle is not followed and REES is observed. The molecular basis of REES is governed by the heterogeneity of the fluorophore−solvent interactions and the rate at which the solvent dipoles reorient themselves around the fluorophore in the excited state, as schematically illustrated in Figure 1. For the illustration, a model[30,32] is considered in which there exists a wide distribution of the fluorophore−solvent interaction energy in the ground state and after a vertical transition to the Franck−Condon (F) excited state, the fluorescence emission with a decay rate τ_f can occur from two distinct solvation states: (i) a completely solvent-relaxed R state, where reorientation of solvent dipoles around the fluorophores with a time constant of τ_s is complete, and (ii) a partially solvent-relaxed R′ state, where solvent-dipole reorientation around the fluorophore has not completely occurred. Because of the heterogeneous distribution of the fluorophore−solvent interaction energy in the ground state, photoexcitation of the ensemble at the center (λ_C) or the red edge (λ_R) of the absorbance spectrum photoselects a different subpopulation of the molecules to the excited state. In the case of a dynamically solvated fluorophore (Figure 1A), the rate of the solvent relaxation is faster than that of the fluorescence timescale ($\tau_s < \tau_f$) which allows the solvent dipoles to realign around the excited fluorophore before the emission. As a result, irrespective of the excitation at the λ_C or the λ_R, the fluorescence emission takes place from the completely solvent-relaxed R state. Thus, the emission spectra become independent of the λ^{ex}. On the other hand, if the fluorophore is placed in a heterogeneous and restricted solvation environment, as in the core of a protein, the solvent relaxation time becomes slower than the rate of the fluorescence emission, that is, $\tau_s > \tau_f$ (Figure 1B). This does not permit the solvent dipolar realignment before the fluorescence emission, and in this case, the emission occurs from a partially solvent-relaxed R′ state. For the low-energy λ_R excitation, the orientations of the solvent dipoles are such that the energy difference between the ground and the R′ state is smaller than that for the high-energy λ_C excitation. As a result, the fluorescence spectra exhibit red spectral shift for the λ_R excitation compared to the λ_C excitation.

One of the unique features of REES is that it provides information about the heterogeneity and the dynamics of the environment, in contrast to other fluorescence techniques that report about the fluorophore itself. This unique feature of REES makes it a reliable method to study the dense fluidity of the side-chain residues inside the protein core. Proteins with heterogeneous side-chain packing can give rise to an ensemble of DMG-like conformations in equilibrium with the N state.[9] If the polar and buried side-chain residues inside the core of the DMG solvate a suitably placed fluorophore on a timescale much slower than the fluorescence timescale, REES can provide useful information on the equilibrium conformational states of a protein. REES has been applied mainly to study the solvation dynamics of water molecules in complex solvation environment.[28,29,33] The dynamics of interfacial water molecules in model membranes is one such example.[32,34,35] In the context of proteins, REES has been utilized to understand the conformational changes in the native,[30,36,37] the unfolded (U),[38] and the molten globule[11,39,40] states. Extrinsic fluorescent probes such as 2-(*p*-toluidinylnaphthalene)-6-sulfonate have been employed to monitor REES in multiple proteins including *β*-lactoglobulin, *β*-casein, and serum albumins.[30] In the case of erythroid spectrin,[38] application of REES revealed the presence of residual structure in the unfolded state. However, its application to study the dynamics of the protein core is highly limited.[28,29]

In this study, we have utilized REES of a buried fluorophore to study the dynamics and conformational heterogeneity of core packing in the native state ensemble of a multidomain protein, human serum albumin (HSA). We observed that the N state of HSA is heterogeneous, and in a subpopulation of protein molecules, some of the side-chain residues in the protein core have different orientations and packings than the mean population. Our results indicate that the N state of HSA is an ensemble of DMG-like states.

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

MATERIALS AND METHODS

Spectroscopic Methods and Instruments. We used a PerkinElmer LAMBDA 650UV/Vis spectrophotometer and a PerkinElmer fluorescence spectrometer LS 55, respectively, to measure all of the absorption and fluorescence emission spectra. In all of the cases, cuvette of path length 1 cm was used. All of the circular dichroism (CD) spectra were obtained using a JASCO J-815 CD spectrometer. Here, cuvettes of path length 1 mm and 1 cm were used for far- and near-UV CD, respectively. For all of the spectra, background signal corrections were done by subtracting the buffer signals.

Chemicals, Reagents, Buffers, and Experimental Conditions. HSA (99% pure, fatty acid and globulin free) and urea (ultrapure grade) were procured from Alfa Aesar. 5-((((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was purchased from Life Technologies. Guanidine hydrochloride (GdmCl) was obtained from Sigma. All other chemicals of highest purity grade were procured from HiMedia and used as received without any further purification. We determined the concentration of HSA by measuring the absorbance at 280 nm and using a molar extinction coefficient[41] of 36 500 M^{-1} cm^{-1}.

Acetate (20 mM), phosphate (20 mM), and Tris-HCl (20 mM) were used at pHs 5, 7, and 8 native buffers, respectively. The unfolding buffers for each pH consisted of 9 M urea in their respective native buffers. We determined the concentrations of urea and GdmCl. from refractive index measurements,[42] using an Abbe refractometer from Rajdhani Scientific Instruments Co. (model: RSR-2).

Preparation of 1,5-IAEDANS-Labeled HSA. HSA was labeled with 1,5-IAEDANS at the sole-free cysteine, C34, as described earlier.[11] In short, the protein was labeled with 15-fold molar excess of dye in 6 M GdmCl and 20 mM Tris-HCl at pH 8. The reaction mixture was kept in dark for 4 h. The protein was then refolded, and the labeling reaction was quenched by diluting the reaction mixture into 20-fold volume excess of refolding buffer (20 mM phosphate buffer, pH 7) followed by overnight incubation at 4 °C. The labeled protein was then separated from the excess free dye by a PD-10 column. The percentage of labeling was estimated as reported previously.[11] The labeled protein (HSA-IAEDANS) was found to be >95% labeled with the dye.

pH Dependence of Protein Structure Monitored by Fluorescence and CD. For pH dependence studies of the protein structure, all of the samples were incubated overnight at different pHs. For fluorescence measurements, the protein concentration was ∼4−5 μM. The aromatic residues were excited at 280 nm, and the emission was collected from 310 to 420 nm. C34-IAEDANS was excited at 337 nm, and emission spectrum was collected from 350 to 550 nm. Slit widths for excitation and emission were kept nominal. The scan speed was 100 nm/min.

For far-UV and near-UV CD measurements, the protein concentrations were 4−5 and ∼20 μM, respectively. The far-UV CD experiment was monitored in the wavelength range of 190−250 nm, whereas for near-UV CD experiment, it was 250−310 nm. Each spectrum was averaged over three scans. For each CD spectrum, the data pitch, data integration time, bandwidth, and scan speed were 1 nm, 1 s, 2 nm, and 100 nm/min, respectively.

Urea-Induced Equilibrium Unfolding Transitions Monitored by Fluorescence. For the urea-induced equilibrium unfolding transitions for HSA and HSA-IAEDANS, the protein samples (∼4−5 μM) were incubated overnight in different concentrations of urea at room temperature at pH 7.0. The equilibrium unfolding was monitored by the change in fluorescence of W214 at 340 nm upon excitation at 295 nm. The transitions were analyzed using a two-state, N ⇌ U model. The data were fitted to eq 1[11,43]

$$y_{obs} = \frac{y_N + y_U\, e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (1)$$

where y_{obs} is the observed fluorescence signal, y_N and y_U are the signals of the native and unfolded states, respectively, and ΔG_{NU} is the free energy of unfolding of N to U. It is assumed that ΔG_{NU} is linearly dependent on the denaturant concentration [D] and is given by

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \quad (2)$$

where $\Delta G_{NU}^{H_2O}$ and m_{NU} are the standard free energy at 0 M urea and slope of the N ⇌ U transition, respectively.

Time-Resolved Fluorescence Measurements. An excitation laser source of 370 nm was used to measure the time-resolved decay kinetics of C34-IAEDANS. The emission decays were collected at 480 nm with 10 000 counts at the peak. An aqueous solution of the milk powder was used to measure the instrument response function (IRF), which was found to be ∼196 ps. We used the DAS6 analysis software supplied by HORIBA Scientific and deconvoluted and fitted the fluorescence decays to a sum of two-three exponentials as per the following equation[31]

$$I(t) = \sum_i \alpha_i\, e^{-\tau/\tau_i}, \qquad i = 1 - 3 \quad (3)$$

In the above equation, $I(t)$ corresponds to the fluorescence intensity at time t and α_i is the amplitude of the ith fluorescence lifetime, τ_i. The mean fluorescence lifetime (τ_m) was calculated using the following equation

$$\tau_m = \sum \alpha_i \tau_i \quad (4)$$

such that $\sum_i \alpha_i = 1$ and $i = 1−3$.

REES Experiments. For C34-IAEDANS, the fluorescence emission spectra were collected by exciting from 337 to 410 nm, whereas for W214, the excitation wavelengths ranged from 295 to 305 nm. The protein concentration was kept between ∼8 and 15 μM. Nominal slit widths for both excitation and emission were used. The scan speed was 50 nm/min. Each spectrum was averaged over three scans. For all of the samples, buffer signal was corrected for solvent Raman peak. The emission maxima were calculated both by the first derivative as well as the manual inspection of the fluorescence spectrum and were found to be similar. The error bars for the values of emission maximum were estimated from the standard deviations of at least three independent measurements.

RESULTS AND DISCUSSION

C34 Is Buried inside the Protein Core and Is Solvated by Polar and Polarizable Residues. HSA is the most abundant protein found in blood plasma having the major function of transportation of fatty acids, hormones, and other compounds through the blood stream. It is an all-helical, three-domain protein where each domain is divided into two

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

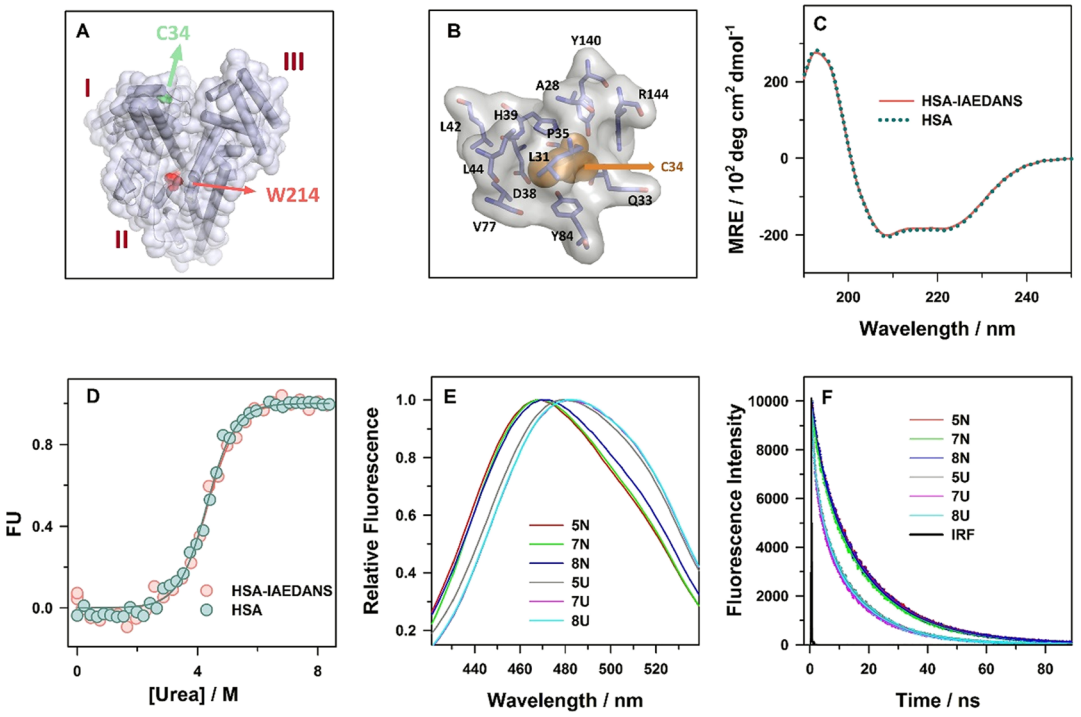


Figure 2. C34 and C34-IAEDANS are buried in the protein structure in the N state, and HSA and HSA-IAEDANS have similar structure and stability. (A) Location of C34 and W214 is shown in the crystal structure of HSA (drawn from PDB file 1AO6 using the program PyMOL). In the three-domain structure of the protein, C34 is completely buried in the core of domain I and W214 is partially buried in the interdomain region. (B) Structure of HSA showing the side-chain residues solvating C34 in the core of domain I. The atoms of C34 are shown in orange spheres. The side chain of the amino acids solvating C34 are shown in blue sticks. The background light gray surface around the amino acid residues shows the water accessible surface area of the residues burying C34. (C) Mean residue ellipticity (MRE) of HSA and HSA-IAEDANS is shown in the far-UV range in the blue dotted and red solid lines, respectively. (D) Change in the fraction of unfolded molecules (FU) as a function of [urea] is shown for HSA (blue circles) and HSA-IAEDANS (red circles). The blue solid line through the circles is a fit to a two-state, N ⇌ U model. (E) Fluorescence emission spectra of HSA-IAEDANS in the N and the U states are shown at different pHs (indicated in the figure). Each spectrum was normalized at its maximum intensity. (F) Fluorescence lifetime decay of C34-IAEDANS in the N and the U states at different pHs as indicated in the figure is shown (IRF: Instrument Response Function).

subdomains (Figure 2A). It has 35 cysteine residues distributed over its three domains. Out of these 35 cysteines, 34 forms disulfide bonds and the sole-free cysteine (C34) is buried in the hydrophobic core of domain I (Figure 2A). From the crystal structure, it was revealed that inside the core of domain I, C34 is solvated by the side chains of residues A28, L31, Q33, P35, D38, H39, L42, L74, V77, Y84, Y140, and R144 (Figure 2B). Out of the 12 amino acid residues solvating C34 in the protein core, two are basic (H39 and R144), one is acidic (D38), and three are polar but uncharged (Q33, Y84, and Y140) in nature. The percentage burial of all of these residues is 95 ± 5; except Q33, P35, and D38, which are around 60% buried. We also calculated the water accessibility of C34 by ASAview[44] and found it to be ∼99% buried.

