SOLUTION THERMODYNAMICS AND INTERACTION STUDIES OF PROTEIN-OSMOLYTE MIXTURES BY NMR

Haritha Haridas

Under the supervision of Prof. Federico Fogolari

and Prof. Gennaro Esposito

University of Udine

A thesis submitted for the degree of

PhD in Biomedical Sciences and Biotechnology

Academic Year 2010-2011

Abstract

Osmolytes are a class of intracellular organic molecules used by living organisms to regulate the cell volume under conditions of osmotic stress. Although the principal property of osmolytes is to function as osmoticants, they are known to play a role as chemical chaperones, protecting proteins against denaturation caused by adverse environmental conditions. The variable and usually high intracellular concentrations of these molecules has raised questions about the possible roles of these molecules in protein folding/unfolding. Studies in this direction have reported that osmolyte-induced protein folding is a very weak protein:osmolyte interaction and the important factor for measuring relative denaturing/stabilizing strength of osmolytes is the transfer Gibbs energy (ΔGtr) of amino acid side chains and backbone when transferred from water to osmolyte solutions. This PhD thesis focuses on the role of osmolytes in stabilizing/destabilizing proteins involved in amyloidoses, i.e. diseases developing as a consequence of protein misfolding.

We have chosen to study three amyloidogenic systems viz.; the eye-lens protein α-crystallin which upon aggregation and precipitation gives rise to cataract, a fragment from the Amyloidβ-peptide (Aβ) a major constituent of the Alzheimers plaque and β2-microglobulin (β2m) the protein responsible for dialysis related amyloidosis. Using NMR experiments we have tried to characterize the type of interactions of the first two systems with a well-studied protein-stabilizing osmolyte, carnosine. Next, we have looked at the unfolding free energy landscape of β2m in the presence of the protecting osmolyte carnosine and the destabilizing osmolyte heparan sulfate by using hydrogen-deuterium NMR exchange experiments. Our results from interaction studies suggests the presence of transient protein:osmolyte interactions. In the case of α-crystallin the interaction was very efficient and

could be traced up to very high concentrations of the osmolyte (a protein-osmolyte ratio of 1:1000). The results from isotope exchange experiments throw light on the global and local unfolding events occurring in the model protein β2m in the presence and absence of the two osmolytes.

To Aachan....

Acknowledgements

This page would fall short of space if I were to extend my acknowledgement to everyone who has helped me successfully completing my PhD. Therefore, I take this opportunity to specifically thank those without whom it would have been impossible for me to complete this thesis work.

First of all I would like to thank the Bio-Physis group at the University of Udine for accepting me and providing me a great experience in the past four years of my stay here. To make one NMR-ist out of a pure biologist, it takes one good teacher. Professor Gennaro Esposito has been this to me and I thank him specially for teaching me the basics of NMR. He has been extremely patient in teaching me all what he did in the time I was here. Every time I sat at the machine with him, I learnt something new and he made sure that I did! His ways at tackling problems have always amazed me and has helped me in becoming a better researcher. Rino, working with you has been great and a truly rewarding experience.

Professor Federico Fogolari, has been the motivation for me to consider doing my PhD in Bio-Physics. Being a computer master-brain, he taught me the basics of programming, operating with Linux, simulations, modeling and so goes the list. Federico, if it had not been for you I would have not known the magnificent world of molecular biophysics! Thank you!

Dr. Allessandra Corazza has been helpful in a variety of ways. I really appreciate the time she took to help me focus on my thesis work and for teaching me to work with tough programs like Mathematica that made no sense at all in the beginning. Sandra, thank you for your efforts, it is highly appreciated. Interacting with Professor Paulo Viglino has been great indeed. His visits to the lab always brought in pleasantness and his

talks and discussions with everyone in the lab has been very motivating and interesting.

Dr Kshitish ACharya has been my professor while in India and also a good source of advises and help during my stay here. I thank you for all what you taught me and I owe a lot to you for making a researcher out of me! Special thanks to you and your wife Mrs.Mangala Acharya for all your support.

Dr. Katiusia Pagano deserves special mention for being the first post-doctoral researcher I met here and for encouraging and helping me when she was here and for continuing to do it while in Milan. She has also been my first friend here and therefore holds a special place in everything that I did here. Katiu, I am really glad to have known you, thanks!

Dr. Maurizio Polano and Dr. Eva Andreuzzi has been the experts in protein expression and purification and have both taught me to work with proteins. Maurizio always had solutions, both in beakers and in his mind for any problem that I encountered while expressing and purifying the "tough" protein that I worked with. Thank you specially to you both. Eva, I also thank you for all the Friuli visits and invites that you offered always. It has all been very memorable.

Thank you Dr.Enrico Rennella and Dr.Luca Codutti for advises about NMR, encouragement shared over cups of coffee and organizing the important friday evening apperitivo. Getting to know you both have been great. Dr Rennella deserves special thanks for the Mathematica notebooks that he created that has been an integral part of my thesis work and his help in making me understand how NMR works. Dr.Chiara Mimmi has been helpful every time I wanted to use the Mass Spectrometer and a very interesting person to talk to over cups of coffees. Thank you! for all you help. I will also remember the travel to Lecce specially!

Thank you Cinzia and Matteo for bring my ex-batchmates. I would first thank you for your friendship. Thanks Matteo for being ever ready to help with any computational problem that I encountered in the days you were here and thank you Cinz for sharing your knowledge of Chemistry with me and for the many many visits to the down-town, coffees, shopping etc etc.

Thanks are due to the current and past students of the lab; Marco Vuano, Tommaso Banelli, Raffaella Picco, Devrim Gumral and others who have visited the lab in between. Interacting with you all have been very nice and you have all made my stay here worthwhile!

I would like to thank all my friends for being with me and would specially like to mention a few names here. Thulasi, Shereen and Juby, it would have been practically impossible to finish this PhD without you! Thank you. Thanks to my friend Alok for the encouragement during the writing of this thesis and otherwise. Special thanks to Melvin for all the motivation sent from across the miles, it worked wonders! Thanks to Nijil and Prasad for being my Indian family in Udine!

Lastly, my love and special thanks to my mom without whom I am nothing! He strength and support has been my light. Talks with her every day kept me going and Im glad I have her beside! Dad, you are fondly remembered and missed! I love you both.

Thank you all!!

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

1

Introduction

1.1 Osmolytes

Osmolytes are a class of organic compounds used by a large variety of living organisms to combat the harsh environmental stress set by factors such as high concentration of salts and organic solutes, drought, extreme temperature and hydrostatic pressure (1, 2, 3, 4). Traditionally osmolytes have been described in water-stressed organisms (5) (commonly found in oceans) and tissues (eg. mammalian kidney (6)) where it accumulates in response to hyperosmolality. More recently (1, 7), emphasis has been on the action of these molecules sometimes called as "chemical chaperones", on macromolecular stability and interactions in stressed and crowded intracellular environments (8, 9, 10).

It is often assumed that osmolytes have evolved to stabilize proteins against adverse environmental conditions. The evolutionary advantage of organic osmolyte systems over inorganic salts is its compatibility with macromolecular structure and function at high and variable concentrations without perturbing cellular processes (7). It has been shown that the protein backbone is effectively osmophobic and osmolytes can provide significant stability to protein by hiding the backbone into the core of folded proteins (11, 12, 13). This unfavorable interaction between the peptide backbone and osmolytes has been termed the osmophobic effect by Bolen and Baskakov (9). They identify this effect as a natural thermodynamic force that complements the well-recognized hydrophobic interactions, hydrogen bonding, electrostatic and dispersion forces that drive protein folding, described by Kauzmann 50years ago.

1.1.1 Osmolytes in protein folding

The proteins synthesized inside the cell are initially highly disordered. This unfolded or unstructured states pass through well-defined partially structured/folded transition/intermediate states, to finally form the native folded and functionally active conformation. Proteins are synthesized on the ribosome and assumed to fold in the endoplasmic reticulum in the presence of several molecular chaperones that deter aggregation of incompletely folded species (14). The correctly folded protein is secreted from the cell and functions normally in its extracellular environment. However, under certain conditions protein unfolds, at least partially, and becomes prone to aggregation. This can result in the formation of fibrils or amyloids and other aggregates that accumulate in tissue/cell. Small aggregates or highly organized fibrils and plaques may give rise to pathological conditions (Figure-Fig. 1.1).

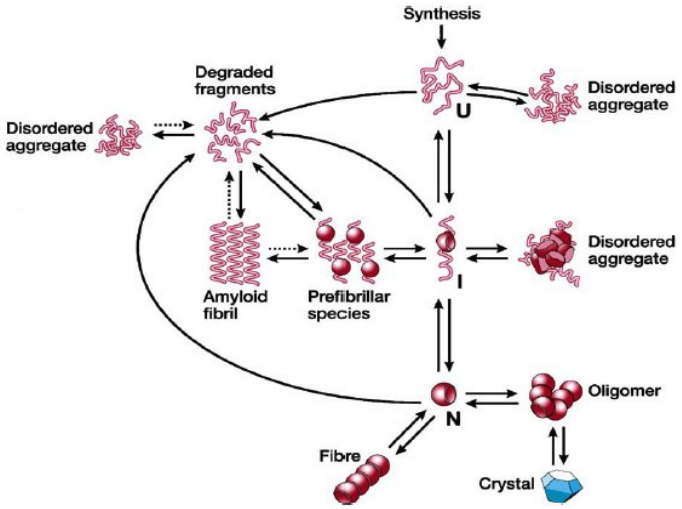


Fig. 1.1 Schematic diagram of some of the states accessible to a polypeptide chain following its biosynthesis. - The relative populations of the different states depend on the kinetics and thermodynamics of the various equilibria shown in the diagram. In living systems the fate of a given protein molecule is closely regulated by molecular chaperones and other quality control mechanisms, rather as metabolic pathways in cells are controlled by enzymes and associated molecules such as cofactors. Naturally occurring osmolytes could facilitate the intermediate reactions by shifting it to the final natively folded conformation, thereby making them less susceptible to aggregation. U, unfolded state; I, partially structured/folded intermediates; N, globular native form. Figure adapted from (15).

The cooperative nature of protein folding is considered to be crucial for avoiding protein/peptide aggregation. In the two-state model of protein folding (explained in section 1.1.2), the partially folded intermediate states are small. It has been proposed that osmolytes aid the cooperative folding of proteins (16). Several studies, have shown that in the presence of certain osmolytes, long stretches of unstructured or intrinsically disordered regions/domains of proteins can be correctly folded in a cooperative manner resulting in conformations that are native-like possessing significant functional activities (16, 17, 18, 19, 20). Reports of another study by Kumar et al (21) show that a relatively folded protein with an active globular structure can adopt a more functionally active conformation in the presence of osmolyte.

However, effects of osmolytes on the protein folding/stability are unique in nature and could vary with the site of action, the osmotic stress involved and availability of the substrates and osmolyes (2). There are many intracellular organic osmolytes in cells, ranging from protecting (eg. N- trimethylamine oxide (TMAO), sucrose, proline) to denaturing (eg. Urea) in action, but effects of differing osmolytes are additive and specific to cells/tissues (22, 23). For example, the high levels of urea in kidney is combated by high concentrations of protecting-osmolytes like betaine, sorbitol, glycerophosphocholine, inositol and taurine. The availability of number of osmolytes in a particular cell may decide the concentration of osmolyte mixtures required to fold protein present in it into a functionally active conformation. In short, specific cells utilize specific osmolyte(s) to encourage optimal functioning conformation(s) of certain proteins (Figure-Fig. 1.2).

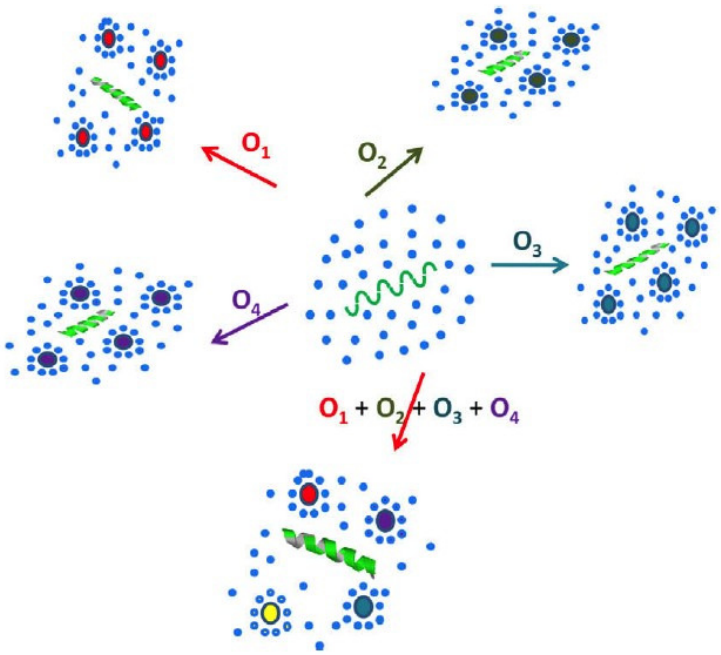


Fig. 1.2 A model showing osmolyte-induced characteristics in thermodynamic terms of solvophobic effects. - Osmolytes interact unfavorably with the unfolded or unstructured states, resulting in preferential depletion of osmolyte proximate to the protein surface, thereby forcing protein to adopt folded conformation(s). Protein is shown with green color. In the center, protein is intrinsically disordered or unstructured, and is surrounded by only water molecules (small blue filled circle). O1, O2, O3, or O4 (shown by large filled circle, red, green, blue, and purple, respectively) represent the presence of various osmolytes (Ox). In the presence of each osmolyte, due to sequestering of water molecules by osmolytes, protein adopts folded conformation, which may have similar or different overall fold depending upon the nature of osmolyte and protein. The final natively folded conformation of protein may be determined by the sum of these osmolytes. Effects of these osmolytes on the folding of protein may be additive in nature. Thus, the availability of number of osmolytes in a particular cell may decide the concentration of osmolytes required to fold protein into a functionally active conformation.The final outcome may also be determined by the nature of other interacting macromolecules and small ligands (if any). Figure adapted from (10).

1.1.2 Thermodynamics of osmolyte aided protein folding

The two-state model of protein folding (N$\longleftrightarrow$U), is not an ordinary chemical reaction as no covalent bonds are made or broken in the process of interconversion between natively folded (N) and unfolded (U) states. Protein denaturation/renaturation is rather a reequilibration process between the two independent species N and U in solution under different solvent conditions. Therefore a thermodynamic process of protein folding can be described in terms of solvent interactions with the unfolded and native states. Protecting osmolytes push the equilibrium toward N, by raising the free energy of the unfolded state thereby favoring the folded population, whereas denaturing osmolytes push the equilibrium toward U, by lowering the free energy of the unfolded state hence favoring the unfolded population (24).

Based on the transfer Gibbs energy Δgtr, measurements of amino acid side chains and backbone from water to osmolyte solutions it has been proposed that the ability of the osmolytes to stabilize the proteins against denaturing stress originates from the unfavourable interaction of the osmolyte with the peptide backbone (11, 12, 13, 24). Since unfavourable interactions between solvent component and protein functional group was traditionally called the solvophobic effect, the interaction in the case of osmolytes was termed the osmophobic effect (9).