Structure and Stability of HSA Does Not Change upon Labeling C34 with 1,5-IAEDANS. In order to probe the core dynamics of the protein, we labeled C34 with a highly sensitive fluorophore, 1,5-IAEDANS. It is important to determine that labeling does not change the structural properties as well as the thermodynamic stability of HSA. We first compared the global secondary structure of the labeled (HSA-IAEDANS) and the unlabeled proteins by far-UV CD spectroscopy (Figure 2C). The mean residual ellipticity (MRE) for HSA and HSA-IAEDANS at pH 7 at 222 nm was −18 670 and −18 262 deg cm^2 dmol^{-1}, respectively, which conveyed that both the labeled and unlabeled HSA have similar secondary structure. We next investigated the effect of labeling on the thermodynamic stability of HSA by urea-induced unfolding experiments (Figure 2D). We observed that the standard free energy of unfolding, $\Delta G_{NU}^{H_2O}$, for HSA was 5.17 ± 0.15 kcal mol^{-1} and for HSA-IAEDANS was 5.11 ± 0.18 kcal mol^{-1}. The similar values of $\Delta G_{NU}^{H_2O}$ for both HSA and HSA-IAEDANS indicated that the structural stability of the N state did not change upon labeling. Similarly, the *m*-value, which is a measure of the solvent-exposed surface area of the protein, remained similar for both HSA (1.15 ± 0.03 kcal mol^{-1} M^{-1}) and HSA-IAEDANS (1.15 ± 0.10 kcal mol^{-1} M^{-1}).

Burial of C34-IAEDANS in the Core as Revealed by Steady-State and Time-Resolved Fluorescence Studies. All of the above data revealed that labeling of C34 did not change the global secondary structure and thermodynamic stability of HSA. Hence, C34-IAEDANS can be used as a reporter to probe the core dynamics of the protein in the N state. HSA is known to maintain N-like conformation in the pH range of 4.8−8.5[11,40] which forms its N state ensemble. In one of our previous studies,[11] we have shown that the N state ensemble of HSA contains a DMG-like intermediate with an alternate side-chain packing arrangement and that the pH modulates the N ⇌ DMG equilibrium. The population of the DMG increases as the pH of the environment decreases from 7 to 5. In order to further understand the equilibrium dynamics of the N state ensemble, we performed REES experiments at

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

different pHs varying from pH 5 to 8. It is, however, important to ensure that the probe C34-IAEDANS is buried in the protein core at all of the pH. Fluorescence Stokes' shift and fluorescence lifetimes are sensitive measures of the solvation environment of a fluorophore. We, therefore, monitored the λ^{em}_{max} and fluorescence lifetime as a function of pH in the N- and urea-unfolded forms at pHs 5, 7, and 8 to probe the burial of C34-IAEDANS. The λ^{em}_{max} of C34-IAEDANS in the solvent-exposed U state at all of the pH is 483 nm and that in the N state is 469 nm (Figure 2E). Because around 50% of the residues solvating C34 is polar or charged as revealed by the crystal structure (Figure 2B), the blue shift is not expected to be very large as in the case of a complete hydrophobic environment. Nevertheless, the observation that C34-IAEDANS shows a large Stokes' shift in the U state which significantly blue shifts in the folded state at all of the pH, indicating that C34-IAEDANS is buried in the hydrophobic core in the N state. Similarly, at all of the pH, the N states exhibited a larger fluorescence lifetime (~19 ns) compared to the U states (~15 ns) in the time-resolved fluorescence measurements (Figure 2F) further supporting the burial of C34-IAEDANS in the pH range 5−8. In previous studies,[40,45] we have measured the solvent accessibility of C34-IAEDANS in the pH range 2−11, using dynamic fluorescence quenching experiments with acrylamide as the quencher. The quenching experiments suggested that C34-IAEDANS is similarly buried in the protein structure at all of the pH measured. The results of the Stokes' shift and the fluorescence lifetime experiments in the current study support our previous observations.

REES as a Reporter of Heterogeneity and Dynamics of the Protein Core. In order to understand the dynamics of core residues in the N states, we excited the HSA-IAEDANS at different wavelengths from the center to the red edge of the absorbance spectrum of C34-IAEDANS (Figure 3A). We observed that λ^{em}_{max} varies as a function of λ^{ex} in the N state ensemble (Figure 3B). For example, the representative fluorescence emission spectra at pH 7, as shown in Figure 3B, displayed a λ^{em}_{max} of 469 nm when excited at 337 nm. However, upon excitation at 410 nm, λ^{em}_{max} shifts by 6 nm to 475 nm. Interestingly, this effect was not observed in the U state (Figure 3B: inset) where the observed λ^{em}_{max} remained at ~483 nm irrespective of any excitation between 337 and 410 nm. When we plotted the λ^{em}_{max} as a function of λ^{ex} at different pHs, we observed that the N state ensemble showed a typical parabolic curve (Figure 3C). When excited from the center up to the red edge (in the wavelength range of 337−370 nm) of the absorbance spectrum, λ^{em}_{max} remained unchanged. Upon exciting at different wavelengths on the red edge (in the wavelength range of 371−410 nm), the λ^{em}_{max} steeply shifted toward red. However, this phenomenon was absent in the U states at different pHs (Figure 3C) where only a very small change in λ^{em}_{max} (1−2 nm) as a function of λ^{ex} was observed (Figure 3D). In the U forms, the protein structure is disrupted and C34-IAEDANS is solvated by bulk water molecules (Figure 2E,F). Because the solvation environment is highly dynamic in this case and the timescale of solvent relaxation (~picosecond) is faster than the fluorescence timescale (~nanosecond), very little REES is observed (Figure 1). In contrast, in the N forms, C34-IAEDANS is buried in the protein core and is solvated by the dipolar side chains of the protein matrix (Figure 2B). The high extent of the REES observed in N forms (Figure 3C,D) indicates that the photoexcitation at the red edge selectively excites a

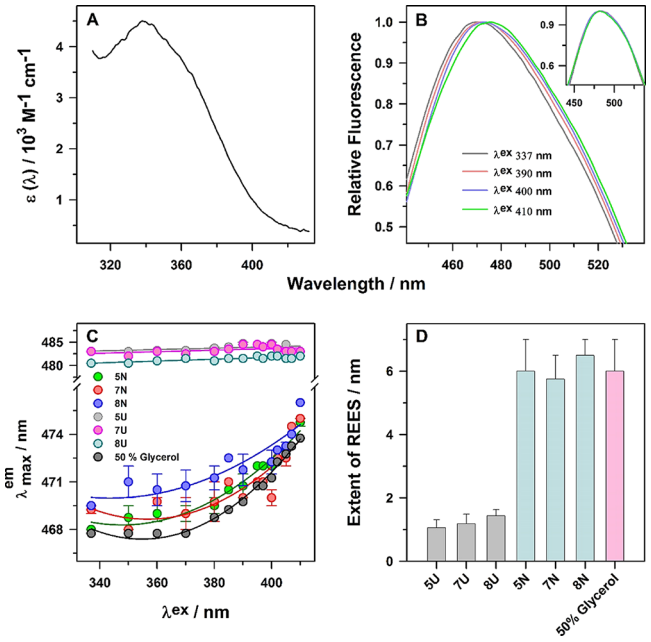


Figure 3. Core dynamics and heterogeneity of HSA as revealed by REES. (A) Molar absorptivity coefficient spectrum, ε (λ), of C34-IAEDANS is shown in panel (A). The absorption spectrum has been divided by the molar protein concentration to obtain ε (λ). (B) Fluorescence emission spectra of C34-IAEDANS when excited at different wavelengths in the red-edge regions in the N and the U states at pH 7 are shown in panel (B) and its inset, respectively. (C) Panel (C) shows the dependence of the wavelength of maximum fluorescence emission (λ^{em}_{max}) on the wavelength of excitation (λ^{ex}) of C34-IAEDANS in the N and the U states at pHs 5, 7, and 8 and in the presence of 50% glycerol at pH 7. (D) Extent of REES as observed in C34-IAEDANS at all of the U and N states at pHs 5, 7, and 8 and in the presence of 50% glycerol at pH 7.

subpopulation of protein molecules in which the fluorophore has a different solvation environment than the mean of the population distribution (Figure 1), and the side chains solvating the fluorophore reorient themselves on a timescale slower than the ~19 ns fluorescence lifetime of C34-IAEDANS. As C34-IAEDANS is majorly solvated by the side chains of buried polar residues (Q33, D38, H39, Y84, Y140, and R144) (Figure 2B), these results indicate that in the N forms, the orientations of these side-chain residues are different in the two population. Hence, our results suggest that the core packing of HSA is heterogeneous and the side chains solvating C34-IAEDANS have relatively restricted motions, similar to a viscous dense liquid.

Observed REES of C34-IAEDANS Is Independent of the Viscosity of the Exterior Solvent. One possible reason for the observed REES of C34-IAEDANS is that it might be measuring the slow dynamics of one or few ordered water molecules around it. This is, however, not the case because of the following two reasons: (i) in the crystal structure of HSA, no water molecules could be seen near C34 which is present in the core and (ii) most importantly, if the REES is due to the interfacial water molecules, then the extent of REES will depend upon the nature and viscosity of the exterior solvent (water in this case), as has been shown previously.[34] We examined this possibility by changing the viscosity and polarity of the exterior solvent with 50% glycerol/pH 7 buffer mixture in the N state. We observed that the λ^{em}_{max} as a function of λ^{ex} showed a similar parabolic trend in both the presence and the

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

absence of 50% glycerol in the external solvent (Figure 3C). The extent of REES remained similar (~6 nm) in the absence and the presence of 50% glycerol (Figure 3D). This result strongly indicates that the observed REES is due to the constrained environment of the protein matrix and not because of the dynamics of the interfacial water molecules around C34-IAEDANS.

REES of a Partially Solvent Exposed Residue, W214, Depends upon the Viscosity of the Exterior Solvent. The sole tryptophan residue, W214, of HSA is located at the interface of domain I and domain II and is partially exposed (~45%) to the exterior solvent as revealed by the crystal structure (Figure 2A). In previous studies, we[11,40] and others[46] have shown that W214 shows a small but distinct amount of REES because of the presence of ordered water molecules around it at the protein−water interface. We took the case of W214 as a positive control to check whether the REES due to interfacial water molecules could be modulated by the change in the viscosity and polarity of the solvent. For W214, in the N state at pH 7, the extent of REES was ~2 nm in response to the excitation from 295 to 305 nm (Figure 4A,B). However,

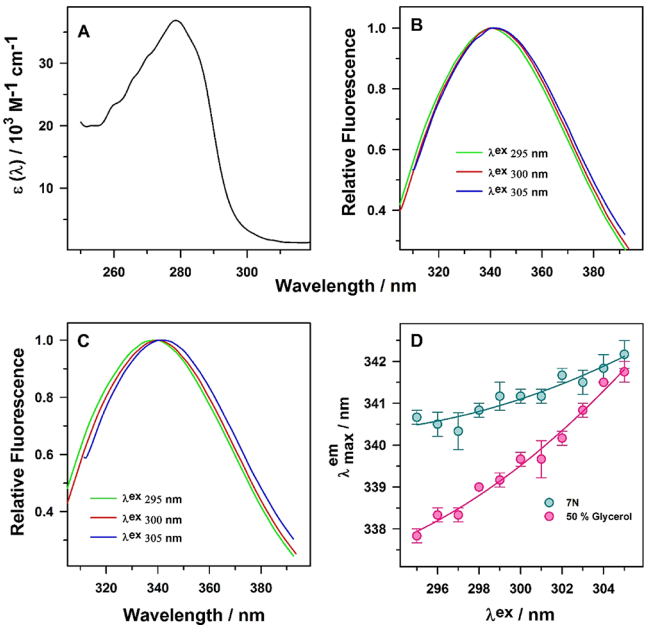


Figure 4. Change in viscosity of exterior solvent modulates the REES of W214 (A) the molar absorptivity coefficient spectrum, ε (λ) of W214. (B) Representative fluorescence emission spectra of W214 at pH 7 when excited at different wavelengths of excitation at the red edge of its absorbance spectrum. (C) Panel (C) shows the representative fluorescence emission spectra of W214 in the presence of 50% glycerol at pH 7 upon excitation at different wavelengths of red edge of its absorbance spectrum. For comparison, all of the spectra were normalized to 1 at their emission maximum. (D) Dependence of λ^{em}_{max} on λ^{ex} of W214 in the N state and in the presence of 50% glycerol at pH 7.

upon changing the exterior solvent to the 50% glycerol/pH 7 buffer mixture, the extent of REES dramatically increased to ~6 nm for the same excitation range (Figure 4C,D). These results show that the amount of REES is markedly perturbed by changing the solvent to 50% glycerol/pH 7 buffer mixture for a fluorophore near the solvent-exposed surface of the protein. Hence, the absence of any modulation of the extent of REES in the case of C34-IAEDANS, by increasing the viscosity of exterior solvent, indicates that the solvation environment of C34-IAEDANS is the protein matrix and not water, and the observed REES is due to the heterogeneity in the orientations of the protein side chains and their slower dipolar relaxation compared to the fluorescence timescale.

pH as a Modulator of Heterogeneity of the Core. Although the extent of REES for C34-IAEDANS remained same at all of the pH for the N forms (Figure 3D), the functional dependence with respect to pH appeared to be marginally different (Figure 3C). One possible reason for this could be the heterogeneity in the N state ensemble. As discussed above and reported earlier,[11] pH is known to modulate the equilibrium between different DMG-like near-native states which together form the N state ensemble. The changes in structural heterogeneity due to the side-chain packing of neighboring residues around the fluorophores, at different pHs, could give rise to this marginally different functional dependence. Another factor that might contribute here is the pK_a of neighboring polar and charged residues. The C34-IAEDANS which is the reporter of the protein core is being solvated mostly by the side chains of charged and polar residues, namely, Q33, D38, H39, Y84, Y140, and R144. These residues might have a different pK_a in the buried form in the N state than that in the water-exposed unfolded state. D38 and H39 could possibly titrate in the pH range 5−8 and that could be another reason for the marginally different trends in REES observed for the N state ensemble at the different pHs.

HSA Has a Compact and Highly Helical Secondary Structure but a Flexible and Loosely Packed Tertiary Structure. Above results indicate that side-chain residues inside the hydrophobic core of HSA are heterogeneously oriented with possibly alternate packing arrangements and that the population distribution might be modulated by pH.[11] We observed that this heterogeneous distribution of side-chain packing does not loosen and expand the N forms enough to allow the penetration of water molecules inside the protein core. HSA has 18 tyrosine residues distributed all over its 3 domains along with the sole tryptophan residue W214. The absorbance spectrum of tryptophan overlaps with the emission spectrum of tyrosine, and they form a fluorescence resonance energy transfer (FRET) pair.[40] Figure 5A shows the fluorescence spectra of HSA in the N and the U forms in the pH range 5−8, after excitation at 280 nm that excites both the tyrosine and tryptophan residues. The U form, at each pH, exhibited similar fluorescence spectra with two peaks with λ^{em}_{max} at 300 and 350 nm, corresponding to the emission because of the tyrosine and tryptophan residues, respectively. In contrast, the N forms at each pH exhibited similar fluorescence spectra, with the single peak with λ^{em}_{max} at 340 nm because of the emission of W214 buried in the protein structure. The peak at 300 nm due to the tyrosine residues is absent as the fluorescence emission of tyrosine residues is heavily quenched due to the spatial proximity to W214 in the N form.[40] These results suggest that at all of the pH the structure in the N form is compact. Site-specific FRET experiments in a previous study also showed similar results.[11] The observation that the λ^{em}_{max} of W214 in the N form is similar at each pH and shifted to blue by ~10 nm compared to the U form indicate that the core is not substantially hydrated in the alternatively packed N forms and it behaves like a DMG.

We observed that HSA has a highly helical but a loosely packed three-dimensional structure. Figure 5B shows the far-UV CD spectrum of HSA in the N and U forms in the pH

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

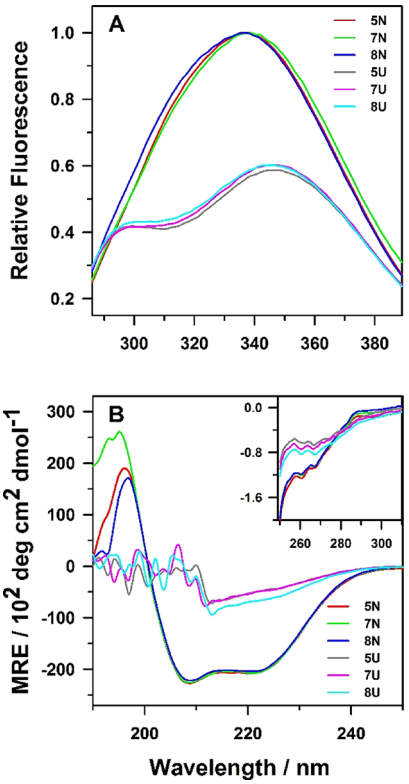


Figure 5. Global structural properties of HSA. (A) Fluorescence emission spectra of the protein at pHs 5, 7, and 8 in both the N and the U forms, when excited at 280 nm is shown. (B) Panel (B) and its inset show changes in MRE with respect to wavelength as a measure of global secondary and tertiary structures, respectively, in the N and the U forms at pHs 5, 7, and 8.

range 5−8. The N forms at pH each have similar far-UV CD spectra with two minima at 208 and 222 nm. The value of MRE at 208 nm is −22149 ± 913 deg cm^2 dmol^{-1} and that at 222 nm is −20331 ± 833 deg cm^2 dmol^{-1}, indicating a highly helical structure.[47] The inset in Figure 5B shows the near-UV CD spectra of HSA in the N and the U forms in the pH range 5−8. HSA absorbs weakly in the near-UV region, and its near-UV CD spectrum, at each pH, is similar and shows absorbance bands mainly in the 255−270 nm region because of the asymmetric packing of its 31 phenylalanine residues in the hydrophobic core. The absorbance in the 275−295 nm region due to the tyrosine and tryptophan residues is almost negligible. The value of MRE at 261 nm in the N form is −121 ± 3 deg cm^2 dmol^{-1} that is only slightly higher than that in the U form (−72 ± 10 deg cm^2 dmol^{-1}). The low values of MRE in the near-UV CD spectrum indicate that the three-dimensional structure of HSA is loosely packed. Hence, these results indicate that HSA has a highly helical structure but a flexible and loose tertiary structure. The loosely packed structure of HSA is likely to give rise to the heterogeneous distribution of side-chain packing indicated by the REES experiments.

N State Ensemble of HSA Contains a Number of DMG-Like States. The N state ensemble of HSA has a flexible structure (Figure 5B: inset), and it contains subpopulation of protein molecules in which the orientation and packing of some of the side-chain residues in the protein core are different than the mean population (Figures 3 and 1). These results indicate that the N state ensemble of HSA contains multiple DMG-like states. Although DMG-like states have been originally postulated to be hypothetical transition states,[48,49] they have been observed as partially unfolded intermediate states on the native side of the free-energy barrier of some proteins, both in simulations[10,50] and in experiments.[14−16,40,45,51−54] It has, however, been challenging to detect them in the N-state ensemble. Notably, DMG-like states have been observed in the N-state ensemble of the small protein HP35 by the triplet−triplet energy transfer method.[8,17] Our observation that N state remains in equilibrium with DMG-like conformations with heterogeneous side-chain orientations might be useful for multiple functions of HSA.[12] For example, HSA performs a multitude of functions in which binding and transportation of ligands hold maximum significance. The ligands are of a variety of chemical nature, shape, and size. The presence of multiple high energy binding sites could facilitate the binding by shifting the N ⇌ DMG equilibrium by preferentially stabilizing one particular high-energy state with respect to others. We observed that the side chains solvating C34-IAEDANS relax on a nanoscale or slower timescale. It is possible that many of these small fluctuations of side chains could couple to give large thermodynamic fluctuations required for conformational conversion and subsequent allosteric and entropy-driven functions.[12,13] These results also indicate that the core of HSA is like a molten dense liquid.

CONCLUSIONS

In summary, in this study, we explored the heterogeneity and dynamics of side-chain packing in the hydrophobic core of the native state ensemble of HSA using REES, a spectroscopic method that is well developed to study the dynamics of solvent molecules. Crystal structure and solvent accessibility calculations showed that C34 is buried in the protein structure and surrounded by side chains of polar and charged amino acid residues. We labeled C34 with a highly sensitive fluorescent dye 1,5-IAEDANS. The immediate solvation environment of C34-IAEDANS is the side chains of the polar and charged residues buried in the protein core that is well supported by fluorescence Stokes' shift and lifetime studies. REES experiments on HSA-IAEDANS revealed that the λ_{max}^{em} varies in a seemingly parabolic fashion as a function of λ^{ex} in the N state ensemble, whereas no such functional dependence was observed in the U forms. The extent of REES for C34-IAEDANS in the N state ensemble was observed to be independent of the viscosity of the exterior solvent. Global structural probes showed that the protein maintains a highly helical structure, but the aromatic side-chain residues in the core are loosely packed and the core is devoid of water molecules. All of these observations indicated that the side-chain residues inside the core have a heterogeneous distribution of spatial arrangement in the N state ensemble. These side-chain residues fluctuate much slower than the fluorescence timescale of C34-IAEDANS, and hence the core behaves like a molten dense liquid. These results together suggest that the N state of HSA is an ensemble of DMG-like states. HSA binds to a variety of ligands of different molecular structures and interconversion between different DMG-like states in the N state ensemble could be crucial to perform its functions.

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

AUTHOR INFORMATION

Corresponding Author

*E-mail: sk.jha@ncl.res.in. Phone: 91-20-25902588. Fax: 91-20-25902615.

ORCID

Santosh Kumar Jha: 0000-0003-1339-7409

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by DST-SERB early career research award (project # ECR/2015/000027) to S.K.J. P.M. is a recipient of Senior Research Fellowship by University Grants Commission, India.

REFERENCES

(1) Schrödinger, E. What is Life? *The Physical Aspect of the Living Cell*; The University press; The Macmillan company: Cambridge England, 1945.

(2) Zhou, Y.; Vitkup, D.; Karplus, M. Native proteins are surface-molten solids: application of the lindemann criterion for the solid versus liquid state 1 1Edited by A. R. Fersht. *J. Mol. Biol.* **1999**, *285*, 1371−1375.

(3) Bowman, G. R.; Geissler, P. L. Extensive Conformational Heterogeneity within Protein Cores. *J. Phys. Chem. B* **2014**, *118*, 6417−6423.

(4) DuBay, K. H.; Bowman, G. R.; Geissler, P. L. Fluctuations within Folded Proteins: Implications for Thermodynamic and Allosteric Regulation. *Acc. Chem. Res.* **2015**, *48*, 1098−1105.

(5) Richards, F. M. The Interpretation of Protein Structures: Total Volume, Group Volume Distributions and Packing Density. *J. Mol. Biol.* **1974**, *82*, 1−14.

(6) Chothia, C. Structural Invariants in Protein Folding. *Nature* **1975**, *254*, 304−308.

(7) Richards, F. M. Areas, Volumes, Packing, and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.

(8) Reiner, A.; Henklein, P.; Kiefhaber, T. An Unlocking/Relocking Barrier in Conformational Fluctuations of Villin Headpiece Subdomain. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4955−4960.

(9) Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry Molten Globule Intermediates and the Mechanism of Protein Unfolding. *Proteins* **2010**, *78*, 2725−2737.

(10) Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein Folding: From Theory to Practice. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22−29.

(11) Mishra, P.; Jha, S. K. An Alternatively Packed Dry Molten Globule-Like Intermediate in the Native State Ensemble of a Multidomain Protein. *J. Phys. Chem. B* **2017**, *121*, 9336−9347.

(12) Baldwin, R. L.; Rose, G. D. Molten Globules, Entropy-Driven Conformational Change and Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 4−10.

(13) Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J. Am. Chem. Soc.* **2017**, *139*, 3599−3602.

(14) Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. Direct NMR Evidence for an Intermediate Preceding the Rate-Limiting Step in the Unfolding of Ribonuclease A. *Nature* **1995**, *375*, 513−515.