The backbone Δgtr value is the important factor for measuring the relative denaturating/stabilizing strength of different osmolytes (24). The backbone unit experiences interactions with the solvent water (reference state). When transferred from water to an aqueous osmolyte solution these interactions differ due to the presence of molecules that interact with the backbone thus raising (protecting osmolyte) or lowering (denaturing osmolyte) the Δgtr value relative to the reference state. Considering the well defined nature of the two solvent system, the resultant Δgtr arises from differences between backbone/water and backbone/water/osmolyte interactions which in-turn is responsible for reequilibrating the native and unfolded species in solution in the absence/presence of the osmolyte (Figure-Fig. 1.3).

Timothy et al (24) have demonstrated that the transfer free energy of protein backbone from water to a water/osmolyte solution, Δgtr, is negatively correlated with an osmolyte's fractional polar surface area. Backbone/solvent interaction energy is a function of interactant polarity, and the number of energetically equivalent ways of realizing

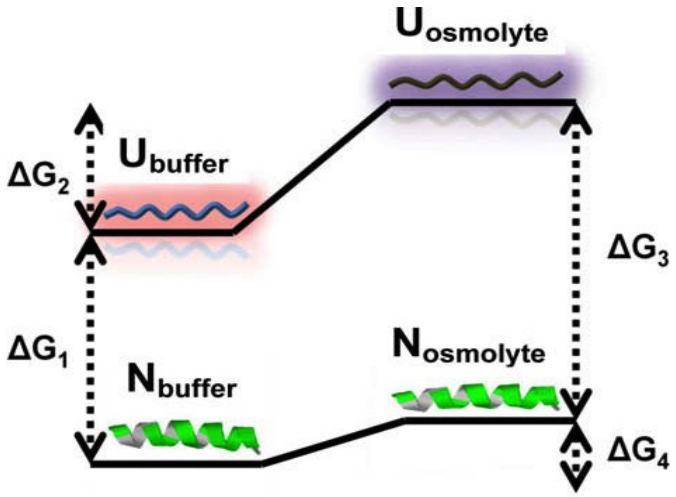


Fig. 1.3 A schematic diagram showing protein stabilization brought about by the action of osmolyte. - Osmolytes act by increasing DG between the natively folded (Nbuffer) and unfolded ensembles. DG1 is the unfolding Gibbs energy difference between native and unfolded (Ubuffer) protein in aqueous buffer solution, and DG3 is the Gibbs energy change for the same reaction in the presence of an osmolyte. Transfer of Nbuffer or Ubuffer from aqueous buffer to osmolyte solution (Nosmolyte and Uosmolyte, respectively) raises the DG2 of the unfolded ensemble much more than it does for the natively folded ensemble (DG4), resulting in a greater stability of the protein in the osmolyte solution than in aqueous solution. Figure adapted from (10).

a given interaction is a function of interactant surface area. Accordingly at the molecular level, stabilizing osmolytes are preferentially excluded from backbone polar groups, whereas denaturing osmolytes are preferentially accumulated there. The reason for this preferential interactions is that in the presence of stabilizing osmolytes like TMAO (having three large apolar groups), water is more likely to interact with the backbone polar groups than the osmolyte itself but in the case of polar molecules like urea (denaturing osmolyte) the situation reverses, as the osmolyte has more propensity to interact with the backbone than water.

It is worth stressing here that osmolyte-induced protein folding is a very weak protein:osmolyte interaction, and as described above do not occupy any significant fraction of the backbone surface (9, 25). This essentially allows no competition for

backbone binding sites (25). This also explains how proteins remain unaffected by the observed high intracellular osmolyte concentrations.

1.1.3 Types of organic osmolytes

As discussed above, based on their intracellular role organic osmolytes can be broadly classified as protecting and denaturing . Examples for protecting osmolytes are TMAO, betaine, sucrose, trehalose, sarcosine, sorbitol, proline, and glycerol and for denaturants are urea and guanidine.

Naturally occurring organic osmolytes can be grouped into few major chemical categories (1, 5, 22) viz; small carbohydrates including sugars (e.g. trehalose), polyols (glycerol, inositols, sorbitol, etc.) and derivatives (such as o-methyl-inositol); amino acids (glycine, proline, taurine, etc.) and derivatives (e.g. ectoine); methylamines (such as N-trimethylamine oxide (TMAO) and betaine (N-trimethyl glycine), glycerophosphorylcholine) and methylsulfonium solutes including dimethylsulfonopropionate (DMSP); and urea. Except urea (a naturally occurring metabolite used by some animals) these compounds are widespread in occurrence across plant and animal kingdoms .

In simple terms, osmolytes are solvent additives that mainly act as stabilizing cosolutes for proteins by affecting the solvent properties of water as related to protein polarity and protein diffusion. Two such osmolytes used in this work viz., Carnosine (an endogenous dipeptide) and Heparan Sulfate (a natural glycosaminoglycan) both of which have been shown to interact and regulate the activities of several proteins are discussed below.

1.1.3.1 Carnosine

Carnosine (β-alanyl-L-histidine) (Figure-Fig. 1.4) is a naturally-occurring, histidine-containing dipeptide present in human brain, kidneys, innervated tissues and eye-lens at concentrations as high as 20mM (26). It acts as a protective agent in such long-lived tissues and is considered to be of potential biochemical and therapeutic significance. Carnosine has been studied in the realm of its antioxidant activities and biological buffering capacity, preventing damage to many biological structures and protecting cells against osmotic shock, oxidative stress or toxic effects of metabolites and free radicals. Carnosine has also been proposed to function as a neurotransmitter, immunomodulator, antiradiation agent, anti-senescence agent, metal ion chelator, wound healing agent, and

antiglycating agent inhibiting sugar mediated protein cross-linking (27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41). As is evident from literature, this molecule is being vastly studied in the realm of its biological activities as a co-solvent and its therapeutic potential is being examined with scrutiny.

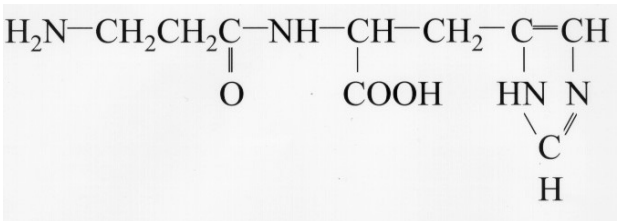


Fig. 1.4 Chemical structure of carnosine - β-alanyl-L-histidine

1.1.3.2 Heparan sulfate

Both Heparin and Heparan sulfate (HS) are members of a family of polyanionic, polydisperse, linear polysaccharides called glycosaminoglycans (GAGs), which perform a variety of crucial biological functions and have been extensively employed as therapeutic agents (42). Heparin and HS-GAGs are structurally related and synthesized in the Golgi by the same biosynthetic enzymes and differ only in the composition of N and O-sulfo groups (Heparan having higher density than HS) (43). They are made up of repeating uronic acidglucosamine disaccharides with as many as 32 different disaccharide building blocks comprising $>10^{14}$ possible sequences for an icosasaccharide chain with a great structural diversity (Figure-Fig. 1.5). Heparan sulfate proteoglycans are expressed by virtually all mammalian cells, and depending on the core protein, they may either be associated with the cell surface or deposited in the extracellular matrix (44). HS interact with a wide array of proteins, modulating their biological activities but the specificity or affinity of these interactions is still an area of debate.

A study (45) using heparin as a cosolvent reported that this polyanion can prevent creatine kinase aggregation and another study by Park et al (46) reported that heparin partially blocked FAS aggregation when added at a low concentration and it stimulated aggregation at higher concentrations. Other studies (47, 48, 49, 50) have shown that HS is a universal component of amyloid and is perhaps playing an important role in

amyloid generation by promoting fibrillogenesis. A recent study by Relini et al (49) showed that heparin promotes β-2 Microglobulin aggregation in solution at pH 6.4.

Given the structural diversity, ubiquitous presence and the multitude of possible heparan/heparan sulfate-protein interactions in biology it is justifiable to think that it plays an important role in the protein folding/unfolding process in a variety of ways.

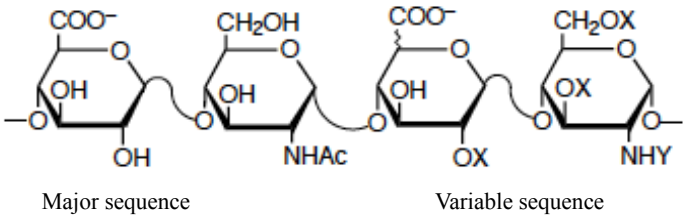


Fig. 1.5 Major and minor disaccharide repeating units in heparan sulfate - X=H or SO_3^-, Y=Ac, SO_3^-, or H

1.2 Protein misfolding and amyloid forming proteins

1.2.1 Protein folding energy landscape- from folding to "misfolding"

Protein folding is the mechanism by which a polypeptide chain attains its unique functional three-dimensional structure. Although theoretically a single protein can exist in an infinite number of ways based on the spatial rearrangement of its chemical groups and functionality, in physiological conditions majority of proteins populate the conformation with the highest functionality, termed the "native" state.

The cooperative nature of protein folding in which the locally folded parts of the protein fold together to stabilize the overall native structure is represented by the funnel shaped energy landscape model (14, 51) (Figure-Fig. 1.6). Proteins are described as following several routes within this funnel to reach their native state conformation. The local roughness of the funnel reflects transient trapping of the protein configurations in local free energy minima. Random fluctuations in the unfolded or partially folded states drive this reaction as different native as well as non-native contacts are sampled. Native interactions between residues are assumed to be more stable than non-native contacts, and as such contacts form, the number of available conformations are reduced,

driving the polypeptide chain towards the native structure. The native conformation thus formed may undergo further structural changes by virtue of its biological functions.

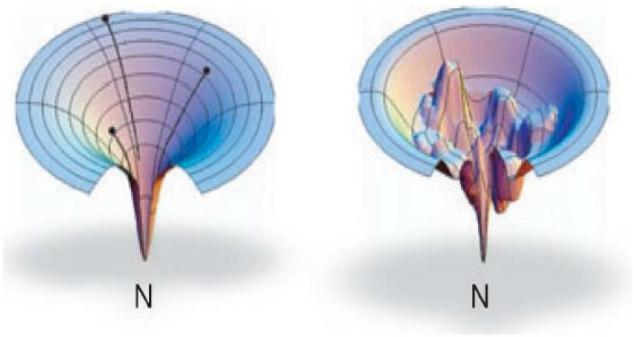


Fig. 1.6 Energy landscape schematic diagrams. - Left, a smooth energy landscape. Right, a rugged energy landscape. "N" indicates the native conformation taken from (52).

In general, small single domain proteins up to 100 amino acids in length fold to their native state by a smooth, two-state mechanism (separated by a single transition barrier between the ensemble of unfolded structures and native state) (53, 54) on a sub-second timescale (Figure-Fig. 1.6) and proteins larger than ∼100 residues in length fold on a much rougher energy surface in which several folding intermediates are populated on its way to the native state. This is because larger protein chains have a higher tendency to collapse in an aqueous solvent, resulting in the formation of compact states that may contain substantial elements of native-like structure containing both native and non-native contacts. Reorganization of these inter-residue contacts may involve a high free-energy barrier, leading to the transient population of partially folded or 'intermediate states. Such species can be productive for folding (on-pathway), or trapped such that the native structure cannot be reached without substantial reorganizational events (the intermediate is off-pathway).

For both small and large proteins, the surface of the folding funnel is unique for a specific polypeptide sequence under a particular set of conditions and is determined by both thermodynamic and kinetic properties of the folding polypeptide chain. Partially folded states on this landscape may be intrinsically prone to aggregation and favorable intermolecular contacts may lead to their association and ultimately to protein misfolding. The minimum energy reached by this "misfolded state" is similar to that reached

by the native protein and the assembly of misfolded monomers may attain a very high stability. (55) (Figure-Fig. 1.7).

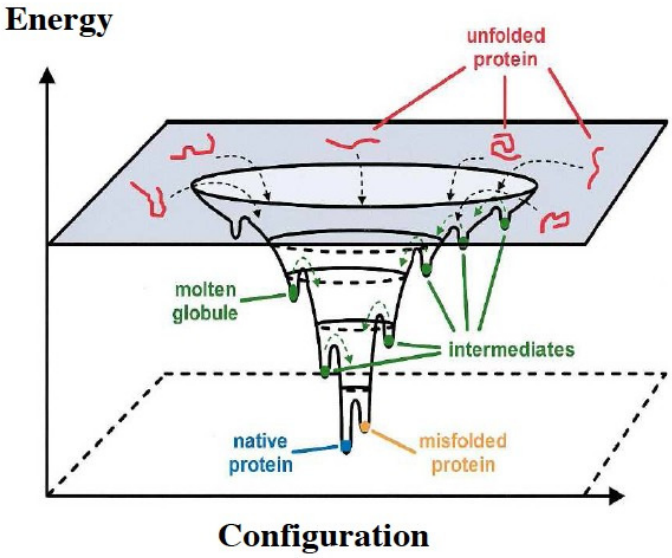


Fig. 1.7 Schematic representation of the different states involved in protein folding displayed as a function of the configuration of the atoms in the energy landscape funnel - The multiple states of the unfolded protein located at the top fall into a folding funnel consisting of a very large number of local minima, each of which describes possible folding arrangements in the protein. Most of these states represent transient folding intermediates in the process of attaining the correct native fold. Some of these intermediates retain a more stable structure such as the molten globule, whereas other local minima act as folding traps irreversibly capturing the protein in a misfolded state. Figure adapted from (55).

Inside the living system this relatively stable state, especially if oligomeric is sometimes very robust and not dissolved by proteolysis (normal physiologic response to misfolded proteins), and hence aggregates to give rise to several pathologies collectively termed amyloidoses. Amyloidoses have been traditionally defined as diseases in which normally soluble proteins accumulate in the extracellular space of various tissues as insoluble deposits of fibrils that are rich in β-sheet structure and have characteristic dye-binding properties (56). About 30 diseases are known today which comply to the typical characteristics of amyloidoses (57).

1.2.2 Amyloidoses and amyloid proteins

Amyloidoses are the largest group of protein misfolding diseases characterized by the conversion of peptides or proteins from their soluble functional states into highly organized fibrillar aggregates showing a cross-beta super-secondary structure termed "amyloid" (Figure-Fig. 1.8) The amyloidoses differ in their protein precursors undergoing aggregation, the target organs involved in amyloid deposition and, consequently, in their clinical features (58). To date, at least 28 different proteins have been identified as causative agents of amyloid diseases (59) affecting either the central nervous system or a variety of peripheral tissues. These conditions include Alzheimers, Parkinsons and the prion diseases, several forms of fatal systemic amyloidosis, and at least one condition associated with medical intervention (haemodialysis) which makes it important in the context of human health and welfare (60).

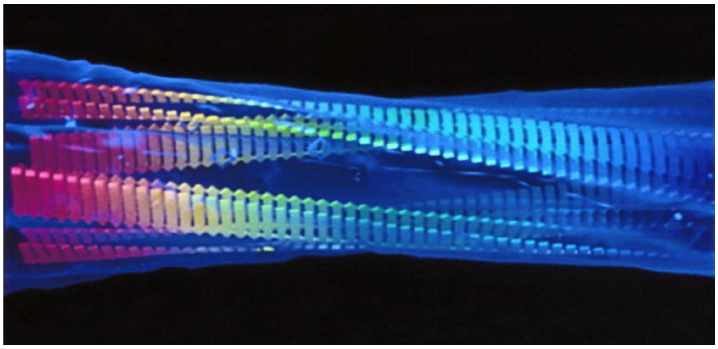


Fig. 1.8 A molecular model of an amyloid fibril. - The fibril consists of four protofilaments that twist around one another to form a hollow tube with a diameter of approximately 6 nm. The model illustrates one way in which regions of the polypeptide chain involved in β-sheet formation could be assembled within the fibrils taken from (14).

This thesis work revolves around three proteins exhibiting pathological aggegation viz. the eye-lens protein α-crystallin which upon aggregation and precipitation gives rise to cataract, the Amyloidβ-peptide a major constituent of the Alzheimer's amyloid plaques and β2-microglobulin the protein responsible for dialysis related amyloidosis. As is the case with other amyloid forming proteins, despite the lack of common features in terms of sequence, structure, function or site of action all of them form fibrillar aggregates with the typical cross-β structure in which continuous β-sheets are formed with β-strands running perpendicular to the long axis of the fibril (61, 62). These three

proteins in the context of amyloidosis and its pathological manifestations are described in the following subsections.

1.2.2.1 α-Crystallin and the cataractous eye lens

The normal functionality of the eye lens is maintained by a set of proteins called crystallins. The α, β and γ crystallins together maintain the stability and transparency of the eye lens. These proteins are laid down in the foetal lens and thereafter the turnover rate is almost zero (63). There is also very little metabolic activity in the lens fibre cells and as a result, crystallin proteins at the centre (nucleus) of the lens are as old as the individual. Clearly the lens crystallins are highly stable proteins and are organized in a supramolecular β-sheet structure(64). The liquid-like short range order present in highly concentrated solutions of the crystallin proteins in the lens is thought to be responsible for its transparency. This, together with the absence of cellular organelles makes it a medium with high refractive index and ensures the proper focusing of light onto the retina.

α-Crystallin is the major structural eye-lens protein. It is a member of the small heat-shock (sHsp) family of intracellular molecular chaperone proteins which interact with partially folded proteins to prevent their mutual association and possible precipitation (65, 66, 67). As a result of this ability, in addition to its structural role in the lens, α-crystallin prevents the aggregation of other lens crystallins (including itself) and acts with various proteins against aggregation induced by heating, chaotropic agents, reduction, and chemical modification (64, 68, 69). In its native form α-crystallin is composed of two closely related subunits, namely αA and αB, each being $\sim$20 kDa, but occurs as a highly heterogenous oligomer with a mass range from 300 kDa to 1000 kDa. In the human lens, the αA to αB ratio is 3:1, although this ratio varies with age. Apart from the lens, the αB subunit is expressed extensively in many other tissues, including the brain, lung, and cardiac and skeletal muscle (70) where it functions solely as a molecular chaperone.

Although the crystallin proteins are highly durable and resistant to aggregation, there is a gradual effect of a combination of destabilizing factors with age (extensive post-translational modification like truncation, phosphorylation, oxidation, deamidation and racemization as well as changes in cellular conditions) that results in eventual

aggregation and precipitation of crystallin proteins out of solution(71, 72). This ultimately leads to the development of age-related cataract, a condition defined as opacity of the lens, which is a leading cause of vision impairment worldwide and the reason behind ∼40-80% of the estimated 45 million cases of blindness that occur across human race (73, 74).

There are two possible mechanisms described in literature for cataractogenesis (75). First one is a condensation phenomenon whereby opacity results from loss of solubility of the crystallins and the second mechanism describes cataract as a conformational disorder resulting from unfolding/destabilization of the crystallin proteins. There have been studies supporting both these mechanisms but of relavance to this thesis is the second hypothesis which highlights the amyloid forming propensity of the crystallin proteins. In one of the studies on murine cataract (76), the investigators identified the presence of intranuclear filamentous inclusions in vivo that stain with the amyloid-detecting dye Congo Red and a ready assembly of the mutant γB^{nop}-crystallin into amyloid fibrils in vitro under conditions where the wild-type protein remained soluble. This made them conclude that amyloid deposition by a γB^{nop}-crystallin was perhaps the cause of the cataract. Later, Meehan (75) and coworkers demonstrated that all three classes of wild-type bovine crystallin proteins, including the molecular chaperone α-crystallin, are capable of forming amyloid fibrils when subject to unfolding conditions in vitro. They explain the amyloid conversion of crystallins resulting from the strain of longevity under destabilizing conditions over many years, along with the diminishing chaperone ability of α-crystallin in old age as the possible cause of cataract.

Despite the widespread nature of this disease, there are no drops or medications to prevent or reverse cataracts. The only treatment available at present is surgical removal of the cloudy lens and replacement with a plastic intraocular lens (73). At a time when cataract induced blindness keeps increasing in poorer regions of the world and the waiting list for cataract surgery in public hospitals gets longer around the globe, a drug developed to treat or reverse cataract could come as a boon to mankind in protecting his sense of sight. A better understanding of the intracellular environment of the eye lens and the molecular mechanism of cataract formation could be directed towards finding a potential molecular drug target endogenous to the eye yet effective in halting/reversing this age-related protein aggregation disease of the eye-lens.

1.2.2.2 β-Amyloid peptide and the Alzheimer's brain

Amyloid beta (Aβ or Abeta) is a peptide normally produced as a soluble, non-toxic metabolic product of the Amyloid precursor protein (APP) (77, 78). It is generated by cleavage of APP at two locations by proteases denoted as the β- and γ-secretases. The initial proteolysis by the β-secretase results in a residual C-terminal fragment containing the transmembrane and cytoplasmic domains of APP, which undergoes an additional, intramembranous cleavage by a γ-secretase to release Aβ. The γ-secretase cleavage site is heterogeneous and produces Aβ fragments that can vary in length from 39-42 residues (79, 80). However, the two predominant species are peptides 1-40 (Aβ40) and 1-42 (Aβ42). Although Aβ is most commonly known in association with Alzheimer's disease (AD), its non-disease functions include activation of kinase enzymes (81), protection against oxidative stress(82, 83), regulation of cholesterol transport (84, 85), functioning as a transcription factor (86, 87) and anti-microbial activity(88). A structural transition of the polypeptide chain from the native to β aggregate conformation is a signature of disease.

In Alzhemeir's disease Aβ aggregates and accumulates in the brain as diffuse and compact plaques (89). Diffuse plaques are not associated with degenerative changes, whereas compact plaques composed of insoluble βA fibrils are associated with pathological changes in the surrounding brain parenchyma, a crucial step that ultimately leads to Alzhemeir's disease as explained by the amyloid cascade hypothesis (90, 91, 92, 93). The hypothesis that underwent modifications over the past twenty years currently suggests that the synaptotoxicity and neurotoxicity of Alzhemeir's may be mediated by intermediary soluble toxic oligomers of multimeric amyloid-β peptide species. The dynamic nature of these species and the poorly defined mechanism (or mechanisms) of toxicity are challenges facing the pharmaceutical industry. In fact, several therapeutics that were purported to reduce amyloid-β production or aggregation have failed in Phase III clinical testing, making the amyloid cascade hypothesis a topic of debate and controversies.

Amidst the trials and controversies surrounding the pharmaceutical research for Alzhemeir's, it continues to be the leading cause of dementia worldwide. The global prevalence of dementia is estimated to be as high as 24 million, and is predicted to double every 20 years through to 2040 (94). Alzheimer's is predicted to affect 1 in

85 people globally by 2050, which would be a costly burden of disease(95). Most often, AD is diagnosed in people over 65 years of age (96), although the less-prevalent early-onset Alzheimer's can occur much earlier. The disease is characterized by a progressive decline in cognitive function, which typically begins with deterioration in memory and eventually ends in death. In the years following diagnosis the patient develops behavioral and psychiatric problems and gradually loses bodily functions. Before death, individuals have usually become dependent on caregivers. On average, the life expectancy following diagnosis is approximately seven years (97)

With the ageing of society (considered as an important social problem) and AD being a degenerative disease affecting the elderly population, finding a cure for AD is an urgent research priority. Interventions that could delay or prevent the onset of AD would have an enormous positive public health impact. Any successful therapeutic strategy would help in reducing individual and societal burden and financial costs caused by this disease. Current therapies for AD provide only mild, transient symptomatic relief. An apparent unmet need exists for improved drugs, which could be based on novel molecular targets that modify the underlying course of the disease. Also, screening any naturally occurring metabolic product found in normal brain tissue for its role in regulating amyloid-β production, facilitating its clearance or preventing its misfolding could be worthwhile in addressing the etiology of this devastating disease and go towards designing a natural therapeutic molecule.

1.2.2.3 β-2 Microglobulin and the related osteoarthropathies

β2-microglobulin (β2m) is a minor plasma protein found associated with the major histocompatibility complex class I (MHC-I) heavy chain. (98). It is a 11.7 kDa protein, necessary for the cell surface expression of MHC class I complex and the stability of the peptide binding groove of the complex. β2m is secreted from the plasma membranes as a result of the continuous regeneration of membrane proteins in the cell surface of all nucleated cells. The concentration of free β2m in the serum is regulated by the equilibrium between its dissociation from the MHC-I complex and its clearance via glomerular filtration. Therefore, the reduced efficiency of the glomerular filtration in renal failures leads to an increase up to 30 or 60 times of the levels of circulating β2m in the serum (99).

Dialysis Related Amyloidosis (DRA) is a complication of critical kidney disease, in which circulating β2m (which is inefficiently removed by hemodialysis or continuous ambulatory peritoneal dialysis) deposits as amyloid plaques in osteoarticular sites, synovial membranes and in some visceral organs. The complication occurs gradually, especially in older adults subjected to dialysis for more than five years. Deposited amyloid plaques induce various destructive osteoarthropathies such as carpal tunnel syndrome, atherosclerosis, flexor tenosinovitis, subchondral bone cysts and erosions as well as pathological fractures (99, 100). Visceral involvement (101) has been reported in organs like heart, gastrointestinal tract and tongue with rare manifestations.

The overall incidence and prevalence of DRA are not clear but reports say that one-third of patients are affected after less than four yeas of undergoing hemodialysis and 90% after more than seven years (102). Risk factors include the time of dialysis, the type of dialysis membrane used and the age of the patient at the onset of dialysis. There is no known cure, however in patients who are able to undergo renal transplantation disease progress is halted but regression is unlikely. Apart from the disabling musculoskeletal complications there have been rare cases of massive gastrointestinal bleeding due to submucosal bowel deposits and reports of severe pulmonary hypertension and heart failure due to β-2-microglobulin amyloid deposits in the interstitium and/or vasculature of the cardiovascular system.

DRA is the third example of amyloidosis addressed in this thesis. Although the prevalence and incidence of this disease is lesser than the two mentioned in the previous subsections, it is a major concern for the sect of population suffering from renal failure and thereby undergoing dialysis. As is the case with the other two, we see that DRA also has a tendency to affect the elderly. This makes us speculate if there is some endogenous component(s) aiding protein folding in normal conditions, whose concentrations reduce with age thereby causing certain proteins to misfold, aggregate and give rise to pathological manifestations. A good guess would be intracellular osmolytes, both protective and denaturing, the physiological balance between them in different tissues and their relative concentrations with age. The hypothesis is very vague and needs to be thoroughly tested before coming to any conclusions, but cannot completely be ruled out. It is worth recalling here that the endogenous dipeptide carnosine, is seen in very high concentrations in the eyes, brain and kidneys and that its levels in these tissues decrease with age.

1.3 Possible roles of osmolytes in amyloidosis

Despite the lack of identifiable common features in the sequences and structures of the proteins implicated in the different amyloid diseases, the fibrillar aggregates associated with them are all characterized by a high degree of structural order and high intermolecular β-sheet content (103, 104, 105, 106). It has therefore been postulated that there is a common underlying mechanism of formation of the various fibrillar aggregates (15). Moreover, there is an increasing number of evidence in support of the hypothesis that the ability to form amyloid structures with ordered intermolecular β-sheets is an inherent property of polypeptide chains and therefore all proteins, at least in principle, have the potential to form such structures under appropriate conditions (15, 105, 107, 108, 109).

This implication has profound consequences for understanding the fundamental origins of the deposition of proteins in diseases (60). It points out to a generalized mechanism of protein misfolding and regulation processes that efficiently maintain the dynamic equilibrium between the various states adopted by polypeptides upon their biosynthesis (including aggregates and their precursors) and conditions that lead to build-up of certain precursor species nucleating their rapid aggregation. Aggregates could result from impairment of specific cellular functions such as the ability to clear misfolded proteins or protein aggregates, or from changes in the intracellular environment such as a decrease in ambient pH or in the antioxidant defences or an increase in the macromolecular crowding. There have been studies showing that the increase in macromolecular crowding as a result of ageing (110) or of progression through the cell cycle (111) may perturb the folding and binding of proteins favouring more compact states of macromolecules, increasing the population of aggregated and precursor species. In this regard it has been shown that every type of cell is equipped with mechanisms designed to maintain or to restore cellular volume, water content, and/or turgor pressure in response to any changes in the composition of the extracellular fluid (112). Unfolded protein response, heat shock proteins, molecular chaperones and osmolytes are all part of the effective quality control mechanisms adapted by cells to prevent aggregation or accumulation of aggregated proteins thereby helping the smooth functioning of the system.

Schein (113) has reviewed the effects of osmolytes on proteins and on the solvent properties of water as related to protein polarity and protein diffusion. Osmolytes as solvent additives can favorably affect protein stability and solubility by changing the viscosity and surface tension of water, reducing the exposure of hydrophobic surfaces thus facilitating the formation of protein hydrophobic collapse, changing the isoelectric points of proteins and affecting the susceptibility to proteases. Results of the studies (45, 114, 115) conducted on the effect of osmolytes in proteins refolding together suggest that although the protective effect of osmolytes on protein activity and structure is equal for different osmolytes, the ability of osmolytes to facilitate the refolding of various proteins differs from case to case. A quick review of literature on the implications of carnosine and HS in amyloidosis is discussed below.

1.3.1 Carnosine

Carnosine is a versatile molecule studied by researchers all around the world for its various biological properties as an antioxidant, immunomodulator, and neuroprotector against free radicals (35, 38, 116). Although there are several studies showing the therapeutic effect of carnosine in treating gastric ulcer and burns, lowering blood pressure and protecting the brain from ischemia/reperfusion injury (117) etc., much interest has been on establishing this molecule as an anti-senescense drug by exploiting its free radical scavenging activity (118). Aging is intimately associated with oxygen free-radical-induced damage to cells, which leads to instability and malfunction of cells and consequently to the occurrence of aging and senile diseases. Diseases like atherosclerosis, diabetes, Alzheimer's disease and senile cataract are all related to free radical induced damage. There have been studies addressing the roles of carnosine as a therapeutic agent against such senile diseases.