(15) Jha, S. K.; Udgaonkar, J. B. Direct Evidence for a Dry Molten Globule Intermediate During the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12289−12294.

(16) Sarkar, S. S.; Udgaonkar, J. B.; Krishnamoorthy, G. Unfolding of a Small Protein Proceeds via Dry and Wet Globules and a Solvated Transition State. *Biophys. J.* **2013**, *105*, 2392−2402.

(17) Neumaier, S.; Kiefhaber, T. Redefining the Dry Molten Globule State of Proteins. *J. Mol. Biol.* **2014**, *426*, 2520−2528.

(18) Cooper, A. Thermodynamic Fluctuations in Protein Molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2740−2741.

(19) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964−972.

(20) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational Entropy in Molecular Recognition by Proteins. *Nature* **2007**, *448*, 325−329.

(21) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The Role of Conformational Entropy in Molecular Recognition by Calmodulin. *Nat. Chem. Biol.* **2010**, *6*, 352−358.

(22) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. Enzyme Dynamics During Catalysis. *Science* **2002**, *295*, 1520−1523.

(23) Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. Intrinsic Dynamics of an Enzyme Underlies Catalysis. *Nature* **2005**, *438*, 117−121.

(24) Ha, J.-H.; Loh, S. N. Changes in Side Chain Packing During Apomyoglobin Folding Characterized by Pulsed Thiol-Disulfide Exchange. *Nat. Struct. Biol.* **1998**, *5*, 730−737.

(25) Jha, S. K.; Udgaonkar, J. B. Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry. *J. Biol. Chem.* **2007**, *282*, 37479−37491.

(26) Jha, S. K.; Dasgupta, A.; Malhotra, P.; Udgaonkar, J. B. Identification of Multiple Folding Pathways of Monellin Using Pulsed Thiol Labeling and Mass Spectrometry. *Biochemistry* **2011**, *50*, 3062−3074.

(27) Bernstein, R.; Schmidt, K. L.; Harbury, P. B.; Marqusee, S. Structural and Kinetic Mapping of Side-Chain Exposure onto the Protein Energy Landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10532−10537.

(28) Demchenko, A. P. The Red-Edge Effects: 30 Years of Exploration. *Luminescence* **2002**, *17*, 19−42.

(29) Chattopadhyay, A.; Haldar, S. Dynamic Insight into Protein Structure Utilizing Red Edge Excitation Shift. *Acc. Chem. Res.* **2014**, *47*, 12−19.

(30) Demchenko, A. P. On the nanosecond mobility in proteins: edge excitation fluorescence red shift of protein-bound 2-(p-toluidinylnaphthalene)-6-sulfonate. *Biophys. Chem.* **1982**, *15*, 101−109.

(31) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Springer: Singapore, 2006.

(32) Lakowicz, J. R.; Keating-Nakamoto, S. Red-Edge Excitation of Fluorescence and Dynamic Properties of Proteins and Membranes. *Biochemistry* **1984**, *23*, 3013−3021.

(33) Arya, S.; Mukhopadhyay, S. Ordered Water within the Collapsed Globules of an Amyloidogenic Intrinsically Disordered Protein. *J. Phys. Chem. B* **2014**, *118*, 9191−9198.

(34) Chattopadhyay, A.; Mukherjee, S. Red Edge Excitation Shift of a Deeply Embedded Membrane Probe: Implications in Water Penetration in the Bilayer. *J. Phys. Chem. B* **1999**, *103*, 8180−8185.

(35) Haldar, S.; Chaudhuri, A.; Chattopadhyay, A. Organization and Dynamics of Membrane Probes and Proteins Utilizing the Red Edge Excitation Shift. *J. Phys. Chem. B* **2011**, *115*, 5693−5706.

(36) Guha, S.; Rawat, S. S.; Chattopadhyay, A.; Bhattacharyya, B. Tubulin Conformation and Dynamics: A Red Edge Excitation Shift Study. *Biochemistry* **1996**, *35*, 13426−13433.

(37) Catici, D. A. M.; Amos, H. E.; Yang, Y.; van den Elsen, J. M. H.; Pudney, C. R. The Red Edge Excitation Shift Phenomenon can be Used to Unmask Protein Structural Ensembles: Implications for Nemo-Ubiquitin Interactions. *FEBS J.* **2016**, *283*, 2272−2284.

(38) Chattopadhyay, A.; Rawat, S. S.; Kelkar, D. A.; Ray, S.; Chakrabarti, A. Organization and dynamics of tryptophan residues in erythroid spectrin: novel structural features of denatured spectrin revealed by the wavelength-selective fluorescence approach. *Protein Sci.* **2003**, *12*, 2389−2403.

(39) Chaudhuri, A.; Haldar, S.; Chattopadhyay, A. Organization and dynamics of tryptophans in the molten globule state of bovine α-lactalbumin utilizing wavelength-selective fluorescence approach:

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

Comparisons with native and denatured states. *Biochem. Biophys. Res. Commun.* **2010**, *394*, 1082−1086.
(40) Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J. Phys. Chem. Lett.* **2016**, *7*, 173−179.
(41) Painter, L.; Harding, M. M.; Beeby, P. J. Synthesis and Interaction with Human Serum Albumin of the First 3,18-Disubstituted Derivative of Bilirubin. *J. Chem. Soc. Perkin Trans. 1* **1998**, 3041−3044.
(42) Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **1986**, *131*, 266−280.
(43) Street, T. O.; Courtemanche, N.; Barrick, D. Protein Folding and Stability Using Denaturants. *Methods Cell Biol.* **2008**, *84*, 295−325.
(44) Ahmad, S.; Gromiha, M.; Fawareh, H.; Sarai, A. Asaview: Database and Tool for Solvent Accessibility Representation in Proteins. *BMC Bioinf.* **2004**, *5*, 51.
(45) Acharya, N.; Mishra, P.; Jha, S. K. A Dry Molten Globule-Like Intermediate During the Base-Induced Unfolding of a Multidomain Protein. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30207−30216.
(46) Demchenko, A. P. Red-Edge-Excitation Fluorescence Spectroscopy of Single-Tryptophan Proteins. *Eur. Biophys. J.* **1988**, *16*, 121−129.
(47) Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006**, *1*, 2876−2890.
(48) Finkelstein, A. V.; Shakhnovich, E. I. Theory of Cooperative Transitions in Protein Molecules. II. Phase Diagram for a Protein Molecule in Solution. *Biopolymers* **1989**, *28*, 1681−1694.
(49) Shakhnovich, E. I.; Finkelstein, A. V. Theory of Cooperative Transitions in Protein Molecules. I. Why Denaturation of Globular Protein is a First-Order Phase Transition. *Biopolymers* **1989**, *28*, 1667−1680.
(50) Mountain, R. D.; Thirumalai, D. Molecular Dynamics Simulations of End-to-End Contact Formation in Hydrocarbon Chains in Water and Aqueous Urea Solution. *J. Am. Chem. Soc.* **2003**, *125*, 1950−1957.
(51) Hoeltzli, S. D.; Frieden, C. Stopped-Flow NMR Spectroscopy: Real-Time Unfolding Studies of 6-^{19}F-Tryptophan-Labeled *Escherichia coli* Dihydrofolate Reductase. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9318−9322.
(52) Rami, B. R.; Udgaonkar, J. B. Mechanism of Formation of a Productive Molten Globule Form of Barstar†. *Biochemistry* **2002**, *41*, 1710−1716.
(53) Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled Motion in Proteins Revealed by Pressure Perturbation. *J. Am. Chem. Soc.* **2012**, *134*, 8543−8550.
(54) Jha, S. K.; Marqusee, S. Kinetic Evidence for a Two-Stage Mechanism of Protein Denaturation by Guanidinium Chloride. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 4856−4861.

DOI: 10.1021/acs.jpcb.8b11151
J. Phys. Chem. B 2019, 123, 1256−1264

Biophysical Chemistry 274 (2021) 106591

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

A pH-dependent protein stability switch coupled to the perturbed pKa of a single ionizable residue

Prajna Mishra [a,b,1], Divya Patni [a,b,1], Santosh Kumar Jha [a,b,*]

[a] Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India
[b] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

ARTICLE INFO

Keywords:
Thermodynamic stability
Electrostatic interactions
Unfolded proteins
Fluorescence spectroscopy

ABSTRACT

The contribution of electrostatic interactions in protein stability has not been fully understood. Burial of an ionizable amino acid inside the hydrophobic protein core can affect its ionization equilibrium and shift its pKa differentially in the native (N) and unfolded (U) states of a protein and this coupling between the folding/unfolding cycle and the ionization equilibria of the ionizable residue can substantially influence the protein stability. Here, we studied the coupling of the folding/unfolding cycle with the ionization of a buried ionizable residue in a multi-domain protein, Human Serum Albumin (HSA) using fluorescence spectroscopy. A pH-dependent change in the stability of HSA was observed in the near native pH range (pH 6.0–9.0). The protonation-deprotonation equilibrium of a single thiol residue that is buried in the protein structure was identified to give rise to the pH-dependent protein stability. We quantified the pKa of the thiol residue in the N and the U states. The mean pKa of the thiol in the N state was upshifted by 0.5 units to 8.7 due to the burial of the thiol in the protein structure. Surprisingly, the mean pKa of the thiol in the U state was observed to be down-shifted by 1.3 units to 6.9. These results indicate that some charged residues are spatially proximal to the thiol group in the U state. Our results suggest that, in addition to the N state, electrostatic interactions in the U state are important determinants of protein stability.

1. Introduction

The role of electrostatics in protein function, solubility and stability has long been recognized [1–7]. The major contributors of electrostatic forces in proteins are the charged states of individual ionizable residues. More often, ionizable amino acid residues are excluded from the interior of the folded proteins because of their intrinsic incompatibility with the hydrophobic environment of the core [8–12]. However, they occasionally get buried inside the protein core to perform certain crucial biological functions [13–15]. This energetically unfavorable process of burial, where ionizable amino acids get sequestered into the less polar protein core from bulk water, allows them to experience different microenvironments, bringing a shift in their pKa. The value of pKa, which gives the information about the equilibrium between the charged and neutral state of an ionizable amino acid, can be modulated by multiple factors. Apart from the dielectric constant of the environment of the ionizable residue, coulombic interactions and hydrogen bonding with nearby polar/ionic residues can affect the pKa of the residue [16]. However, the structural adaptations to facilitate the burial of an ionizable residue inside the core and its effect on protein stability are not fully understood [17–20].

The stability of a protein is determined by the difference in the free energies of its native (N) and the unfolded (U) states. Therefore, the knowledge of the parameters that influence the stabilities of the N and the U states becomes crucial. pH, being a well-known modulator of electrostatic interaction, can affect the thermodynamic stabilities of the N and the U states in different ways. A buried ionizable amino acid can titrate differently as a function of pH in the folded and the unfolded conformations of the protein [17,19,21]. In addition, contrary to the initial notion, recent reports on the U states have emphasized on the presence of nonrandom conformers with local and long range interactions similar or different to their respective native structures [22–24]. These interactions shift the pKa of unfolded protein from its standard value. This disparity in the pKa values of the residue in the N and the U states, therefore, gives rise to a pH dependent protein stability. The site-specific measurement of the pKa values of the buried ionizable

* Corresponding author.
E-mail address: sk.jha@ncl.res.in (S.K. Jha).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bpc.2021.106591
Received 4 February 2021; Received in revised form 15 March 2021; Accepted 7 April 2021
Available online 20 April 2021

residue becomes essential for the detailed understanding of the thermodynamic stability of the protein.

Limited methods have been able to measure the pKa of an ionizable residue in the N and the U states of a protein reliably. Although theoretical studies have succeeded in calculating the pKa of the ionizable residues of a protein, the calculations are yet to attain the experimental accuracy [25,26]. The major complications arise in determining the molecular force field factors due to the heterogeneous environment experienced by the buried residues and generating the precise conformational unfolded state ensemble of the protein [25–27].