We have already seen that the fibrillar form of Aβ is a defining feature of Alzheimer's disease (AD). Aβ is known to increase oxidative stress in endothelial cells and smooth muscle cells of cerebral blood vessels, which accumulate the peptide during AD (119). It inhibits endothelial cell replication and is directly toxic to both peripheral and cerebral vascular endothelium. In the brain, Aβ could result in impairment of the blood-brain barrier. Preston et al. (35) studied the protection by carnosine from toxic effect of truncated form of Aβ on immortalized rat brain vascular endothelial cells (RBE4). Using a mitochondria dehydrogenase activity reduction assay, lactate dehydrogenase release,

and glucose consumption, they found that addition of 20 mM carnosine immediately before Aβ treatment significantly protected the cells from toxic effects of 200 and 300 μg/ml of Aβ. They postulate that the mechanism of carnosine protection lies in its antiglycating and antioxidant activities, both of which are implicated in neuronal and endothelial cell damage during AD.

Wang et al (118) investigated the curative effect of carnosine on senile cataract by administrating eye drops containing 20mM carnosine to 96 patients aged 60 years having senile cataract of various degrees of maturity, with the duration of the disease from 2 to 21 years. Based on the level of eyesight improvement and the change of lens transparency they reported that carnosine gave a pronounced effect on primary senile cataract with the effective rate being 100%. For mature senile cataract, the effective rate reported was 80% and positive effects were also observed with other types of cataract. They concluded that carnosine has the potential of being a drug to prevent and treat senile cataract and to delay eyesight senescence. At a molecular level, Attanasio and coworkers (120) have shown invitro that both L and D enantiomers of carnosine are capable of inhibiting α-crystallin fibrillation and of restoration of the chaperone activity of α-crystallin lost after denaturant and/or heat stress.They also found a disassembling effect of carnosine on α-crystallin amyloid fibrils and demonstrated the ability of carnosine to restore most of the lens transparency in organ-cultured rat lenses exposed to similar denaturing conditions that were used for in vitro experiments.

A study by Kang et al (121) stated that carnosine can efficiently inhibit the oligomerziation of both wild type and mutant alpha-synuclein. α-synuclein is a major component of Lewy bodies that are found in the brains of patients with Parkinson's disease (PD). Another recent study (122) reported that carnosine inhibited the β-sheet formation and neurotoxicity of synthetic prion protein PrP106-126.

1.3.2 Heparan Sulfate

Heparan sulfate proteoglycans (HSPG) are ubiquitous macromolecules associated with the cell surface and extracellular matrix (ECM) of mammalian tissues. HS interacts with a multitude of proteins and plays important roles not only in development and homeostasis, but also in pathological processes (123, 124) thereby making it an important molecule for pharmaceutical research. Studies have shown that HS directly interacts with amyloid proteins in vitro and investigations of in vivo amyloid deposits

have shown that HS is a co-deposit in all types of amyloids (125). Although there is a lot of evidence suggesting an important role for HS in amyloidosis its exact functions in the pathogenesis of amyloidosis is not clear.

It has been proposed that HS may promote formation of the nidus and/or protofilament around which amyloid fibrillogenesis takes place and impart stability to the amyloid fibril in vivo (126). This hypothesis was supported by the marked increase in fibrillogenesis in vitro exerted by HS or its analogue heparin on various amyloidogenic polypeptides including Aβ (127, 128), tau (129, 130), PrP (131), and β2m (48, 49). Also, fibril formation of β2m was found to involve GAGS. Rellini et al (49) have pointed out that Heparin has a strong enhancing effect on β2m fibril formation in the presence of collagen. Another study demonstrated that HS is an important cell surface factor mediating uptake of Aβ140, and the uptake event coincides with Aβ cytotoxicity (132)

Based on the studies so far, there is a proposition that HS may function as a scaffold to facilitate amyloid deposition (133) Understanding the molecular mechanism of HS-amyloid interaction in vivo will help us to get a better idea about the onset and progress of amyloid deposition. HS and HSPG may be functionally engaged in amyloid deposition through common mechanisms, but they can also exert distinct functions in other aspects of amyloid pathology, for example, cell toxicity (134).

2

NMR as a Tool to Study Protein-Osmolyte Interactions

Apart from its widely recognized role in the structural characterization of biological macromolecules, Nuclear Magnetic Resonance (NMR) spectroscopy is a unique tool to study molecular interactions in solution. Today, bio-molecular NMR has re-invented itself as an essential technique to characterize events of molecular recognition by providing us quantitative information on molecular flexibility and dynamics. Ever since the disclosure of the "Structure Activity Relationship (SAR) by NMR" (135) method in literature, by the group at Abbott Laboratories in 1996, NMR techniques have been applied in a systematic way to screen for target binding ligands in pharmaceutical research.

Any small molecule in solution reversibly interacting with the binding domain of a macromolecule shows notable differences in its relaxation behavior and molecular mobility characteristics, that are exploited in NMR-based techniques used to identify and characterize protein-ligand interactions. These techniques (136, 137, 138, 139, 140, 141, 142, 143, 144) involve chemical shift perturbation, differential line broadening, transferred NOE, and pulsed field gradient diffusion measurements in solution mixtures. The macromolecule studied can be a protein or nucleic acid and the ligand of interest may be an organic molecule, a peptide, a carbohydrate, or a combination of these. The measurements are usually performed in aqueous (mostly buffered) solutions that are fairly close to those found in biological systems which makes it easy to add ligands and observe the effects that these ligands have on the target macromolecule.

2. NMR AS A TOOL TO STUDY PROTEIN-OSMOLYTE INTERACTIONS

NMR-based interaction studies can be can be roughly subdivided into two groups, namely; those in which the focus is on the macromolecule spectrum and the changes in chemical shift by ligand titration is followed, and those in which the NMR signals of the usually low molecular weight-ligand and its behavior upon binding to the target is observed.

The first approach that uses techniques involving protein (macromolecule) detection, exploit the site-specific information afforded by assigned protein NMR spectra along with a priori knowledge of its 3D structure (either from X-ray diffraction or NMR). For this, the use of large amounts of labeled protein and the access to high magnetic fields is required. Chemical shift alterations due to ligand interaction can be easily followed by 2D ^{1}H correlation experiments or Heteronuclear Single Quantum Coherence (HSQC) experiments with 15N-labeled or/and 13C-labeled proteins and by identifying binding-induced perturbations of assigned protein resonances, ligands can be identified and their binding sites can be localized. However in this case, NMR is limited by the size of the protein, need for over-expressing and purifying milligram (mg) quantities of soluble, non-aggregating, protein which should be stable throughout the battery of multi resonance experiments followed by the time consuming process of resonance assignments. For these reasons, target-based detection methods are normally used for proteins with solved structures and the less time-consuming ligand-based detection approach is generally used for novel protein-ligand binding characterization studies.

The major attractions of the ligand-based approaches (145, 146, 147) in addition to lesser experimental time are that spectra can be easily recorded by low magnetic field spectrometers using samples with small amounts of unlabeled protein and prior knowledge of the resonance assignment of the protein is not needed. Since ligand-based methods can compare the NMR parameters (variations in peak intensities, relaxation behavior, diffusion or saturation transfer) of a single or a mixture of compounds in the presence and absence of the receptor molecules, the size of the protein receptor becomes irrelevant. In fact the techniques are more sensitive when dealing with larger molecular weight proteins.

Four such methods that I have used to study the interaction behaviour of the osmolyte carnosine with α-crystallin and Aβ-peptide, namely Saturation Transfer Difference (STD), Diffusion Ordered Spectroscopy (DOSY), Nuclear Overhauser effect

spectroscopy (NOESY) and Total correlation spectroscopy (TOCSY) are described in the following sections.

2.1 Saturation transfer difference (STD) spectroscopy

STD (148, 149, 150, 151, 152) is a key method to identify and characterize ligand-receptor interactions and is based on the transfer of saturation from the protein to the bound ligand. Macromolecules like proteins have a large system of protons tightly coupled by dipole dipole interactions. The longitudinal relaxation rate of protons in proteins is dominated by cross relaxation rate (σintra). As the name suggests, STD takes the difference of two experiments (Figure-Fig. 2.1). In a first experiment ('on-resonance' experiment), one selectively saturates the various receptor proton magnetizations via a train of selective radio-frequency (rf) pulses. The rf train (usually a cascade of soft Gaussian pulses) is applied to a frequency window that contains receptor resonances, but lacks resonances from the ligands (e.g. 0.0 to -1.0 ppm for proteins). The saturation propagates from the point of application to other receptor protons through spin-diffusion via the vast network of intramolecular 1H-1H cross-relaxation pathways. This is quite efficient due to the larger molecular weight of the receptor. If a ligand resides in the binding site of the protein, saturation may be transferred progressively to this ligand (σinter), by intermolecular 1H-1H cross relaxation at the ligand-receptor interface. Protons which are close in space to the protein surface (e.g. proton Ha in Figure-Fig. 2.1) receive higher degree of saturation, while protons that are not in intimate contact with the protein (e.g. protons Hb and Hc in Figure-Fig. 2.1) receive little or no saturation. They then dissociate back into free solution where the saturated state persists due to the small free state R1 relaxation values. At the same time, more 'fresh' ligand exchanges on and off the receptor while saturation energy continues to enter the system through the sustained application of rf. Thus, saturated free ligands accumulate during the saturation time.

A complementary reference experiment ('off-resonance' experiment) is then recorded that applies the identical rf train far off-resonance; thereby not perturbing any NMR resonances. The 'on-resonance' and 'off-resonance' experiments are recorded in an interleaved fashion and subtracted. The resulting difference spectrum yields only those

resonances that have experienced saturation which includes precisely the protein receptor and the binding compound resonances. The receptor resonances are not visible because of their minimal concentration and because of R2 relaxation-filtering (T1ρ) just prior to detection. The result is a simple 1D 1H spectrum that reveals only the binding compounds. In order to improve the spectrum the solvent signal is also suppressed.

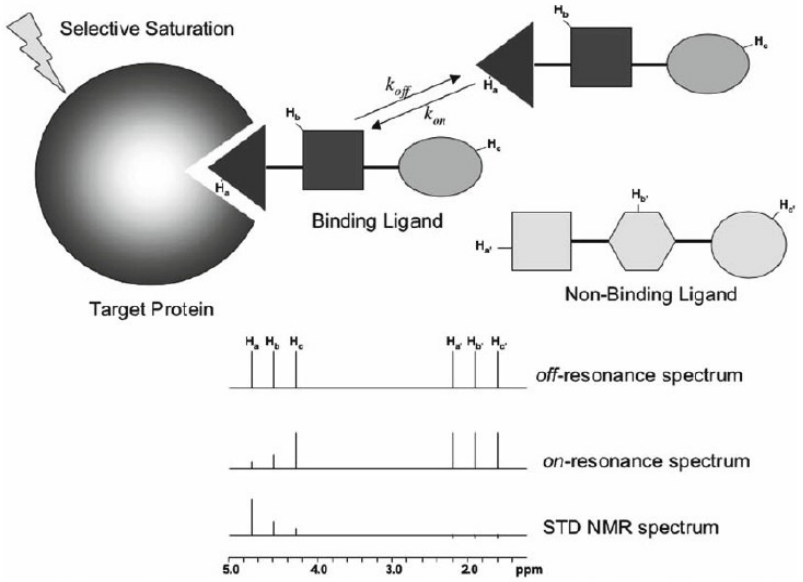


Fig. 2.1 Scheme showing principle of the STD NMR method. - Target protein is saturated with a cascade of selective pulses. Saturation is transferred quickly through the protein mediated by spin diffusion. A ligand binding that is in fast exchange with the target protein receives saturation from the protein. Ligand protons that are in close contact with the protein (Ha) receive more saturation than ligand protons that are solvent exposed (Hc) indicated by the colour scheme of the moiety. Protons from a non-binding ligand (Ha' , Hb' and Hc') do not receive any saturation from the protein and are therefore not attenuated in the on-resonance spectrum. Subtraction of the on -resonance and off -resonance spectrum reveals the final difference spectrum (STD) showing only signals from binding ligand protons that are in close proximity with the protein surface. Protons that are exposed to solvent (Hc) show none or weak STD NMR intensity and protons from non-binding ligands (Ha', Hb' and Hc') do not show any STD NMR signals.Figure adapted from (151).

The efficiency of ligand saturation largely depends on several factors, such as the disassociation rate constant, concentration ratio of ligand and target, saturation duration, irradiation offset and protein properties (like molecular tumbling rate) and the

rate of saturation depends on protein mobility, ligand/protein complex lifetime, and geometry.

An example of a pulse sequence of an STD experiment is given below in Figure-Fig. 2.2.

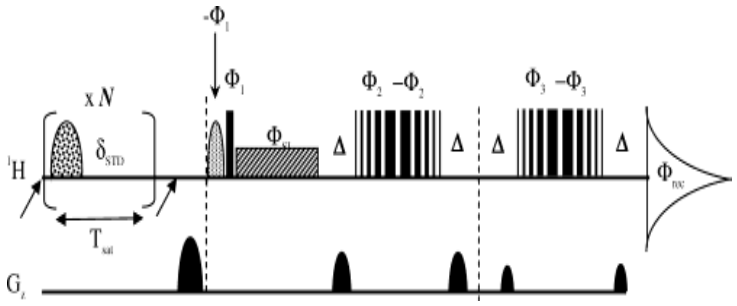


Fig. 2.2 An example of an STD pulse sequence - The upper and lower staffs show proton rf and field gradient pulses, respectively. Selective rf saturation occurs for a time T_{SAT} = 1-3s by a train of N frequency selective rf pulses (e.g. 50 ms Gaussian or 50ms Eburp 90° pulses separated by an interpulse delay of 1 ms). Two experiments are performed, which apply saturation on and off resonance with receptor protons in an interleaved fashion. The 1D signals are stored in separate locations, and their spectral differences are taken via inverting the receiver phase. A WATERGATE-5 readout sequence polls the residual z magnetization. Suppression of receptor signals is achieved by relaxation filtering during the spin-echo delay Δ and the optional $R_1\rho = 1/T_1\rho$ spin-lock. An optional water flipback 90° pulse precedes the first hard 90° pulse (e.g. 2 ms Seduce-1 90° pulse). Phase cycling is as follows: $\Phi_1 = (16x, 16{-}x)$;$\Phi_2 = (x, y, -x, -y)$;$\Phi_3 = (4x, 4y, 4{-}x, 4{-}y)$;$\Phi_{SL} = y; and \Phi_{rec} = 2(x, -x, x, -x, -x, x, -x, x), 2(-x, x, -x, x, x, -x, x, -x).\Phi_{rec} flips 180°$ between the on and off resonance spectra. Figure adapted from (153).

2.2 Diffusion ordered spectroscopy (DOSY)

Molecules in a solution exhibit random translational motion (brownian motion) by virtue of their internal kinetic energy. Self-diffusion of a molecule depends on its size and shape in solution and is altered upon binding to a large receptor. This makes the determination of self-diffusion coefficients (154) an important methodology for molecular interaction studies in solution. Pulsed field gradient (PFG) NMR diffusion experi-

ments (155) is one such powerful technique used to measure the translational diffusion of both small molecules and macromolecules and for identifying, quantifying, and mapping ligand-protein interactions in drug discovery . It is a non-invasive, cheap and easy method that can obtain physical, chemical and qualitative information about the components in the sample mixture.