Experimentally pKa can be determined by studying the coupling of thermodynamics of folding and unfolding of a protein to the ionization equilibria of a buried ionizable residue (Fig. 1). For the illustration, the model protein in its N state is assumed to have an acidic amino acid (A–H) buried inside its hydrophobic core and there exists an equilibrium between the protonated / neutral and deprotonated / charged states of the acidic residue in the folded and unfolded conformations of the protein (Fig. 1). The four species in equilibrium are represented as N^P, N^D, U^P and U^D in which N and U are native and unfolded conformations, respectively, and the superscripts P and D, respectively, denote the protonated and deprotonated states of the ionizable residue. K^P_{NU} and K^D_{NU} are the equilibrium constants of the equilibrium between the folded and unfolded conformations, respectively, in their protonated and deprotonated states. K^N_a is the acid dissociation constant of $N^P \rightleftharpoons N^D$, while K^U_a are the acid dissociation constant of the equilibrium between U^P and U^D. At any pH, the standard free energy of unfolding, which is a measure of the thermodynamic stability of the protein, $\Delta G^{H_2O}_{NU}$ is given by [17]:

$$\Delta G^{H_2O}_{NU} = \Delta G^{H_2O,NE}_{NU} + \Delta G^{H_2O,pH}_{NU} \quad (1)$$

where $\Delta G^{H_2O,\ NE}_{NU}$ and $\Delta G^{H_2O,\ pH}_{NU}$ are the non-electrostatic contribution (involving folding / unfolding equilibria) and electrostatic contribution (involving protonation / deprotonation equilibria) to $\Delta G^{H_2O}_{NU}$. The

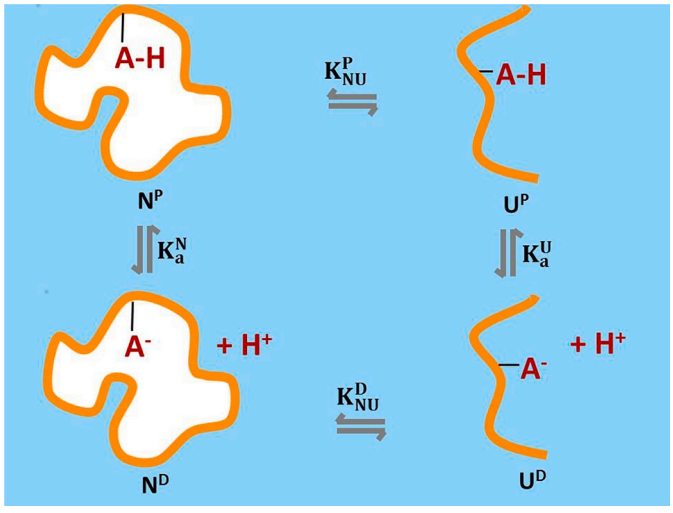


Fig. 1. The coupling of the folding / unfolding cycle of proteins with the ionization of an ionizable amino acid residue buried in the protein core. The orange lines in the figure represent the protein backbone. A-H is the protonated acidic ionizable residue in its neutral form buried inside the core (white patch) of the fully folded protein and exposed in its unfolded counterpart. A^- refers to the deprotonated / charged state of the acidic residue. N^P and U^P, respectively, refer to the folded and unfolded state of protein when the ionizable amino acid is protonated and K^P_{NU} is the equilibrium constant between them. N^D and U^D are, respectively, the folded and unfolded conformations of the protein when the ionizable amino acid is deprotonated / charged and K^D_{NU} refers to their equilibrium constant. K^N_a is the equilibrium acid dissociation constant between the charged and neutral states of the native protein. Similarly, the acid dissociation constant for the equilibrium between the charged and the neutral states of the unfolded protein is represented by K^U_a.

electrostatic contribution to the standard free energy can be calculated from the following [17]:

$$\Delta G^{H_2O,pH}_{NU} = -2.303RT\ log\left(\frac{[H^+] + K^U_a}{[H^+] + K^N_a}\right) \quad (2)$$

This thermodynamic cycle that links the equilibria between the charged and uncharged states of a protein to its folding / unfolding cycle has been implemented in various cases to determine the pKa of an ionizable residue in the N and the U states of a protein. NMR has been the method of choice for most of these studies [17,19,24,28]. Although an elegant approach, NMR can only be applied to small and soluble proteins and has limited applications for large, multi-domain and aggregation-prone proteins.

In this work, we have tried to understand the pH-dependence of stability of a multi-domain protein, human serum albumin (HSA). HSA is an all helical protein found abundantly in blood plasma and plays a distinct role in the transportation of metabolites, drugs and ions in the bloodstream and the maintenance of the pH of the plasma and extravascular fluids [29,30]. HSA maintains its native structure in the pH range of 5.0–9.0 [29,31,32]. However, it is also known to undergo several pH dependent conformational transitions. The N form of HSA converts to a fast migrating form (F) in the pH range of 5.0–3.5 followed by a transition into the acid expanded or the extended (E) form below pH 3.5 [29,31,33,34]. Above pH 9.0, the N form transforms to a basic (B) form [29,31,35]. We explored the pH dependence of the thermodynamic stability of HSA in the N state ensemble, *i.e.*, in the near neutral pH range of 6.0–9.0. This is an important pH range because the functional pH of HSA in plasma is ~7.4. The crystal structure of HSA reveals that it consists of three domains and each domain is divided into two subdomains. A cluster of hydrophobic residues holds together the three domains of HSA in its N state [32]. The sole tryptophan residue (W214) is located in the interdomain cluster (Fig. 2A). We used the fluorescence method to understand the coupling of the folding / unfolding cycle of HSA with the ionization of the ionizable residues. We show that the pH dependent changes in the stability of HSA in the pH range of 6.0–9.0 are due to perturbed ionization of a single thiol group (C34). C34 is known to impart antioxidant properties to HSA as its reduced form is highly abundant in blood plasma [36–38]. This makes C34 an important target for reactive oxidizing species [37] and the determination of its pKa in the N and the U states of the protein becomes crucial considering its redox potential. We calculated the pKa of C34 in the N and the U states. Our results show that the pKa of the thiol is also perturbed in the U state in addition to the N state indicating that the charged interactions in the U state are also important for the thermodynamic stability of the proteins.

2. Materials and methods

2.1. Spectroscopic methods and instruments

Fluorescence experiments were performed on Fluoromax-4 spectrofluorometer from HORIBA Scientific, while absorption studies were done using UV 3200 spectrophotometer procured from LABINDIA Analytical. Quartz cuvettes of path length of 1 cm were used for signal acquisition for both fluorescence and absorption studies. Circular Dichroism (CD) studies were performed on Jasco J-815 CD spectrometer. For far-UV CD experiments, cuvettes of path lengths 1 mm were used. An Abbe refractometer from Rajdhani Scientific Instruments Co. (model: RSR-2) was used to calculate the denaturant concentration by measuring the refractive index [39].

2.2. Chemicals and buffers

HSA (99% pure, fatty acid and globulin free) and urea (ultrapure grade) were sourced from Alfa Aesar. 5-((((2-Iodoacetyl)amino)ethyl)

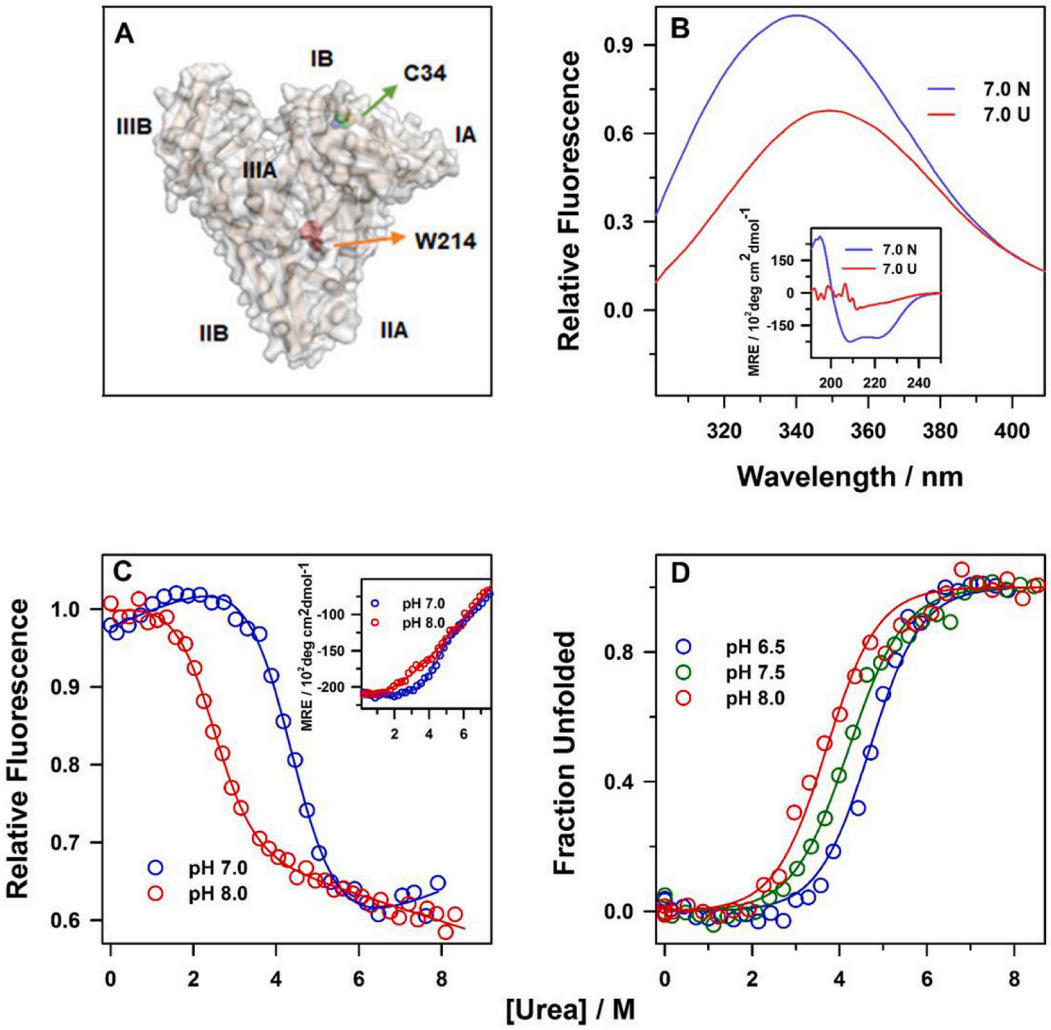


Fig. 2. The thermodynamic stability of HSA is pH dependent. (A) Structure of HSA, depicting its six sub-domains and the locations of free cysteine residue (C34) and single tryptophan residue (W214). This image was drawn from the pdb file 1AO6 using the PyMol Molecular Graphics System, Version 2.0 Schrodinger, LLC. (B) Fluorescence emission spectra of W214 in the native (N) and unfolded (U) states of HSA at pH 7.0. The inset of panel B shows the change in the mean residue ellipticity (MRE) of HSA in the N and U states at pH 7.0 in the far-UV region. (C) Urea induced equilibrium unfolding transitions of HSA at pH 7.0 and pH 8.0 as monitored by the changes in the fluorescence signal of W214 at 340 nm and change in the MRE at 222 nm (inset). (D) The fractions of the unfolded protein at pH 6.5, pH 7.5 and pH 8.5 are plotted as a function of the [urea]. The solid lines through the data in panel C, inset of panel C and panel D are fits to a two-state, N⇌U model (Eqs. (3) and (5), Section 2.5). The plots in Fig. 2C, C inset and 2D are representative plots of one set of experiments.

amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was purchased from Life Technologies. All other chemicals of the highest purity grade were purchased from HiMedia and used directly without further purification. HSA concentration was measured using an extinction coefficient of 36,500 M^{-1} cm^{-1} by measuring absorbance at 280 nm [40]. For all the pH titration studies, native buffers consisted of 20 mM MES, 20 mM phosphate and 20 mM Tris-HCl for pH 6.0, pH 6.5 - pH 7.5 and pH 8.0 – pH 9.0, respectively. The unfolding buffers constituted of 9 M urea in their respective native buffer.

2.3. 1,5-IAEDANS labeled HSA preparation

We followed the previously reported protocol for the site-specific covalent labeling of HSA [32]. In brief, 20 fold molar excess of 1,5-IAEDANS was added to HSA, which was previously unfolded in 6 M Guanidine Hydrochloride (GdmCl) and 20 mM Tris-HCl at pH 8.0. The reaction mixture was stirred for 4 h in the dark followed by refolding in 15 fold volume excess of 20 mM phosphate buffer at pH 7.0 with an overnight incubation at 4 °C. Refolded protein was then concentrated by centrifugal concentrator (GE healthcare, molecular weight cutoff-30 kDa) to 2.5 mL. To remove the remaining unbound dye and GdmCl, the concentrated reaction mixture was passed through a PD-10 desalting column (GE healthcare). The percentage of labeling in the labeled protein (HSA-IAEDANS) was estimated as discussed previously [32] and was observed to be >95%.

2.4. Fluorescence and CD experiments

For fluorescence experiments, the protein concentration used was 4–6 μM. In HSA and HSA-IAEDANS, W214 was excited at 295 nm and the emission spectra were collected from 310 nm – 550 nm. 1,5-IAEDANS attached to the C34 (C34-IAEDANS) was excited at 337 nm. The excitation slit widths for all the fluorescence experiments were kept between 1.0 nm to 1.5 nm, while the emission slit widths used were in the range of 8–12 nm.

For far-UV CD measurements, 3–4 μM of protein concentrations were used. Scans were collected from 190 nm - 250 nm. The data pitch, data integration time, scan speed and bandwidth were kept at 1 nm, 1 s, 50

nm/min and 2 nm, respectively, for all the experiments. Background signal corrections have been done for all the CD and fluorescence experiments by subtracting the buffer signal.

2.5. Urea induced equilibrium unfolding experiments

For urea induced equilibrium unfolding experiments, protein samples (HSA and HSA-IAEDANS) were subjected to a gradient of urea concentration and incubated for 3 h at room temperature in the pH range of 6.0–9.0. For fluorescence measured experiments, W214 and C34-IAEDANS were excited at 295 nm and 337 nm, respectively, and equilibrium unfolding was measured by the change in respective fluorescence signal at 340 nm and 469 nm, respectively. For far-UV CD measured equilibrium unfolding experiments, the signals were measured at 222 nm. The data were fitted to the following equation based on a two state, N⇌U model [32,41].

$$y_{obs} = \frac{y_N + y_U\, e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (3)$$

In the above equation, y_{obs} is the observed fluorescence / CD signal; y_N and y_U are the signals of the N and the U states, respectively; ΔG_{NU} is the free energy of unfolding of N⇌U, which has a linear dependency on denaturant concentration and is given by:

$$\Delta G_{NU} = \Delta G_{NU}^{H_2O} + m_{NU}[D] \quad (4)$$

where $\Delta G_{NU}^{H_2O}$ is the standard free energy of unfolding in 0 M urea and m_{NU} is the slope of the N⇌U transition.

The fractions of unfolded states, f_U, at any particular denaturant concentration were determined as per the equation mentioned below:

$$f_U = \frac{e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NU}}{RT}}} \quad (5)$$

All the urea induced denaturation experiments shown in Figs. 2C, D, 3C and D were repeated two or more times. The error in thermodynamic parameters shown in Figs. 4 and 5 were determined from the spread of two or more independent measurements.

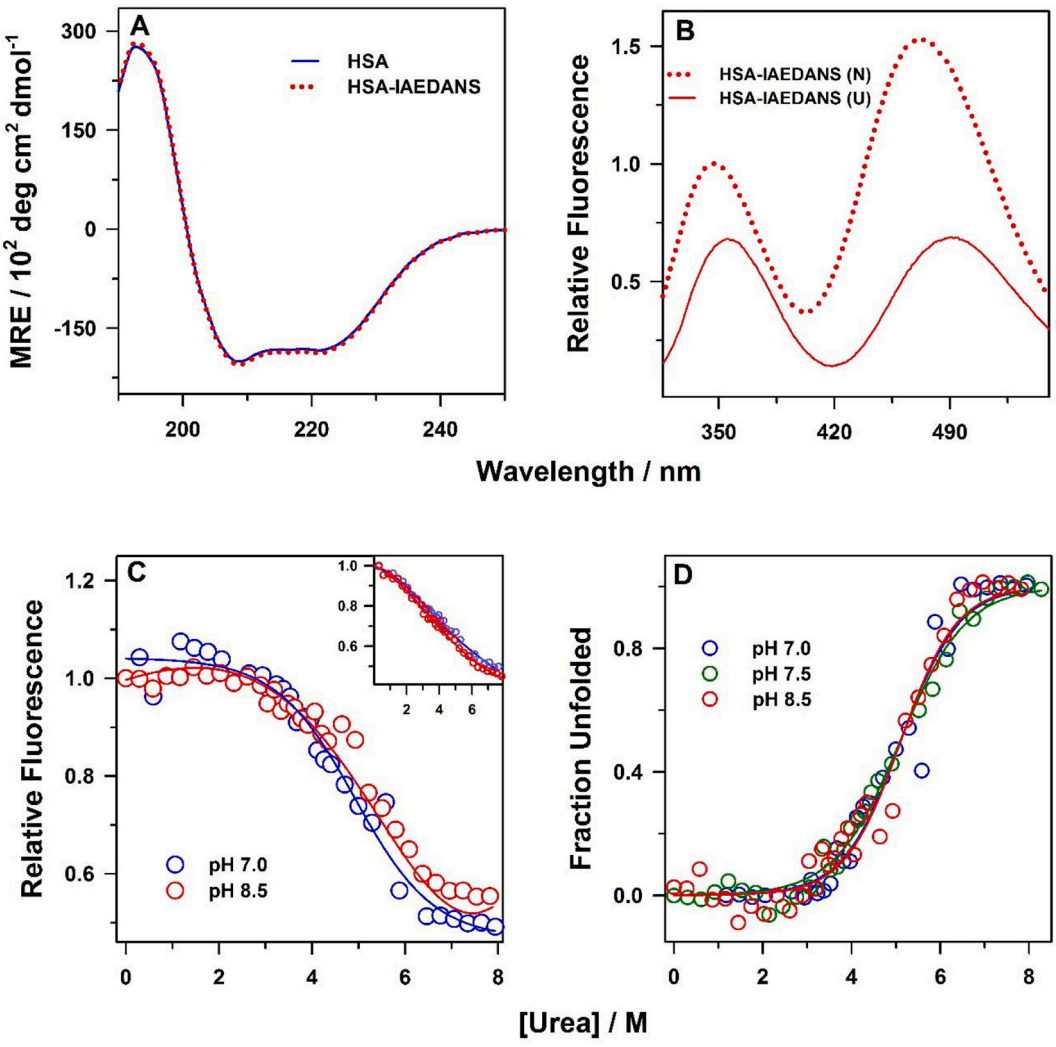


Fig. 3. The thermodynamic stability of HSA-IAEDANS is independent of pH. (A) far-UV CD spectra of HSA and HSA-IAEDANS at pH 7.0 as a measure of their global secondary structure content are shown in panel A. (B) Fluorescence emission spectra of W214 and C34-IAEDANS, when excited at 295 nm, in the N and the U states of HSA-IAEDANS at pH 7.0 are shown. (C) Urea induced equilibrium unfolding transitions of HSA-IAEDANS as monitored by the change in W214 fluorescence at 340 nm at pH 7.0 and pH 8.5. The inset shows the pH dependence of the urea induced equilibrium unfolding transitions as monitored by the change in fluorescence signal of C34-IAEDANS at 469 nm at pH 7.0 and pH 8.5. (D) The fractions of unfolded protein were plotted as a function of [urea] at pH 7.0, pH 7.5 and pH 8.5. The solid lines through the data in panel C, inset of panel C and panel D are fit to a two-state, N⇌U model. In the Fig. 3C, C inset and 3D, representative plots of one set of experiments are shown.

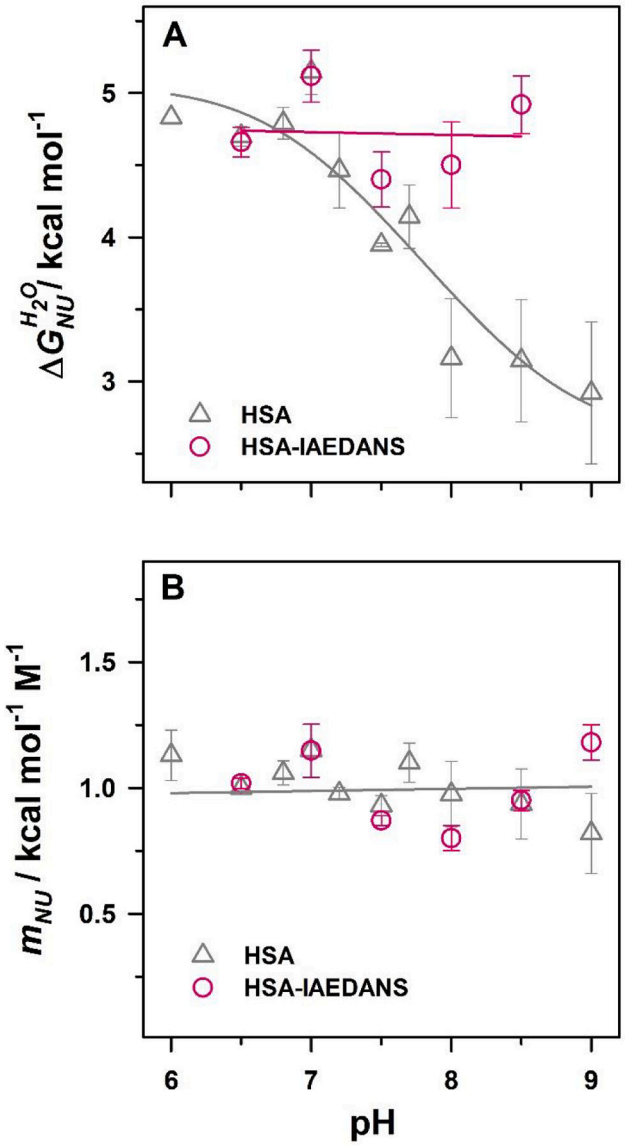


Fig. 4. pH dependence of the thermodynamic parameters. (A) Changes in the standard free energy of unfolding ($\Delta G_{NU}^{H_2O}$) of HSA and HSA-IAEDANS as a function of pH. The solid line through the data of HSA is a least-square fit to Eq. (6), while that of HSA-IAEDANS was drawn to guide the eye. (B) Dependence of change in solvent exposed surface area (m_{NU}) during N⇌U transition of HSA and HSA-IAEDANS with respect to pH. Error bars shown in panel A and panel B represent the spread for at least 2 or more independent sets of experiments.

2.6. Analysis of titration of thermodynamic parameters as a function of pH

Assuming that the single titratable group (C34) in HSA has different pKa values in the native (pK_a^N) and the unfolded states (pK_a^U), the dependency of the free energy of unfolding, $\Delta G_{NU}^{H_2O}$ on pH was analyzed using the following equation [17,20]:

$$\Delta G_{NU}^{H_2O} = \Delta G_{NU}^{H_2O,NE} - 2.303RT\ log\left(\frac{1+10^{pK_a^U-pH}}{1+10^{pK_a^N-pH}}\right) \quad (6)$$

The above equation is a combined and modified form of Eqs. (1) and (2) (see Section 1).

3. Results and discussion

3.1. The N and U states of HSA have distinct spectroscopic properties

We first compared the fluorescence signal of W214 in the N and the U states of HSA at pH 7.0 and we observed that the signal from the U state was highly quenched and the wavelength of the maximum fluorescence emission (λ_{max}^{em}) in the U state (349 nm) was red shifted as compared to the N state (340 nm), pertaining to the hydration of W214, which lead to the quenching of its fluorescence (Fig. 2B). Similarly, the global secondary structure content of HSA in the N and U states were distinctly different, as revealed by the far-UV CD spectra with the global secondary structure being entirely lost in the U state (Fig. 2B, inset). The mean residual ellipticity (MRE) at 222 nm for the N and the U states of HSA were −18,690 and −4868 deg. cm^2 dmol^{-1}, respectively. All these results conveyed that the N and the U states of HSA have distinct spectroscopic properties.

3.2. The thermodynamic stability of HSA is pH dependent

We utilized the difference in spectroscopic properties of the N and the U states of HSA to monitor their equilibrium population distribution as a function of urea concentration and the thermodynamic stability of HSA over a range of pH (pH 6.0 – pH 9.0). We first used W214 fluorescence as a measure of structure loss upon unfolding during the N⇌U transition. Urea induced single sigmoidal equilibrium unfolding curves for HSA were obtained in the pH range of 6.0 – 9.0. Two representative plots at pH 7.0 and pH 8.0 are shown in Fig. 2C. We observed that the midpoint of the N⇌U transition (C_m) changed as a function of pH in the pH range of 7.0 – 8.0. We further compared the thermodynamic stability of the secondary structure of HSA as measured by far-UV CD at pH 7.0 and pH 8.0 (Fig. 2C, inset). The slopy nature of the unfolded baseline in the equilibrium unfolding experiments monitored by far-UV CD limited the correct estimation of the thermodynamic parameters. Nevertheless, we observed that, at pH 7.0, the loss of the protein secondary structure was initiated ~3 M [urea] and C_m was ~4.5 M [urea]. However, at pH 8.0, the unfolding of HSA started ~1.5 M [urea] and the observed C_m was ~3 M [urea]. These results were in very well correspondence with the W214 fluorescence measured unfolding transitions (Fig. 2C). W214 fluorescence is also a measure of the change in the tertiary structure. These results suggested that the global secondary and tertiary structure of the protein dissolved similarly during urea induced unfolding studies. The unfolding curves as measured by W214 fluorescence were converted into fractions of unfolded protein and were analyzed using a two-state N⇌U model (Fig. 2D, Section 2.5). The mean value of standard free energy of unfolding of the N⇌U transition,$\Delta G_{NU}^{H_2O}$, at pH 6.5 was 4.8 kcal mol^{-1}, at pH 7.5 was 4.0 kcal mol^{-1} and at pH 8.0 was 3.1 kcal mol^{-1}. The mean value of the change in the solvent accessible surface area (m_{NU}), as measured by the slope of the N⇌U transition, was 1.00 kcal mol^{-1} M^{-1} at pH 6.5, 0.94 kcal mol^{-1} M^{-1} at pH 7.5 and 0.98 kcal mol^{-1} M^{-1} at pH 8.0. These results show that the thermodynamic stability of HSA is dependent on pH and decreases upon increasing the pH above 6.5. However, the value of m_{NU} remains unperturbed.