The basic principle (156, 157) of diffusion edited NMR experiments is that the appication of externally controlled magnetic field gradients G to a sample in an NMR probe in the same direction as the dominant magnetic field B_0, will induce a change in the magnetization M as the molecules diffuse to different regions of the gradient. The gradients spatially encode the position of each nuclear spin. This spatial distribution can be decoded after waiting a short time by applying a second gradient. If the waiting time is very short between the encoding and decoding gradients the spins will not have had a chance to change position and in turn the magnetization will refocus without a net phase change. However the spins that have moved during the waiting time between the encoding and decoding steps will acquire a net phase change. The summation of the accumulated phases over the entire sample will lead to partial cancellation of the observable magnetization and hence an attenuation of the NMR signal. This signal attenuation is dependent on the strength of the applied gradients, the time between the gradients (diffusion time) and the diffusion coefficient of the molecules under study. In a typical Diffusion-ordered spectroscopy (DOSY) NMR experiment the signal decays exponentially leading to two dimensions; one dimension accounts for conventional chemical shift and the other for diffusion behaviour.

Since the diffusion behaviour of a molecule in solution depends on various properties such as size, shape, mass and charge as well as its surrounding environment, including solution temperature and aggregation state, it is possible to seperate each component in a mixture solely based on its diffusion coefficient in the diffusion dimension of the two-dimensional DOSY experiment.

We know that the diffusion coefficient is related to the size of a molecule by the Stokes-Einstein equation as,

D= $k_B T/f$ where k_B is the Boltzman constant, T is the temperature and f is the friction coefficient. For the simple case of a spherical particle of a hydrodynamic radius R_h in a solvent of viscosity η, the friction factor is given by $f = 6\ \pi\ \eta\ R_h$.

A pulse sequence of a simple PFG NMR experiment is given below in Figure-Fig. 2.3.

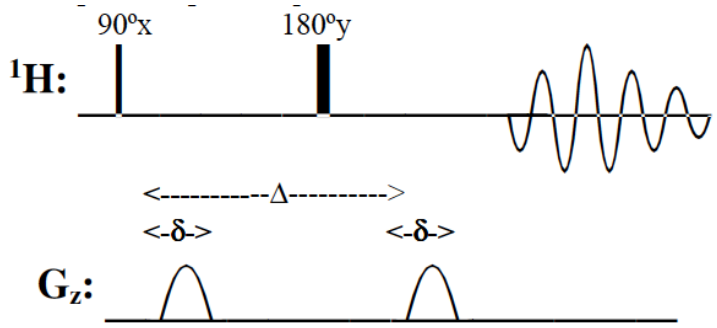


Fig. 2.3 The Stejskal and Tanner PFG spin echo pulse sequence. - On ^1H *the thin black solid line denotes a* 90° pulse and the thick solid black line denotes a 180°. G_z gradients for coding and decoding are denoted by a sine bell. Figure taken from Bruker Manual: Measuring Diffusion by NMR

In this experiment a gradient is applied right after the first 90 degree pulse has been executed. The spins experience a change in the phase angle proportionate to the amplitude of the applied gradient. After a time $\Delta/2$ a 180 degree pulse is applied which refocuses chemical shifts. The molecules are allowed to diffuse for a period of Δ. The second gradient then allows the magnetization to be refocused and the FID is subsequently recorded. Since the spins are allowed to move during the diffusion time, Δ, the net total phase change after the second gradient summed over the entire sample will not be zero. This leads to an attenuation in the observed signal. In order to measure the diffusion coefficients a series of experiments with varying echo time, Δ,=are recorded. The same effect can be obtained by varying the gradient amplitude while keeping the duration of the gradient constant. The overall length of each experiment will remain constant. The signal intensity will decrease with increasing gradient strength. Since the echo time for a series of experiments is the same the affects of T2 relaxation should remain constant within a given series of experiments.

The decrease in signal intensity as a function of gradient strength can be described by the general equation,

$I(q) = I_o e^{-Dq^2\Delta\prime}$

where D is the diffusion coefficient

$\Delta\prime = \Delta \ \ \delta/3$ where

$\Delta$is the diffusion time and δ is a correction factor for finite gradients

$q = \gamma g \delta$

γ is the gyromagnetic ratio

g is the amplitude of the applied gradient

δ is the duration of the applied gradient

A plot of I(q) versus $q^2\Delta\prime$ yields an exponential decay curve which is fitted to obtain the diffusion coefficient.

2.3 Nuclear overhauser effect spectroscopy (NOESY)

NOESY (158) is perhaps the most important tool in small protein NMR. It is a 2D experiment to probe intramolecular distances and is therefore used widely in structural NMR. The Nuclear Overhauser Effect (NOE) arises through the rf saturation or pertubation of one spin that causes perturbation of further nuclear spins through dipolar interactions (due to spatial proximity) between nuclei via cross relaxation. This results in a time-dependent change in the intensities of these spins. The relaxation effect strongly vary depending on the distance between the nuclei as the NOE $\propto r^{-6}$, where r is the distance between two spin systems and NOE is the observable signal pertubation intensity. In other words, the observed 2D peak volume is related to the distance between two nuclei, i.e. the lower the intensity, the larger is the distance. The sensitivity of the experiment can be used to probe distances of around 5 - 6 $\mathring{A}$ or shorter as the dipole-dipole interaction rapidly decreases with increased distance.

The maximum possible NOE depends on the molecular correlation time (or the inverse of the rate of molecular tumbling), which is in large part determined by the molecular weight and solvent viscosity. Larger molecular weights and higher viscosities lead to larger correlation times.The NOE is positive for small molecules if $\omega\tau_c \gg 1$ (molecular weight<600 for B_0 11.7 T), goes through zero (theoretically) for medium-sized molecules if $\omega\tau_c = 1.12$ (MW range 700-1200 for B_0 11.7 T) and becomes negative for large molecules $\omega\tau_c \ll 1$ (MW>1200 for B_0 11.7 T). Therefore, large molecules generally give better NOEs at higher field and small molecules may actually give NOEs at lower field. A 2-D NOESY of a small molecule will have cross peaks of opposite phase to the diagonal and that of a large molecule will have cross peaks of the same phase as the diagonal. Theoretically these experiments should be symmetrical but it is common to see more intense peaks on one side of the diagonal than the other due to departures from equilibrium and acquisition assymmetries.

A typical pulse sequence of a NOESY (Figure-Fig. 2.4) consists of three $\pi/2$ pulses. The first pulse creates transverse spin magnetization. This precesses during the evolution time t_1, which is incremented during the course of the 2D experiment. The second pulse produces longitudinal magnetization equal to the transverse magnetization component orthogonal to the pulse direction. Thus, the basic idea is to produce an initial situation for the mixing period τ_m (the time during which cross relaxation occurs) where the longitudinal polarization of each spin is labelled by its resonance frequency. The longitudinal magnetization is allowed to cross-relax during the mixing time τ_m meaning, transferring part of labeled longitudinal magnetization to another nucleus. For the basic NOESY experiment, τ_m is kept constant throughout the 2D experiment. The third pulse creates transverse magnetization from the remaining longitudinal magnetization, thereby enabling an additional and different frequency labeling of the transferred magnetization extent. Acquisition begins immediately following the third pulse, and the transverse magnetization is observed as a function of the time t_2. The NOESY spectrum is generated by a 2D Fourier transform with respect to t_1 and t_2.

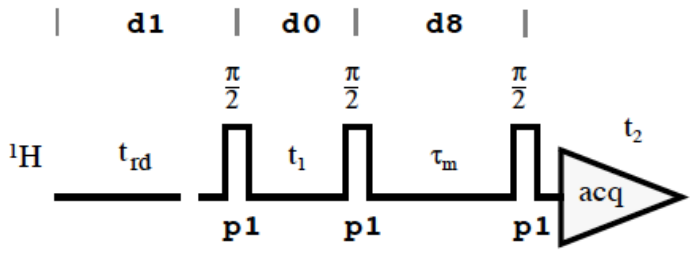


Fig. 2.4 NOESY pulse sequence - D0 is the incremented delay (t1) and D1 is the relaxation delay (should be about $1.25*T_1(^1H)$). D8 determines the length of the mixing period during which NOE buildup occurs. This should be on the order of T1. Figure taken from Bruker Manual: NOESY

A well known approach for the quantitative determination of distance between two nuclei using the NOE peak intensity is by using the following equation;

$r_{ij}=r_{ref}(a_{ref}/a_{ij})^{1/6}$; where

a_{ij} is the NOE cross peak volume and

r_{ij} is the interproton distance of the two protons i and j.

Given r_{ref} (known distance between two protons) and its NOE volume a_{ref}, r_{ij} can be calculated.

2.4 Total correlation spectroscopy (TOCSY)

The key to obtain useful chemical or structural information from multi dimensional NMR spectroscopy is the transfer of coherence from one spin to another during the mixing period of an NMR experiment. Coherence transfer can be produced by evolution of the spin system under a series of rf pulses and free precession delays (COSY-type coherence transfer) or by cross-polarization of the spin system by using continuous, time-varying rf fields (TOCSY-type or homonuclear Hartmann-Hahn (HOHAHA)-type coherence transfer). Both COSY and TOCSY mixing generate through-bond coherence transfer, but COSY identifies spins which are coupled to each other while TOCSY identifies nuclei that are connected by a continuous chain of couplings. This makes TOCSY useful for identifying the larger interconnected networks of spin couplings. As long as successive protons are coupled, magnetization is transferred successively over up to 5 or 6 bonds in a TOCSY experiment. Transfer is interrupted substantially by small or totally by zero proton-proton couplings.

In TOCSY (159), oscillatory exchange is established which proceeds through the entire coupling network so that there can be net magnetization transfer from one spin to another even without direct coupling. The isotropic mixing which occurs during the spinlock period of the TOCSY sequence exchanges all in-phase as well as antiphase coherences. A variety of schemes can be used to achieve mixing, but the most effective and therefore now universally used methods involve composite pulse spin locking. The coherence transfer period of the TOCSY sequence occurs during this multiplepulse spin-lock period. The duration of the mixing period depends on two factors, the size of the coupling constants among the various spins in the system, and the number of transfer steps required to see the desired correlations. Thus the length of the spin-lock period determines how far the spin coupling network will be probed. A general rule of thumb is that 1/(10 JHH) should be allowed for each transfer step, and five transfer steps are typically desired from the TOCSY spectrum.

The most commonly used multiple-pulse spin-lock TOCSY pulse sequence MLEV-17 (160) is given below in Figure-Fig. 2.5.

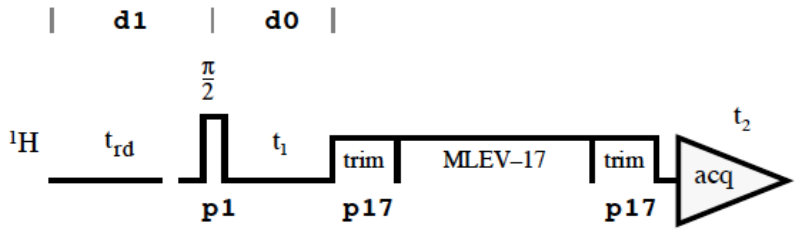


Fig. 2.5 TOCSY pulse sequence - D0 is the incremented delay (t1) and D1 is the relaxation delay (should be about 1.25*T_1(^{1}H)). The parameter p17 determines the length of the trim pulses at the beginning and end of the mixing period. The trim pulses are used to ensure that the final 2D spectrum can be phased easily. MLEV spinlock sequence determines the length of the TOCSY mixing period. Figure taken from Bruker Manual: TOCSY

2.5 Materials and Methods

2.5.1 Sample Preparation for NMR experiments

Bovine α-crystallin was purchased from Sigma and purified using a using Amicon Ultra Centrifugal filter device (3000MWCO) when needed. All samples were prepared in 20 mM aqueous phosphate buffer (pH 7.2) with 7% D2O for the NMR spectrometer frequency lock and 0.1% NaN3 to prevent microbial contamination. The protein concentration, assessed by UV absorption at 280 nm, ranged from 0.1 to 0.5 mM. A 1 M stock of L-carnosine (Fluka) in 20 mM aqueous phosphate buffer (pH 7.2) was prepared in water and was diluted to the required concentration for the different STD titration experiments. Aliquots of α-crystallin solutions (mentioned above) were prepared and 1 M guanidine-HCl added alone or in the presence of different concentrations of L-carnosine and used as such or after incubation at 60°C for 24 hours. The thermal treatment generates fibrillar assembly of α-crystallin (75). A slight solution opacity occurred after thermal treatment, independent of the presence or absence of carnosine. Samples of hen egg white lysozyme and bovine βL-crystallin (both purchased from Sigma), respectively 0.11 mM and 0.05 mM, were prepared in the same buffer conditions as for α-crystallin. All NMR experiments (on the α-crystallin-carnosine system) were carried out at 25°C.

Aβ12-28 was purchased from Anaspec. According to a previously reported protocol

(161) the lyophilized Aβ12-28 fragment was dissolved first in water/d6-DMSO 80/20 (v/v), at concentrations between 0.8 and 1.0 mM, with 0.03% (w/v) NaN3, and the pH was adjusted to 3.85 by means of very small additions (μL) of 1 M HCl. Carnosine, from concentrated stock solutions (10-100 mM) at pH adjusted to 3.7-3.8 (with HCl), was mixed through additions of few μL to Aβ12-28 solution, to reach a two-to- fourfold stoichiometric excess. To control better the pH upon carnosine addition, preparations with 20% d6-DMSO were made also using 20 or 50 mM d5-glycine buffer, at pH 3.80 or 3.30, respectively, for the Aβ fragment solution and the carnosine stock solution. In order to compare the output of NMR experiments with the other experimental results, additional experiments were carried out at neutral pH without d6-DMSO. To avoid Aβ12-28 aggregation and precipitation, the peptide was first dissolved at basic conditions, i.e. 2.3 mM in 20 mM NaOH at pH ¿ 10. Then a fivefold dilution in phosphate buffer led to final concentrations of peptide, phosphate and NaOH of 0.46, 16 and 4mM, respectively, in H2O:D2O 92:8, at pH 7.36. Carnosine additions were performed with 5-10 μL aliquots from a 50 mM stock solution in 50 mM phosphate (pH 7.42), to reach stoichiometric ratios carnosine: Aβ12-28 of 1:1 and 2:1, with final pH of 7.48. All operations were performed in ice bath and, when not in use for spectra acquisition, solutions were stored in the fridge at 4°C. NMR measurements with water/d6- DMSO samples were performed at 21°C, whereas for neutral aqueous phosphate samples the temperature was kept initially at 6.7°C and subsequently increased stepwise within limited intervals, first 6.7-12°C and then 25-30°C.

All chemicals were of the highest available grade and were used without further purification unless otherwise stated above.

2.5.2 NMR spectroscopy

All NMR experiments were carried out at 11.7 T with a Bruker Avance-DRX 500-MHz spectrometer equipped with triple-axis magnetic field gradients.

2.5.2.1 STD NMR

The pulse program for Saturation Transfer Difference (STD) experiments employs a shaped pulse train for saturation alternating between on and off resonance, a spoil module to destroy unwanted magnetization, a spin-lock module to suppress protein signals

and water suppression using excitation sculpting (148, 151, 152). The on-resonance irradiation of the protein was performed at a chemical shift of 0.814 ppm, far removed from any carnosine resonances, and the off-resonance irradiation was applied at 60 ppm where no protein and ligand signals were present. The STD spectrum was obtained by subtraction of the on-resonance and off-resonance time domain data acquired in an interleaved mode to minimize artifacts arising from homogeneity and temperature drifts. Selective presaturation of the protein signal was achieved by a train of 90°E-Burp pulses (162) of 50 ms length covering a frequency band of about 20 Hz. The total length of the saturating train of pulses was 1 s. Spectra were collected with a relaxation delay of 10 s to allow the establishment of an equilibrium state. A 20 ms spin- lock pulse with a $\gamma B2/2\pi$ of 10 kHz was used to substantually attenuate the background protein resonances thereby facilitating analysis. The spectra were processed using TOPSPIN software. An exponential line broadening function of 14 Hz was applied to the difference spectrum prior to Fourier transformation. The STD intensity (I_{STD}) of $H^{\epsilon 1}$ aromatic resonance of carnosine histidine in the generated difference spectrum was measured relative to the same resonance intensity in the off-resonance spectrum to estimate the saturation transfer enhacement (η_{STD}).

$\eta_{STD} = I_{STD}/I_0 = (I\text{-}I_0)/I_0$;

which in turn defines the STD amplification factor A_{STD} () as;

$A_{STD} = \varepsilon x \eta_{STD}$ where ε= [L- carnosine]$_{TOT}$ / [α-crystallin]$_{TOT}$

Since the η_{STD} was calculated as a self normalized ratio within each pair of on-resonance and off-resonance spectra, it was possible to instantly compare spectra acquired with a different number of scans and different concentrations of protein and/or ligand without any ambiguity. Qualitatively similar results were obtained also by considering the saturated protein resonance in the unperturbed spectra as reference (I_0) with an integral always calibrated to unity.

2.5.2.2 DOSY

Diffusion Ordered SpectroscopY (DOSY) (155) experiments for the measurement of the diffusion coefficients were run on the same solutions of α-crystallin and L-carnosine used for the STD experiments, and in addition to blank solutions, i.e. L- carnosine alone and α-crystallin alone, following or not a 24-hours thermal incubation at 60°C in the presence of 1 M guanidine-HCl. A sequence including bipolar gradients and

longitudinal eddy current suppression, in a double stimulated echo module delay for cancellation of convection artifacts was employed (163) to collect 80 points with linear increments from 2 to 95% of the z-axis magnetic field gradient (65 G/cm), with Δ (big delta) = 100-150 ms, δ (little delta) = 2-6 ms and 64-128 scans/increment was used. All data were processed using the Bruker TOPSPIN software. Solvent suppression was obtained by WATERGATE (164) applied in the excitation sculpting mode (165).

The translational diffusion coefficient measured by NMR DOSY is:

D= k_BT/f where k_B is the Boltzman constant, T is the temperature and f is the friction coefficient. For the simple case of a spherical particle of a hydrodynamic radius R_h in a solvent of viscosity η, the friction factor is given by $f = 6\ \pi\ \eta\ R_h$.

2.5.2.3 NOESY and TOCSY

2D TOCSY (159) and NOESY (158) experiments were collected typically over 2048x512 (t2xt1) data-point matrices for spectral widths of 69kHz in either dimensions with 64 scans/ t1-point, 64 dummy scans and relaxation delay of 1s. The TOCSY MLEV17 (160) isotropic mixing intervals were 80 ms long at $\gamma B2/2\pi$ = 8.3 kHz, whereas for NOESY, mixing times of 250 ms were employed. Quadrature in the indirect dimension was obtained by time proportional phase incrementation (166) and solvent suppression was implemented by appending to the sequences a WATERGATE spin-echo module performed in the excitation-sculpting mode (164, 165). All data were processed using the Bruker TOPSPIN software. 10

The chemical shift differences ($\Delta\delta$) for the 1Hα protons were analysed with respect to the random coil values. From a detailed study performed by Wishart and co-workers (167) on 70 proteins, it is known that in all 20 naturally occurring amino acids 1Hα resonances move upfield ($\sim$0.39 ppm) from the random coil value when placed in α helical configuration, and downfield ($\sim$0.37 ppm) when the residue is placed in a β-strand or extended configuration.

3

NMR as a Tool to Study the Thermodynamics of Protein Folding

Despite our intuitive knowledge that conformational variability or dynamism is an intrinsic property of a polymeric chain, the view that we get from structural studies of the past is largely static. Although NMR and X-ray crystallography has provided important insights into the structure and function of proteins, we now know that it has mostly been brief snapshots from the complex life of proteins. Static protein structures focus on "ground-state" conformation with the lowest energy but function may be dependent on brief excursions to higher energy conformations of the protein and its binding ligands (Figure-Fig. 3.1). Today, proteins are viewed as an ensemble of alternative substructures in equilibrium with their "native state" (originally proposed by R. L. Baldwin and K. A. Dill in relation to protein folding and was later extended to describe native state ensembles (168)) and we have come to appreciate the intimate relationship between dynamics and molecular function. Understanding protein dynamics is of prime importance, not only in revealing its biological activity but also in understanding its role in ligand recognition that will help in future structure-based drug design programs. In this regard, experimental methods that has been developed to probe protein dynamics attract special attention.

Many spectroscopic and computational techniques have been designed to analyze various aspects of protein dynamics. X-ray crystallography by itself, X-ray and neutron

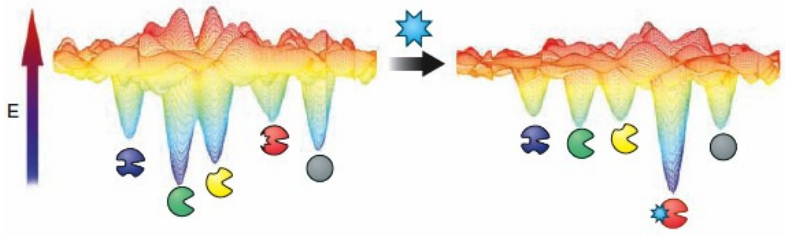


Fig. 3.1 Shallow wells and transitional barriers of a dynamic free-energy landscape - Figure shows the multiple preexisting conformational states of a macromolecule(an enzyme), corresponding to the catalytically active (green) and inactive (blue, red, yellow, and gray) forms. The global energy minimum shifts to favor the inactive conformation (red) when an inhibitor (cyan star) binds to an allosteric site. Figure adapted from (169)

scattering, fluorescence, hydrogen-deuterium exchange mass spectrometry, specialized techniques such as Mossbauer, Raman and 2D infrared spectroscopy all provide new insights into protein dynamics (170). Increasing computer power has seen more powerful methodologies emerging with increased time scales and protein sizes amenable to simulations. Computer simulations also form a theoretical basis for predicting molecular motion in solution and by incorporating inputs from various experimental techniques it outputs dynamic information beyond what can be accessed experimentally. NMR is an especially powerful technique for studying protein dynamics as the time scale available to NMR techniques cover all the relevant dynamic motions in proteins (Figure-Fig. 3.2). The multitude of time frames covered by NMR ranges over 17 orders of magnitude (10^{-12} to 10s) over length scales of 10^{-2} to $>10\mathring{A}$ (170), providing site specific information for motions from rapid bond librations (picoseconds) to events that take several seconds or hours (secondary structure folding/unfolding and solvent accessibility)(171, 172, 173, 174). Today NMR is vastly used to study kinetics, thermodynamics and structures of "native" and weakly populated protein states, thus filling gaps in the structure-dynamics-function paradigm.

3.1 Probing the protein unfolding landscape

The biophysical basis of how a polypeptide chain folds into a stable, native protein within a biologically relevant timescale has been a problem fascinating researchers for decades. The states of the conformational ensemble of a protein are populated

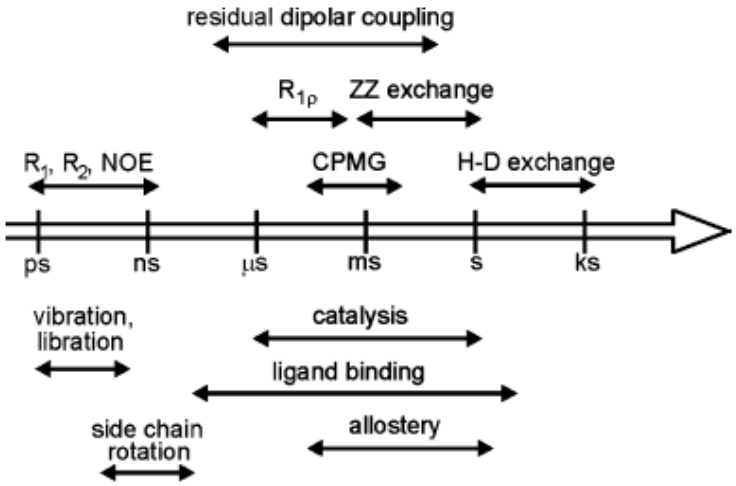


Fig. 3.2 Solution NMR techniques cover the complete range of dynamic events in enzymes - Figure adapted from (170)

according to the Boltzmann distribution, i.e. the population of a conformation j is proportional to exp[-E_j], where E_j is the Gibbs energy of the conformation. In normal conditions, the most populated form of the protein is the native state, with a Gibbs energy several kcal/mol lower than all the other states kinetically accessible in standard lifetimes. This implies that the native state forms more than 95% of the equilibrium population of a protein in solution. Since chemical equilibria are dynamic, every single protein molecule has a finite probability to transiently populate conformational states with higher energy. This process of transient conformational transitions, that do not change the overall population of the native state, is called equilibrium unfolding. Understanding the equilibrium unfolding landscape involves figuring out the dynamics and thermodynamics of the most populated conformational states, other than the directly observable native state. Knowledge about the equilibrium unfolding landscape of a protein is especially important in cases where non-native states play a crucial role in biological function or disease.

Investigations of protein folding by NMR have included measurements of protein folding kinetics and characterization of the ensemble of unfolded conformations. The use of isotopically labeled molecules has revolutionized NMR and has allowed the vast array of NMR techniques to be applied routinely with site specific resolution. These approaches are particularly important for unfolded or non-native states that frequently

display reduced chemical shift dispersion. Kinetics of protein folding have been investigated by techniques based on chemical exchange phenomena which provide unique information about the unfolded ensemble of conformations at equilibrium under native state conditions.

3.1.1 Hydrogen exchange

Hydrogen exchange is a powerful NMR technique used to gain insight into macromolecule dynamics and thermodynamics (175, 176, 177) and can be successfully used to characterize the equilibrium unfolding landscape of proteins. Hydrogen exchange is a simple chemical reaction in which a hydrogen atom, covalently bonded to nitrogen or oxygen, is replaced by another hydrogen/deuterium atom. The major attraction of hydrogen exchange experiments are the following (178); it is very easy to execute, the exchange can be monitored simultaneously at many individual residues in proteins, the replacement of hydrogen by deuterium has little or no effect on protein conformation and stability and a simple kinetic model for slow exchange in native proteins that has been proved to be accurate provides a sturdy framework for interpreting exchange data (179).

In a typical hydrogen/deuterium (H/D) exchange experiment, protein in H_2O is dissolved or diluted into D_2O and the replacement of protein hydrogens by solvent deuterons is monitored by NMR. In the case of proteins, the most useful hydrogen exchange process occurs in the amide protons of the backbone. This follows from the fact that most backbone NHs are involved in hydrogen-bonded structure, which appears to be the major determinant of slow NH exchange in proteins. Slow exchange of hydrogen-bonded and buried NHs is generally interpreted with the two-state model proposed by Hvidt (179). The chemical rate of hydrogen exchange for an amide proton of the residue j in a protein depends on several factors viz.; inductive electronic and steric effects arising due to the primary structure of the protein directly depending on the nature of the side chains at the immediate left (residue j-1) and right (residue j itself) of each residue and steric effects arising from the secondary, tertiary and quaternary structure of the protein.

The contributions deriving from the primary structure have already been parameterized by Bai et al. (180), using model dipeptides. From these standard values it is possible to calculate the contributions to the exchange rate arising from the primary

structure of a given aminoacid sequence. A zero deviation between the experimental exchange rates and the calculated ones would mean that the protein has a random coil conformation, therefore these standard values of rates are called k_{rc} (rc- random coil) values. The deviation between the experimental exchange rates and the k_{rc} are only due to the sterical effects arising from the conformation of the protein (the secondary, tertiary and quaternary structure of the protein). Amide hydrogens involved in intramolecular hydrogen bonds are inaccessible to the bulk water molecules of the solvent and are the major determinants of slow NH exchange in proteins. However, protein molecules are involved in the equilibrium unfolding processes, leading to the transient, though minimal, population of solvent-accessible conformations, or opened states. Several opening processes are possible, ranging from very local openings, leading to the exposure of just a single amide hydrogen, to the maximum, global opening, leading to the exposure of all the amide hydrogens together. This implies that it is the equilibrium unfolding dynamics that allows the hydrogen exchange of the buried amide protons and therefore the hydrogen exchange rates can be used in probing the equilibrium unfolding landscape of proteins (176).

A simple kinetic two-state model proposed by Hvidt (179, 181) has been helpful in analyzing slow exchange of hydrogen-bonded and buried NHs under the various conditions of stability. The model consists of two components: (1) a structural factor that takes into account the rate at which the protein becomes exchange competent or open (O) by losing any structural barriers to exchange, and (2) a chemical factor accounting for the exchange rate of protons that are fully exposed and unstructured (deduced from k_{rc} values). According to this model, the exchange rate of a proton that is intrinsically incompetent to exchange from the native state or closed (C) is described by:

$$\mathrm{NH}_{(C)} \quad \underset{k_{\mathrm{cl}}}{\overset{k_{\mathrm{op}}}{\rightleftharpoons}} \quad \mathrm{NH}_{(O)} \xrightarrow{\mathrm{krc}} \mathrm{ND} \tag{3.1}$$

where, the NH_C (closed state) and NH_O (open state) are the two possible conformations adopted by the slowly exchanging NHs and the rates of the opening and closing reactions given by k_{op} and k_{cl}. Exchange takes place only from the open state, with the rate constant denoted krc which is predicted from the method given by Bai et al (180). According to the model, the observed exchange-rate constant for slow NH exchange,

k_{ex}, can be defined as:

$$k_{ex} = (k_{op} \cdot k_{rc})/(k_{op} + k_{cl} + k_{rc}) \tag{3.2}$$

It is assumed that $k_{op} \ll kcl$ for native proteins and therefore equation 3.2 simplifies to

$$k_{ex} = (k_{op} \cdot k_{rc})/(k_{cl} + k_{rc}) \tag{3.3}$$

If $kcl \gg krc$, then kex can be expressed as:

$$k_{ex} = (k_{op}/k_{cl}) \cdot k_{rc} = K_{eq} \cdot k_{rc} \tag{3.4}$$

where K_{eq} is the equilibrium constant for the opening reaction. This kinetic limit where the equilibrium constant of the opening reaction dictates the rate of exchange is known as EX2-type exchange (pseudo-bimolecular reaction). EX2-type exchange relies entirely on a priori knowledge of the chemistry of exchange under the experimental solution conditions and is a powerful method to estimate the relative stability of various regions of protein molecules thus throwing light on the structural basis for the thermodynamic stability of proteins at residue-level resolution. If $krc \gg k_{cl}$, then kex will equal k_{op}.

$$k_{ex} = k_{op} \tag{3.5}$$

This kinetic limit is known as EX1-type exchange (pseudo-unimolecular reaction). Here, the rate of H/D exchange directly reflects the rate at which individual NHs are exposed to solvent. In the EX1 limit (equation 3.5), the observed exchange rate is equal to the opening rate. In the EX2 limit (equation 3.4), the observed exchange rate (k_{ex}) is equal to random coil calculated rate (k_{rc}), multiplied by the equilibrium constant of the chemical equilibrium between N and O (k_{op}/k_{cl}), i.e. the relative population of the opened state with respect to the closed native state. EX1 exchange may provide more direct experimental access to the dynamics of hydrogen-bond interactions in native proteins than the aforementioned EX2-type exchange thus giving us useful information regarding the kinetics of conformational interconversions.