3.3. C34-IAEDANS is buried in the protein core

Since pH is a well-known modulator of the structure of proteins and, by extension, their thermodynamic stabilities [33,35,42], it can be argued that the observed dependence of thermodynamic stability of HSA on pH was due to the change in its structural properties. However, it has been previously reported that HSA conserves its N-like structure in the pH range of 4.8 to 8.5 [29,31–33]. The similar values of m_{NU} at all the pH (Fig. 2D), which measures the change in solvent exposed surface area, also indicated that the pH dependence of the stability was not due to a structural change. The other possibility for the destabilization of protein in the pH range of 7.0 to 8.0 could be due to the titration of a

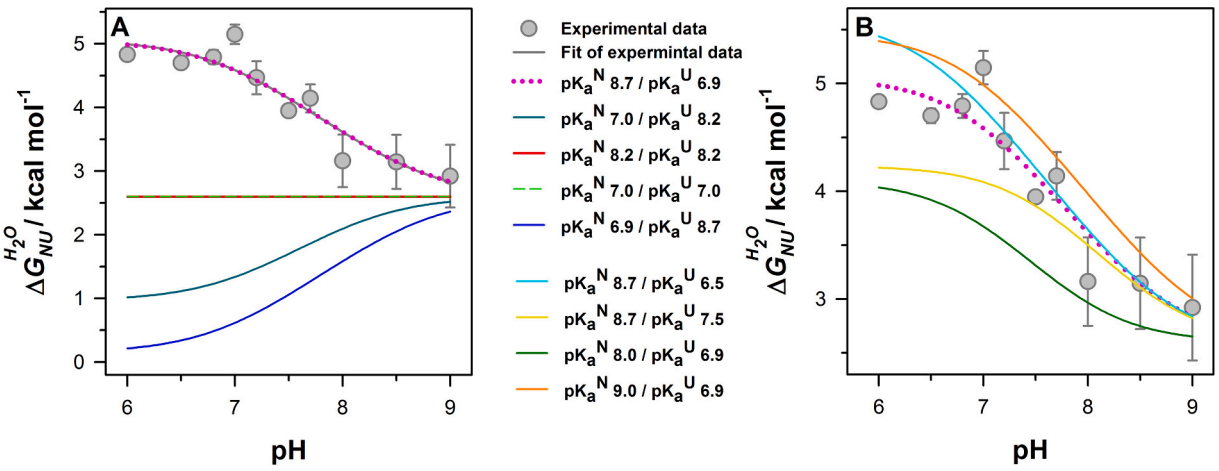


Fig. 5. The pKa of C34 in the N state is upshifted, while in the U state, it is downshifted with respect to a free thiol group. (A) The simulated $\Delta G_{NU}^{H_2O}$*versus* pH curves when i) $pK_a^U = 6.9$ and $pK_a^N = 8.7$ (pink dotted line) and hence are equal to the experimental value (gray filled circles); ii) $pK_a^U = pK_a^N$ (red line and green dashed line); and iii) $pK_a^U > pK_a^N$(blue line and dark cyan dashed line). (B) Comparison of simulated $\Delta G_{NU}^{H_2O}$*versus* pH curves by keeping pK_a^Nconstant while varying pK_a^U and *vice versa.* All the curves were simulated using Eq. (6). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

buried ionizable amino acid residue inside the core of the protein. Since free cysteine has a pKa ~8.2 and upon burial inside a hydrophobic core, its pKa is likely to get modulated, we hypothesized that the titration of a buried cysteine might be leading to the protein destabilization in this pH range. HSA has 35 cysteines, out of which 34 forms disulfide linkages in the protein. The sole free cysteine, C34, has its side chain buried in the hydrophobic core of protein (Fig. 2A) [43]. The titration of the protonation-deprotonation equilibrium of cysteine could be blocked in two ways: i) by site-directed mutation of C34 to a non-ionizable residue and ii) test-tube engineering in which C34 is covalently labeled with a molecule that prohibits its ionization. As we had successfully labeled the C34 and performed multiple experiments with labeled HSA, previously [32,33,35,43], we opted for the second alternative. To block its titration with pH, we labeled C34 with a fluorophore 1,5-IAEDANS (C34-IAEDANS) (Section 2.3). It is important to confirm that labeling does not alter the conformational properties of the protein. We compared the global secondary structural difference between the unlabeled HSA and the labeled HSA (HSA-IAEDANS) by far-UV CD (Fig. 3A). The MRE for HSA and HSA-IAEDANS at 222 nm were −18,690 deg. cm^2 dmol^{-1} and −18,282 deg. cm^2 dmol^{-1}, respectively, which implied that the labeling of HSA did not perturb its secondary structural content. We further compared the tertiary structural difference of HSA and HSA-IAEDANS as measured by W214 fluorescence and we observed that λ_{max}^{em} in the N and the U states of HSA did not change upon labeling (Fig. 2B, Fig. 3B). In previous global tertiary structural studies, low values of MRE at 261 nm in the N state (−121 ± 3 deg. cm^2 dmol^{-1}) was observed which was marginally higher than the MRE value of the U state (−72 ± 10 deg. cm^2 dmol^{-1}) [43]. This result conveyed that the global tertiary structure of HSA is flexible and loosely packed. So it is not surprising that the loose packing of the N state of HSA is being able to accommodate 1,5-IAEDANS without disrupting its conformation. In our earlier studies, we have shown that W214 and C34-IAEDANS form a FRET pair [32,33,35]. Upon excitation of W214 (Donor) at 295 nm, the fluorescence of C34-IAEDANS (Acceptor) increased significantly in the N state with a λ_{max}^{em} of 469 nm due to FRET as compared to the U state (λ_{max}^{em} of 483 nm) (Fig. 3B). The differences in the fluorescence intensity and λ_{max}^{em} in the N and U states of HSA-IAEDANS implied that the N and the U states have characteristic spectroscopic properties and C34-IAEDANS is buried in the core of the protein in the N state.

3.4. Thermodynamic stability of HSA-IAEDANS does not change with pH

In order to investigate the dependence of the thermodynamic stability of HSA-IAEDANS on pH, we performed the urea induced equilibrium unfolding experiments on HSA-IAEDANS in the pH range of 6.0 – 9.0. We first monitored the W214 fluorescence as a function of [urea] at different pH and observed that the sigmoidal curves at 340 nm for the N⇌U transition perfectly overlapped on each other (Fig. 3C). We also checked C34-IAEDANS fluorescence to compare the thermodynamic stability of HSA-IAEDANS (Fig. 3C, inset). The equilibrium values at 469 nm showed a gradual and non-cooperative dependence on [urea] which limited the correct estimation of the thermodynamic parameters. This observation could be due to the high fluorescence sensitivity of C34-IAEDANS. Nevertheless, the C_m values obtained from W214 and C34-IAEDANS fluorescence measurement studies in the pH range of 6.0 – 9.0 were analogous. We converted the unfolding curves of HSA-IAEDANS acquired from W214 fluorescence measurements to fractions of unfolded protein (Fig. 3D) and analyzed them using a two-state N⇌U model. The obtained thermodynamic parameters, both $\Delta G_{NU}^{H_2O}$ and m_{NU} were strikingly similar in the pH range of 6.0–9.0. All these results suggested that the thermodynamic stability of HSA-IAEDANS was pH independent.

3.5. Titration of C34 was responsible for the destabilization of HSA in the pH range of 6.0–9.0

We compared the change in thermodynamic parameters, $\Delta G_{NU}^{H_2O}$ and m_{NU}, governing the denaturation of HSA and HSA-IAEDANS as a function of pH (Fig. 4A and B). Fig. 4A shows that the dependency of $\Delta G_{NU}^{H_2O}$ on pH in HSA was sigmoidal, while in HSA-IAEDANS, this dependency was abolished. Above pH 6.0, the thermodynamic stability of HSA decreased in a sigmoidal manner and plateaued near pH 8.5 to pH 9.0. The m_{NU} values for HSA and HSA-IAEDANS did not change with pH (Fig. 4B). All the above results clearly conveyed that the titration of buried C34 was responsible for the pH dependency of $\Delta G_{NU}^{H_2O}$ of HSA. We fitted the pH dependency of $\Delta G_{NU}^{H_2O}$ of HSA (Fig. 4A) to Eq. (6) to obtain the pKa values of C34 in the N and the U states (Section 2.6). For a cysteine residue buried in the protein core, an increase in its pKa value is expected. For the U state, the anticipated pKa was similar to that of free cysteine in water (8.2). The pKa value obtained for the buried C34 in the N state was 8.7 ± 0.5. Interestingly, in the open U state, the pKa value of C34 was 6.9 ± 0.3.

The observed decrease in the mean pKa of the C34 in the U state by 1.3 units was quite significant which could be due to the presence of electrostatic interactions near C34 in the U state. However, it is important to confirm that the observed pKa of the thiol group in the N and the U states were not an artifact of the fitting errors. In order to gauge the reliability of the values of pK_a^N and pK_a^U obtained from the fit, we simulated and compared the graphs of $\Delta G_{NU}^{H_2O}$ as a function of pH for different values of pK_a^N and pK_a^U, using Eq. (6) (Fig. 5). Fig. 5A shows that when $pK_a^N = pK_a^U$, no pH-dependent change in the values of $\Delta G_{NU}^{H_2O}$ will be observed. Fig. 5A also shows that when $pK_a^N < pK_a^U$, the value of values of $\Delta G_{NU}^{H_2O}$ will increase with pH and not decrease, as observed experimentally. Hence, for the experimentally observed dependence of $\Delta G_{NU}^{H_2O}$ on pH, pK_a^N must be greater than pK_a^U. Fig. 5A shows that for the values of $pK_a^N = 8.7$ and $pK_a^U = 6.9$, the simulated curve fits the experimental data very well.

We further checked the robustness of the observed values of pK_a^U and pK_a^N by simulating $\Delta G_{NU}^{H_2O}$*versus* pH plots, keeping one parameter constant and varying the other by about ±0.5 pH units (Fig. 5B). The generated $\Delta G_{NU}^{H_2O}$*versus* pH curves markedly deviated from the experimental data (Fig. 5B). Collectively these results confirmed that the experimentally obtained values of pK_a^U and pK_a^N of C34 were quite reliable.

3.6. The electrostatic interactions in both the N and the U states are important for the thermodynamic stability of proteins

Our results show that the protonation-deprotonation equilibrium of a single ionizable residue, C34, modulates the thermodynamic stability of HSA in the near neutral pH range of 6.0 – 9.0 (Fig. 4). This result is significant considering the vital role HSA plays in maintaining the pH of the blood plasma, extravascular fluids and ascitic fluid [30]. In the crystal structure of HSA, ~99% of the total surface area of C34 is buried inside the core of domain I of HSA in the N state and is solvated by the side-chains of 12 amino acid residues out of which 2 basic residues (H39 and R144), 1 acidic residue (D38) and 3 polar residues (Q33, Y84 and Y140) directly affect the pKa of C34 [43]. The interactions of C34 with H39 and R144 will be stabilizing and will lead to a decrease in its pKa; whereas, D38 will destabilize and increase the pKa of C34. The observed pKa of C34 in the N state of HSA could be a combination of its burial as well as its interactions with adjacent acidic/basic amino acid residues. Such observations have been previously reported for other proteins [18–20,44]. For HSA, the pKa of C34 in the N state of the protein have been reported to range from 8.0–9.0 depending upon the methodologies and experimental conditions used (buffer, ionic strength, temperature *etc.*) [45–47]. The estimated pKa of C34 in the N state of HSA from our study, in its error range (8.7 ± 0.5), is in good agreement with the previously reported pKa values.

Another significant result from the present study is that the mean pKa of the C34 in the U state is ~1.3 unit less than the expected pKa of thiol in water (~8.2). We conjectured this could be due to the presence of some polar residues in the primary sequence of HSA near C34, thereby reducing its pKa. Assuming the U state to be completely unfolded and random coil-like structure, we checked for the adjacent amino acid residues of C34 in the sequence. The pentapeptide segment of HSA containing C34 in the middle has other residues as Q32, Q33, P35 and F36. These residues, being neutral in nature, are unlikely to affect the pKa of C34. The only other possibility is that there exists some residual structure in the U state in which some charged residues are interacting with C34, thereby, lowering its pKa. It is important to note that the m_{NU} value of a large protein like HSA is only ~1 kcal mol^{-1} M^{-1} (Fig. 4B). This result indicates that only a small amount of the buried surface area is exposed to solvent upon unfolding and there exist substantial residual structure in the U state. It appears that the residual structure brings one or more charged residue in the spatial proximity of C34 and hence decreasing its pKa. Multiple reports have demonstrated the presence of both favorable and unfavorable electrostatic interactions in the U states of proteins largely affecting their energetics [22–24,48–51].