For both types of exchange, apparent first-order kinetics are expected. The value of k_{ex} can be derived by fitting the experimental proton occupancy as a function of time to a simple exponential function.

Exchange data for a given NH are often reported as a protection factor (182) which is defined as the ratio of the rate of exchange of a given backbone amide when it is in a fully solvent-exposed state to the rate of exchange of that amide in some state of the protein or in some intermediate on the folding pathway of the protein. Therefore the protection factor P, is given by;

$$\mathrm{P} = \mathrm{k}_{rc}/\mathrm{k}_{ex} = \mathrm{k}_{cl}/\mathrm{k}_{op} = 1/\mathrm{K}_{eq} \tag{3.6}$$

Under EX2 conditions, the protection factor is related to the apparent free energy of exchange, ΔGex, as shown in equation 3.7.

$$\Delta\mathrm{G} = \text{-}\mathrm{RT}\ \ln\ (1/\mathrm{P}) \tag{3.7}$$

where R is the gas constant and T is the absolute temperature.

The two state model as mentioned above is described for slowly exchanging NHs. The most slowly exchanging NHs that are involved in secondary structure and buried in the protein interior exchange only upon complete unfolding and therefore their free energy of exchange, $\Delta\mathrm{G}_{\mathrm{HX}}$, is equal to their free energy of unfolding, $\Delta\mathrm{G}_{\mathrm{u}}^{\circ}$ over a wide range of experimental conditions. For those amide hydrogens that exchanges faster (typically involved in hydrogen-bond interactions at the periphery of secondary structure and in loops) the $\Delta\mathrm{G}_{\mathrm{HX}}$ is less than $\Delta\mathrm{G}_{\mathrm{u}}^{\circ}$. The molecular phenomena responsible for slow-exchange is referred to as global unfolding reaction and those for fast exchange is called as partial openings or local fluctuations (175, 182). Protein destabilization brought about by physical factors like temperature and pH or chemical denaturants leads to a decrease in $\Delta\mathrm{G}_{\mathrm{u}}^{\circ}$ and is immediately manifested in the $\Delta\mathrm{G}_{\mathrm{HX}}$ of the amides exchanging by global unfolding. Amides exchanging by local fluctuation on the other hand are largely unaffected by decrease in $\Delta\mathrm{G}_{\mathrm{u}}^{\circ}$ until the point where global unfolding becomes the dominant contributor to slow exchange also at these sites resulting in a situation where the $\Delta\mathrm{G}_{\mathrm{u}}^{\circ}$ becomes equal to $\Delta\mathrm{G}_{\mathrm{HX}}$.

A protein in solution experiences both local and global unfolding and may be represented as;

$$\mathrm{N} \quad \underset{k^1_{\mathrm{PN}}}{\overset{k^1_{\mathrm{NP}}}{\rightleftharpoons}} \quad \mathrm{P}_1 \xrightarrow{\mathrm{krc}} \mathrm{P}_1^*$$

$$\mathrm{N} \quad \underset{k^2_{\mathrm{PN}}}{\overset{k^2_{\mathrm{NP}}}{\rightleftharpoons}} \quad \mathrm{P}_2 \xrightarrow{\mathrm{krc}} \mathrm{P}_2^* \ldots\ldots$$

$$\mathrm{N} \quad \underset{k^j_{\mathrm{PN}}}{\overset{k^j_{\mathrm{NP}}}{\rightleftharpoons}} \quad \mathrm{P}_j \xrightarrow{\mathrm{krc}} \mathrm{P}^*_{\mathrm{j}}$$

$$\mathrm{N} \quad \underset{k_{\mathrm{DN}}}{\overset{k_{\mathrm{ND}}}{\rightleftharpoons}} \quad \mathrm{D} \xrightarrow{\mathrm{rpxkrc}} \mathrm{D}^*$$

where, D is the globally unfolded exchange-competent state and Pj denotes all the partially opened species; the asterisk indicates exchange, while k_{rc} is the contribution of the primary structure (random coil) to the exchange rate as in the previous case and rp accounts for residual structure and protection in the global unfolded state (182).

According to this model the k_{ex} has contributions from both the partial and global unfolding and can be written as,

$$k_{ex} = \frac{k_{ND} \cdot k_{rp} \cdot k_{rc}}{k_{DN} + rp \cdot k_{rc}} + \sum_j \frac{k^j_{NP} \cdot k_{rc}}{k^j_{PN}} \tag{3.8}$$

assuming that $k^j_{NP} \ll k^j_{PN}, k_{ND} \ll k_{DN} and k^j_{PN} \gg k_{rc}$. The first two assumptions hold true as native state is the most stable conformational state under the HX experimental conditions of folded proteins. The third relation is based on the assumption that the relaxation to the native state from a partially open conformation is very fast, because only a single residue or, at most, small regions of the protein are involved in the local opening process. If we disregard this third assumption the second term in the equation above changes itself to,

$$\sum_j \frac{k^j_{NP} \cdot k_{rc}}{k^j_{PN} + k_{rc}}$$

The EX1 and EX2 limits described above for the two state model in this case is,

$$k^{EX2}_{ex} = \frac{k_{ND} \cdot k_{rp} \cdot k_{rc}}{k_{DN} + rp \cdot k_{rc}} + \sum_j \frac{k^j_{NP} \cdot k_{rc}}{k^j_{PN}} for (rp \cdot k_{rc}) \ll k_{DN} \tag{3.9}$$

$$k^{EX1}_{ex} = k_{ND} + \sum_j \frac{k^j_{NP} \cdot k_{rc}}{k^j_{PN}} for k - DN \ll (rp \cdot k_{rc}) \tag{3.10}$$

Acquiring 1H-15N HSQC spectra has been the standard way to monitor the exchanging protons for proteins with already assigned spectra. Typically HSQC spectra are recorded as a function of time (typically 10-100 hours, depending on the conditions) while the hydrogen is exchanging with the deuterium. The apparent exchange rates of

the individual residues are then extracted by exponential fitting and the corresponding thermodynamic parameters calculated according to Eqs. 3.9 or 3.10. The series of experiments are repeated at different temperatures with freshly prepared samples and the changes in the individual equilibrium constants can be assessed to determine the enthalpy and entropy differences for each local process and for the global one. Identifying the global process relies on observing similar thermodynamic parameters (ideally the same single set) from a number or a few amides. This method gives residue-specific information for all the residues in the protein simultaneously and helps in extracting the free energy, enthalpy and entropy changes of global unfolding.

Recently in Udine (Rennella et al., 2011 submitted) a method was developed that allows the acquisition of all the series of experiments at different temperatures with a single series collected as a function of time and temperature. A single sample was used (that was previously exchanged in D2O and lyophilized) that was dissolved in an H2O buffer and subjected to a thermal ramp during the exchange process while acquiring a series of 2D NMR spectra. The same sample was monitored at the end of the exchange process by repeating the thermal ramp and acquiring the same series of experiments. In the first series, the time evolution of the signals from amide groups, denoted S1 signals, would contain not only the build-up due to the hydrogen exchange process, but also the unwanted contribution that arises from the thermal dependence of spin relaxation. The time evolution of the signals from the second series (denoted S2 signals) however would contain only the contribution from the thermal dependence of spin relaxation. The authors thus described a method to extract a quantity S3 = (S2 S1)/S2 to account for the hydrogen exchange process. Therefore the buildup of signals in S1 was normalized and the time evolution curves of S3 were called decays.

I have used the same experimental method to monitor the exchange process of β2m with carnosine and with heparan sulfate. The decays have been obtained in this way, subjected to analysis and the thermodynamic parameters extracted considering an EX2 limit .

3.2 Materials and methods

3.2.1 Sample preparation for Hydrogen/Deuterium (H/D) exchange experiments

Wild-type human β2m was expressed and purified as previously reported,21 with additional 15N labelling. An additional methionine (Met-0) was always present at the N-terminal of the protein. Carnosine was bought from Fluka and a 1M solution was prepared and diluted to the required concentrations as mentioned below.

For experiments with B2m-carnosine, 4.5mg of B2m was dissolved in 1.5ml of buffer prepared (concentration measured as 0.47mM) in D2O (7mM sodium phosphate, 10mm sodium chloride pH 6.65) and left at 37°c for 16 hours and split into 2 tubes. The protein was lyophilized and the powder was redissolved in D2O and left at 37°c for 16 hours twice to ensure the complete exchange in D2O. The final lyophilized powder in one tube was dissolved in 500μl of H2O:D2O (93:7) buffer containing 63 mM sodium phosphate buffer, 90 mM sodium chloride (pH 7.0) and 0.471mM carnosine (after carnosine addition the pH of buffer was 6.96) for the b2m:carnosine (1:1) sample. The pH measured after experiments was 6.88. For the b2m:carnosine (1:100 sample) the final buffer in H20:D20 (93:7) was prepared at pH 6.43 and 47mM carnosine was added. The pH of the sample after experiments in this case was 6.99. The experiment blank was pre-exchanged and lyophilized in the same way and did not contain carnosine in the final buffer. The pH measured was 7.0.

For experiments with B2m-Heparan sulfate, 6mg B2m was dissolved in 2ml of buffer prepared in D2O (concentration measured as 0.46mM)(10mM sodium phosphate, 10mM sodium chloride), left at 37°c for 16 hours and split into 4 tubes. The protein was lyophilized and the powder was redissolved in D2O and left at 37°c for 16 hours twice to ensure the complete exchange in D2O. Final buffer (10mM sodium phosphate, 90mM sodium chloride) was prepared in H2O:D2O (93:7) at two different pH of 6.7 and 7.0. 500μl of each buffer was used as it is for the two blanks and 10% heparan sulfate (46μM) was added to the other two tubes alonwith the respective buffers just prior to spectra acquisition. The blank samples read 6.44 and 7.0 after experimentation and the corresponding samples with heparan sulfate read a pH of 6.25 and 7.0

In all the above cases, care was taken to add the final buffer (in H2O) immediately before the experiments and acquisition was started within 4minutes of the buffer

addition.

3.2.2 H/D exchange experiments by NMR spectroscopy

All NMR experiments were carried out at 11.7 T with a Bruker Avance-DRX 500-MHz spectrometer equipped with triple-axis magnetic field gradients.

Before starting an experiment, the sample tube containing the buffer to be used for sample dissolution (500μl) was inserted into the machine and the acquisition parameters optimized. This was done in order to save time with the real sample and to start recording the hydrogen exchange from time zero (T $_0$). The same tube was washed with Milli-Q water and dried well before using again for the experiment. After dissolving the pre-exchanged powder the exchange process was monitored by running a series of 1H,15N-SOFAST-HMQC experiments each lasting less than 8 minutes over a range of temperatures from 293.15K to 313.13K, with an increment of 0.1K per experiment. The pulse program with reduced scan times were used to get a sensitivity advantage rather than a reduction in the acquisition time required for two-dimensional spectra. In these experiments, the H^N excitation pulse was obtained with a polychromatic PC9 (183) shaped pulse with a flip angle of 120°, while for the H^N refocusing pulse a REBURP(184) shape was used; these two band-selective H^N pulses were centred at 9.5 ppm for β2m and covered a bandwidth of 4 ppm.^{15}N decoupling was applied during signal detection (55 ms) through a GARP-4 sequence(185) with a field strength of g B1/(2 p) = 625 Hz. The inter-scan recovery delay, optimized in order to maximize the signal-to-noise ratio per unit time for β2m, was 135 ms, while 40 complex points were acquired in the heteronuclear dimension for each experiment over a spectral width of 1,550 Hz, with 24 scans/t1-point and 768 points in t2, over a spectral width of 7,000 Hz. The multiple acquisition of 1H,15N correlation spectra as a function of temperature was performed by means of the standard multi_zgvt macro of the Topspin (Bruker) interface. At the end of the ramp, subsequent HSQCs were acquired every two hours at 293.15K to ensure completion of any residual exchange and when the sample seemed stable with no new peaks arising or peak intensities remaining constant , the same cycle of experiments was repeated on the exchanged sample in order to acquire the reference data set.

NMR data processing (Refer the paper "Single-shot NMR measurement of protein unfolding landscapes" (submitted) by Enrico et al for a detailed description of the

method given below) from time-domain data to frequency-domain data was performed using NMRPipe (186) Calculations for conversion of frequency-domain NMR data into experimental peak decays were performed with Mathematica 6 (Wolfram Research, Inc.) running on a 2.80 GHz Intel® Core™ i7 CPU processor mounted on a Linux-based computer. Four mathematica notebooks were used, the first two were to convert the frequency domain spectra to the corresponding peaks, the third one was to convert the raw data to peak decays. This algorithm efficiently picked required points for each peak, calculated the the total intensities for each reference spectrum as sum of all the intensities of the constituent points. It then used the same points to calculate the intensities in the exchange spectra. Therefore from each pair of reference and exchange spectra the out put consisted of a list of peaks at each acquisition time/temperature with peak positions and relative intensities that was calculated as;

$$\frac{Intensity_{reference} - Intensity_{exchange}}{Intensity_{reference}} \tag{3.11}$$

The next step was to compare the created peaklists and to collect peaks having correlated positions in both ^{15}N and 1H dimensions versus time/temperature, that most likely belong to the same amide site. The dependance of the resonance chemical shifts on temperature was taken into account for tracing the positions. The algorithm finally assigned peaks to single amide sites, using the chemical shifts reported in BMRB database under the accession codes 307825 for β2m (187).

Next step was to calculate the free energy of unfolding (ΔG_u) using the decays. Only the decay curves that were good (decayed to zero) were chosen for this calculation. The thermal dependance of k_{ex} in a single time evolution curve can be described by the formula;

$$N(t) = N(0) \cdot exp\left(-\int_0^{t\prime} k_{ex}(T(t)dt)\right) \tag{3.12}$$

where, N(t) is the time course of the N species exchange. The variation of temperature (T), during time (t) propagates into the variation of k_{ex} rates and is expressed by the term $k_{ex}(T(t))$. To estimate the $k_{ex}(T)$ from the time course of the peak decays, the experimental signal for each decay was to be fitted with the function;

$$N(t) = c \cdot exp(a \cdot exp(b \cdot t)) \tag{3.13}$$

where a, b, and c are the fitting parameters. By taking the negative first derivative with respect to time of the logarithm of eqn 3.13, we arrived at the function for the exchange rates as;

$$k_{\mathrm{ex}}(T) = -a \cdot b \cdot exp(b \cdot t(T)) \tag{3.14}$$

Assuming an EX2 limit, the thermodynamic parameters were calculated from the derived k_{ex} according to the eqn 3.7, ΔG = -RT ln (k_{ex}/k_{rc}). Residue-specific values of k_{rc} used for the calculations were computed using the spreadsheet available online (Englander Lab Website. http://hx2.med.upenn.edu/download.html) (180).