The thermodynamic stability of proteins is a difference in the free energy of their N and the U states. Therefore, understanding the thermodynamic and structural properties of the U state is also essential in addition to that of the N state for decoding the factors that govern the stability of proteins. Generally, the free energy of the U state is considered zero and the difference between the free energies of the N and the U states is taken as the thermodynamic stability of the N state. However, our observations that the pH-dependent thermodynamic stability of HSA is controlled by the perturbed pKa of a single ionizable group in the N state and the U state and that in the U state, the ionizable residue participates in electrostatic interactions indicate that the electrostatics of the U state is also an important determinant of protein stability.

The findings of this study also point towards the significant contribution of interdomain coupling to the thermodynamic stability of HSA. In this study, the far-UV CD monitored unfolding experiments at pH 7 and pH 8 suggested that the global secondary structure of HSA changed differently as a function of urea concentration at two different pH conditions. The C_m of the far-UV monitored transition was observed to be in very well correspondence with that of the W214 fluorescence monitored transitions. It is important to note that W214 is placed at the interdomain region of the protein and reports about the site-specific structural change of that region. Excitingly, the observation that the abolition of the titration of C34 (which is placed in the core of domain I) is eliminating the pH dependence of the stability of HSA as monitored by W214 fluorescence (at the interdomain region) shows that the protein stability is coupled to the interdomain coupling. In one of our previous studies, we have observed that in the E form of HSA at low pH, domain III was fully unfolded while domain I and II were expanded but retained N-like secondary structures [33]. Interestingly, the thermodynamic stabilities of the protein in its E form as estimated from the changes in W214 fluorescence and far-UV CD were in very well agreement. Both these studies highlight the importance of interdomain coupling in protein stability.

4. Conclusion

In this study, we have explored the pH dependence of the thermodynamic stability of the human serum protein, HSA. We used fluorescence spectroscopy to understand the coupling of folding/unfolding of the protein with the ionization of the ionizable residues. We observed that the thermodynamic stability of HSA was pH dependent and changed in a sigmoidal fashion in the near neutral pH range (pH 6.0 – 9.0). In order to understand the cause of this dependency, we labeled a buried ionizable residue C34 with a dye, 1,5-IAEDANS and blocked its titration due to pH. The thermodynamic stability of the labeled protein became pH independent. These results conveyed that the pH dependent changes in the stability of HSA were due to the aberrant titration of the single thiol group. We calculated the pKa of the thiol in the N and U states. In the N state, the mean value of the calculated pKa was ~0.5 unit higher than that of the free cysteine in water, whereas, in the U state, there was a significant decrease in the pKa of C34 (by ~1.3 unit). The pKa in the N state could be attributed to the burial of the ionizable thiol group inside the hydrophobic core and its interactions with multiple nearby polar/ionizable residues. However, the perturbation of the pKa of the thiol in the U state indicated some local electrostatic interactions near C34 in the U state. These results together suggest that the electrostatic interactions regulate the ionization equilibrium of the ionizable residues and therefore affect their pKa values. The knowledge of pKa of these residues in both the N and the U states is essential in understanding the pH dependence of the thermodynamic stability of proteins.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgments

We thank Minnu M. Lal for helping in some of the equilibrium unfolding experiments. This work was funded by SERB-DST core research grant (project CRG/2019/002922) to S.K.J. P.M. is a recipient of the Senior Research Fellowship by University Grants Commission, India. D.P. is a recipient of the Senior Research Fellowship by the Council of Scientific and Industrial Research, India.

References

[1] C. Tanford, Protein denaturation. C. Theoretical models for the mechanism of denaturation, Adv. Protein Chem. 24 (1970) 1–95.
[2] J.B. Matthew, Electrostatic effects in proteins, Annu. Rev. Biophys. Biophys. Chem. 14 (1985) 387–417.
[3] N.M. Allewell, H. Oberoi, Electrostatic effects in protein folding, stability, and function, Methods Enzymol. 202 (1991) 3–19.
[4] C.N. Pace, Polar group burial contributes more to protein stability than nonpolar group burial, Biochemistry 40 (2001) 310–313.
[5] S. Kumar, R. Nussinov, Close-range electrostatic interactions in proteins, Chembiochem 3 (2002) 604–617.
[6] K. Takano, J.M. Scholtz, J.C. Sacchettini, C.N. Pace, The contribution of polar group burial to protein stability is strongly context-dependent, J. Biol. Chem. 278 (2003) 31790–31795.
[7] G.I. Makhatadze, V.V. Loladze, D.N. Ermolenko, X. Chen, S.T. Thomas, Contribution of surface salt bridges to protein stability: guidelines for protein engineering, J. Mol. Biol. 327 (2003) 1135–1148.
[8] G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee, M.H. Zehfus, Hydrophobicity of amino acid residues in globular proteins, Science 229 (1985) 834–838.
[9] S. Miller, J. Janin, A.M. Lesk, C. Chothia, Interior and surface of monomeric proteins, J. Mol. Biol. 196 (1987) 641–656.
[10] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[11] G.J. Lesser, G.D. Rose, Hydrophobicity of amino acid subgroups in proteins, Proteins 8 (1990) 6–13.
[12] L. Lins, A. Thomas, R. Brasseur, Analysis of accessible surface of residues in proteins, Protein Sci. 12 (2003) 1406–1417.
[13] S. Iwata, C. Ostermeier, B. Ludwig, H. Michel, Structure at 2.8 a resolution of cytochrome c oxidase from Paracoccus denitrificans, Nature 376 (1995) 660–669.
[14] H. Luecke, H.T. Richter, J.K. Lanyi, Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution, Science 280 (1998) 1934–1937.
[15] Y. Jiang, A. Lee, J. Chen, V. Ruta, M. Cadene, B.T. Chait, R. MacKinnon, X-ray structure of a voltage-dependent K+ channel, Nature 423 (2003) 33–41.
[16] C.N. Pace, G.R. Grimsley, J.M. Scholtz, Protein ionizable groups: pK values and their contribution to protein stability and solubility, J. Biol. Chem. 284 (2009) 13285–13289.
[17] M. Tollinger, K.A. Crowhurst, L.E. Kay, J.D. Forman-Kay, Site-specific contributions to the pH dependence of protein stability, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 4545–4550.
[18] D.G. Isom, C.A. Castaneda, B.R. Cannon, P.D. Velu, E.B. Garcia-Moreno, Charges in the hydrophobic interior of proteins, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 16096–16100.
[19] D.G. Isom, C.A. Castaneda, B.R. Cannon, B. Garcia-Moreno, Large shifts in pKa values of lysine residues buried inside a protein, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 5260–5265.
[20] N. Aghera, I. Dasgupta, J.B. Udgaonkar, A buried ionizable residue destabilizes the native state and the transition state in the folding of monellin, Biochemistry 51 (2012) 9058–9066.
[21] C.A. Fitch, S.T. Whitten, V.J. Hilser, E.B. Garcia-Moreno, Molecular mechanisms of pH-driven conformational transitions of proteins: insights from continuum electrostatics calculations of acid unfolding, Proteins 63 (2006) 113–126.
[22] S.T. Whitten, E.B. Garcia-Moreno, pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state, Biochemistry 39 (2000) 14292–14304.
[23] C.N. Pace, R.W. Alston, K.L. Shaw, Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, Protein Sci. 9 (2000) 1395–1398.
[24] S. Lindman, M.C. Bauer, M. Lund, C. Diehl, F.A. Mulder, M. Akke, S. Linse, pK(a) values for the unfolded state under native conditions explain the pH-dependent stability of PGB1, Biophys. J. 99 (2010) 3365–3373.
[25] A.S. Yang, B. Honig, On the pH dependence of protein stability, J. Mol. Biol. 231 (1993) 459–474.
[26] J. Makowska, K. Baginska, A. Liwo, L. Chmurzynski, H.A. Scheraga, Acidic-basic properties of three alanine-based peptides containing acidic and basic side chains: comparison between theory and experiment, Biopolymers 90 (2008) 724–732.
[27] C.N. Schutz, A. Warshel, What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 44 (2001) 400–417.
[28] D.L. Luisi, C.D. Snow, J.J. Lin, Z.S. Hendsch, B. Tidor, D.P. Raleigh, Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9, Biochemistry 42 (2003) 7050–7060.
[29] D.C. Carter, J.X. Ho, Structure of serum albumin, Adv. Protein Chem. 45 (1994) 153–203.
[30] T.U. Peters, All About Albumin: Biochemistry, Genetics, and Medical Applications, Academic Press, 1996.
[31] M. Dockal, D.C. Carter, F. Ruker, Conformational transitions of the three recombinant domains of human serum albumin depending on pH, J. Biol. Chem. 275 (2000) 3042–3050.
[32] P. Mishra, S.K. Jha, An alternatively packed dry molten globule-like intermediate in the native state ensemble of a multidomain protein, J. Phys. Chem. B 121 (2017) 9336–9347.
[33] N. Acharya, P. Mishra, S.K. Jha, Evidence for dry molten globule-like domains in the pH-induced equilibrium folding intermediate of a multidomain protein, J. Phys. Chem. Lett. 7 (2016) 173–179.
[34] J.R. Olivieri, A.F. Craievich, The subdomain structure of human serum albumin in solution under different pH conditions studied by small angle X-ray scattering, Eur. Biophys. J. 24 (1995) 77–84.
[35] N. Acharya, P. Mishra, S.K. Jha, A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein, Phys. Chem. Chem. Phys. 19 (2017) 30207–30216.
[36] M. Steglich, R. Lombide, I. López, Expression, Purification and Initial Characterization of Human Serum Albumin Domain I and Its Cysteine 34 15, 2020 (e0240580).
[37] L. Turell, R. Radi, B. Alvarez, The thiol pool in human plasma: the central contribution of albumin to redox processes, Free Radic. Biol. Med. 65 (2013) 244–253.
[38] L. Turell, S. Carballal, H. Botti, R. Radi, B. Alvarez, Oxidation of the albumin thiol to sulfenic acid and its implications in the intravascular compartment, Brazil. J. Med. Biol. Res. 42 (2009) 305–311.
[39] C.N. Pace, Determination and analysis of urea and guanidine hydrochloride denaturation curves, Methods Enzymol. 131 (1986) 266–280.
[40] L. Painter, M.M. Harding, P.J. Beeby, Synthesis and interaction with human serum albumin of the first 3,18-disubstituted derivative of bilirubin, J. Chem. Soc. Perk. T 1 (1998) 3041–3044.
[41] T.O. Street, N. Courtemanche, D. Barrick, Protein folding and stability using denaturants, Methods Cell Biol. 84 (2008) 295–325.
[42] R. Khurana, A.T. Hate, U. Nath, J.B. Udgaonkar, pH dependence of the stability of barstar to chemical and thermal denaturation, Protein Sci. 4 (1995) 1133–1144.
[43] P. Mishra, S.K. Jha, Slow motion protein dance visualized using red-edge excitation shift of a buried fluorophore, J. Phys. Chem. B 123 (2019) 1256–1264.
[44] A. Giletto, C.N. Pace, Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1, Biochemistry 38 (1999) 13379–13384.
[45] A. Bocedi, G. Cattani, L. Stella, R. Massoud, G. Ricci, Thiol disulfide exchange reactions in human serum albumin: the apparent paradox of the redox transitions of Cys(34), FEBS J. 285 (2018) 3225–3237.
[46] J. Bonanata, L. Turell, L. Antmann, G. Ferrer-Sueta, S. Botasini, E. Méndez, B. Alvarez, E.L. Coitiño, The thiol of human serum albumin: acidity, microenvironment and mechanistic insights on its oxidation to sulfenic acid, Free Radic. Biol. Med. 108 (2017) 952–962.
[47] O. Spiga, D. Summa, S. Cirri, A. Bernini, V. Venditti, M. De Chiara, R. Priora, S. Frosali, A. Margaritis, D. Di Giuseppe, P. Di Simplicio, N. Niccolai, A structurally driven analysis of thiol reactivity in mammalian albumins, Biopolymers 95 (2011) 278–285.
[48] M. Guzman-Casado, A. Parody-Morreale, S. Robic, S. Marqusee, J.M. Sanchez-Ruiz, Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of Thermus thermophilus ribonuclease H, J. Mol. Biol. 329 (2003) 731–743.
[49] J.H. Cho, S. Sato, D.P. Raleigh, Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, J. Mol. Biol. 338 (2004) 827–837.
[50] B.E. Bowler, Thermodynamics of protein denatured states, Mol. BioSyst. 3 (2007) 88–99.
[51] E. Arbely, T.J. Rutherford, H. Neuweiler, T.D. Sharpe, N. Ferguson, A.R. Fersht, Carboxyl pK(a) values and acid denaturation of BBL, J. Mol. Biol. 403 (2010) 313–327.