In the more rigorous method described in the paper (Enrico et. al, submitted), the estimated rates are then used to derive the thermodynamic equilibrium parameters for the various opening process taking place according to equations 3.9 and 3.10. This is followed by a combination of clustering and selection steps to identify the kinds of openings, viz. local, sub-global and local that are involved. The clustering step searches for groups of amide sites with correlated rates, i.e. exhibiting exchange rates that could be explained within the errors, by a single global opening process with a defined value of $k_{ND}(T) and k_{ND}(T)$. The result is a set of clusters each made by several amide sites that can exchange through global unfolding only, without any contribution from partial openings or residual protections. In the selection step, one of the clusters is chosen as the apt representative of the global unfolding process. In the clusters identified the amide sites are grouped into three categories; viz. hydrogens exchanging through global unfolding only, hydrogens with residual protection in the globally unfolded state and/or hydrogens with contributions from partial openings. So each cluster corresponds to a different candidate model that is the result of grouping the amide sites differently into the above three classes. The different candidate models are then compared using the Bayes Information Criterion (BIC), that is defined as;

$$BIC = N \cdot ln(RSS) + kln(N) \tag{3.15}$$

where, N is the total number of observations, RSS is the sum of the squares of the residuals relative to the model and k is the number of fitting parameters for the model according to (188). The most plausible model is described as the one with the lowest BIC value. Once the proper model is selected, the clustering and selection procedures

are repeated to identify possible sub-global openings. The errors of the estimated thermodynamic parameters are finally calculated through Monte Carlo simulations.

Under the selected experimental conditions in the range 293.15K to 313.13K, β2m HX data can be analysed by single EX2 process, or by dissecting global and local processes (with the inclusion of the selection and clustering steps). In the case of the exchange studies with β2m-carnosine, the data were analyzed in both the ways (explained in Chapter 4) whereas in the case of β2m-HS the data was analyzed by assuming a single EX2 process.

4

Results and Discussions

4.1 L-Carnosine interacts efficiently with α-crystallin

4.1.1 Titration studies by STD NMR

Titration studies were performed as a function of L-carnosine concentration at different concentrations of α-crystallin, on a monomer subunit basis, to determine the effect of the ligand on STD signal intensities. Fig. 4.1 shows an overlay of the difference STD spectra at various stages of the titration study. The intensity of the reference peak can be seen as remaining constant and that of carnosine increasing with the increase in its concentration. As described before, the protein resonances are not seen in the difference spectra and the resonances of carnosine that experienced saturation due to its proximity to the protein are seen, the intensity of which increases with the increase in the carnosine concentration. It is clear that the system exhibits efficient saturation transfer.

Within the explored conditions, the effects proved related to the protein:ligand ratio, rather than to the absolute α-crystallin concentration. Therefore, only the most extensive group of data obtained at 0.1 mM α-crystallin concentration are further described. The titration plot in Fig. 4.2 shows that the increase of the STD signal is more pronounced at low ligand protein ratios and that it slowly levels off at high ligand protein ratios. The typical STD pattern expected from a ligand exchanging in the fast regime, i.e. with a k_{off} between 10^2 and $10^7 s^{-1}$, is a Langmuir isotherm that can provide, by proper fitting, the value of the thermodynamic dissociation constant of the ligand-protein complex. The occurrence of a fast interaction regime between

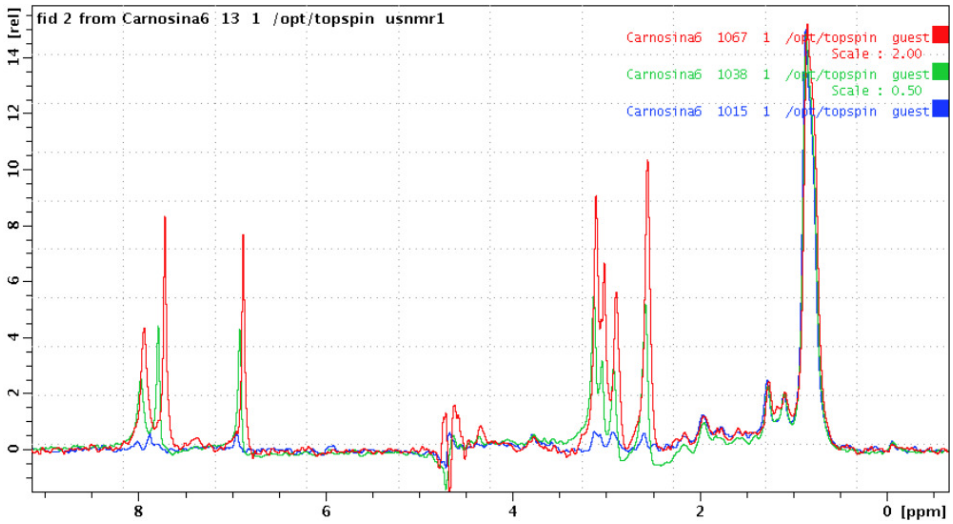


Fig. 4.1 Overlay of 1-D STD difference spectra at increasing concentrations of carnsoine - The reference peak is seen at 1ppm and the aromatic peaks of carnosine used for the analysis are seen between 6.5 and 8ppm

L-carnosine and α-crystallin is confirmed by the observation of a single spectroscopic species and saturation transfer effects that increase with L-carnosine concentration. However, the extent of such increase is far greater than the typical stoichiometric ratios that are explored when studying 1:1 interactions, i.e. the experimental STD amplification factor (A_{STD}) values start to decrease towards the theoretically expected plateau at L-carnosine concentrations about three magnitude orders above the α-crystallin concentration (Fig. 4.2). We do not see the typical plateau phase, as expected for a specific binding equilibrium, though the increase in the saturation transfer effect reduces considerably at very high ligand concentrations. This points out to a very efficient turnover of bound and free states and possibly the occurrence of multiple equivalent binding sites on the protein, in addition to the T1 relaxation time of the free ligand and the exchange kinetics of the system (149). Large saturation effects could also be because of weak interactions and long relaxation times. Also, small molecules with fast exchange kinetics from the bound to the free states show this trend in saturation transfer experiments. An efficient turnover shows maximum saturation transfer effects at high ligand concentrations, as ligand molecules having received no saturation increase with

increasing ligand excess.

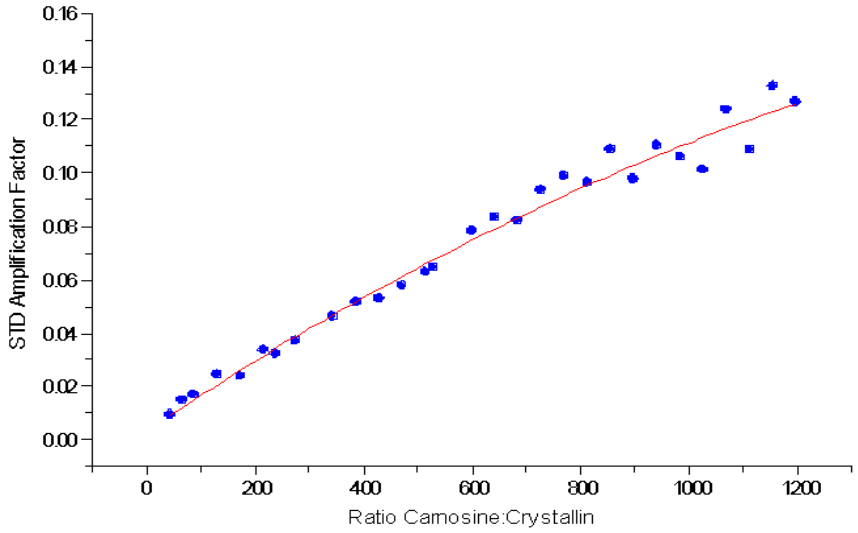


Fig. 4.2 Saturation transfer of α-crystallin (0.1 mM) probed by L-carnosine at 25°C. - The intensity of the STD spectra is plotted against the ligand-to-protein concentration ratio.

4.1.2 STD experiments under destabilizing conditions

We studied the saturation transfer effects of α-crystallin treated with guanidinium hydrochloride (with and without heat treatment) using L-carnosine. As discussed previously, guanidinium is a deaturant that unfolds proteins and alters their three-dimensional structure. In solution the positively charged guanidinium ions bind to the proteins thereby shielding electrostatic and polar interactions and establishing preferential interactions with the polypeptide. At high denaturant concentrations this pushes the protein conformational equilibrium towards the unfolded state (189). Thermal denaturation, instead, essentially weakens and breaks the hydrogen bonds, the electrostatic interactions and eventually the hydrophobic interactions that stabilize the secondary and tertiary structure of a protein, thereby favouring the unfolding process.

Our results showed that incubation of α-crystallin samples with 1M Guanidinium hydrochloride at room temperature did not affect the saturation transfer to L-carnosine

and the intensities of the STD signals are same as in the samples without denaturant. This has been repeatedly tested at specific protein/ligand ratios (1:25 and 1:200).

However, α-crystallin samples pretreated with 1M guanidinium hydrochloride and heated at 60°C for 24 hours to form amyloid fibrils (75) did not interact with L-carnosine at low concentrations (ratios of 1:5 and 1:15) but interacted at higher ratios (1:50, 1:100), although with considerably reduced intensity of the STD signals as compared to the untreated samples (Figure-Fig. 4.3).

To check if L-carnosine has any protective effects against thermal treatment in the presence of denaturant, we pretreated α-crystallin with varying concentrations of L-carnosine and then submitted it to 1M guanidinium hydrochloride treatment and heating at 60°C for 24 hours. Again, saturation transfer was observed only at higher concentrations of L-carnosine (ratios of 1:50, 1:100) although the transfer in this case was reduced with respect to that observed when L-carnosine had not undergone the whole treatment applied to the protein (Figure-Fig. 4.3). It can not be ruled out that the thermal treatment in the presence of denaturant could induce partial degradation of L-carnosine.

4.1.3 STD NMR control experiments (Lysozyme and βL-Crystallin)

To check if the saturation transfer effects seen in the α-crystallin-L-carnosine system arise from a real, albeit weak, interaction and are not due to a non-specific process (statistical collisions) because of the co-existence of both species in solution at high ligand:protein ratio, we conduced control STD experiments on L-carnosine with lysozyme and with βL-crystallin. The latter lens protein is a heterogeneous oligomeric mixture of related subunits from dimmers through to octamers (mass range about 50-200kDa). That that is a natural target protein for α-crystallin in the lens. Lysozyme and β-crystallin posses no chaperone activity. We conducted these experiments at two different ratios of protein:ligand, viz. 1:20 and 1:200 and observed that L-carnosine interacted with lysozyme although the intensity of this interaction was considerably lower when compared with that obtained with α-crystallin (Fig. 4.4), but is comparable to that observed for α-crystallin treated with guanidinium chloride (Fig. 4.3). However, in the case of β-crystallin, no STD effects were observed at both the ratios (Figure-Fig. 4.4), suggesting that β-crystallin does not interact with carnosine and that statistical encounters do not account for the results obtained with α-crystallin.

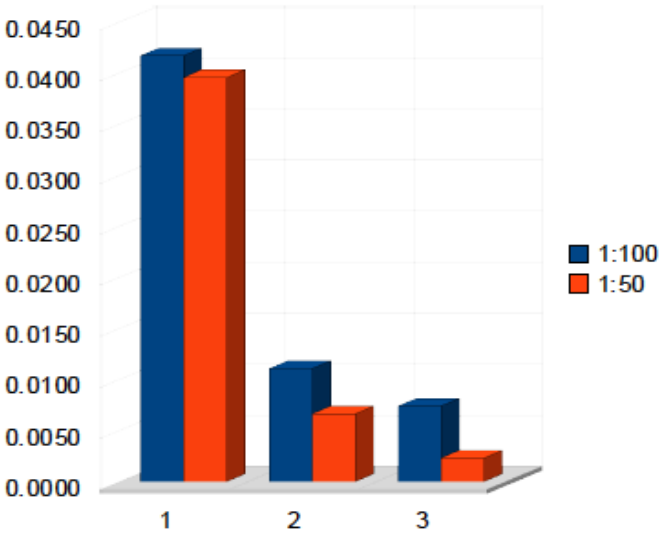


Fig. 4.3 Comparative saturation effects at two different ratios of α-crystallin:L-carnosine plotted against the A_{STD} values - colours: blue bars = ratio 1:100; red bars = ratio 1:50, The three sample sets are; 1: untreated a-crystallin, 2: α-crystallin incubated with 1M guanidinium hydrochloride and heated at 60°C for 24 hours before addition of L-carnosine, 3: α-crystallin incubated with L-carnosine before treatment with 1M guanidinium hydrochloride and heating at 60°C.

4.1.4 Diffusion measurements

NMR DOSY experiments on isolated α-crystallin (0.5 mM at 25°C) give a diffusion coefficient of 3.3×10^{-11} m^2/s. For an isotropic spherical model, t his rather slow translational diffusion would be consistent with a hydrodynamic radius (R_h) of 7.3 nm and a molecular weight of $1.3\text{-}1.4 \times 10^3$ kDa. There are some caveats that one should consider when interpreting DOSY experiments. First, for large molecule assemblies, estimates may be substantially affected by anisotropy contributions. Second, the type and size of aggregates that are seen by NMR may be different from the aggregates that determine the largest contributions to light or X-ray scattering.

Therefore it is safer to evaluate the DOSY results only from a qualitative viewpoint. The evidence obtained through DOSY indicates that carnosine has no major effect on the hydrodynamic parameters of α-crystallin, except for aged samples. For long incubations with L-carnosine at room temperature (say, above 2-4 weeks, depending on the

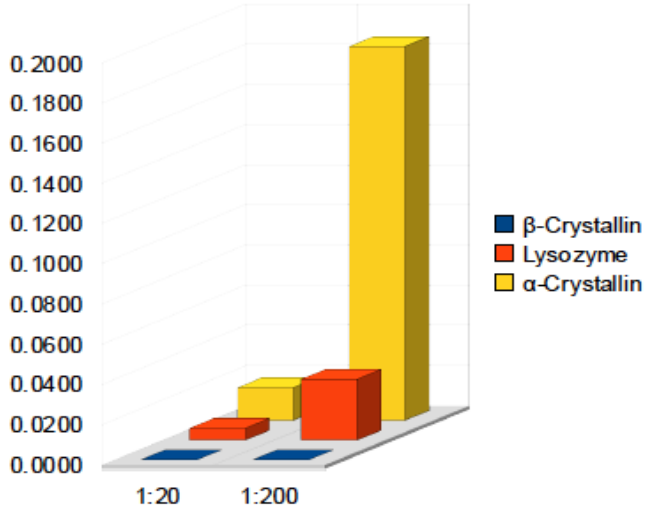


Fig. 4.4 Comparative saturation transfer effects at two different ratios of protein:ligand for different systems studied. - α-crystallin (yellow bars), βL-crystallin (blue bars) and lysozyme (red bars) L-carnosine systems. On the x-axis are the protein-ligand ratios an on teh Y axis the calculated A_{STD} values.

preparations), the α-crystallin aggregate diffusion rate increases which may either be consistent with smaller aggregate dimensions or anisotropy decrease Fig. 4.5. These results do not demonstrate, however, any causality relationship between incubation with L-carnosine and loss of native aggregate dimensions. The observed behaviour of crystallin samples rather prompts prudence when assessing the results with old samples that may be affected or determined by processes that are not under control such as a slow onset of fibrils. Pre-incubation with 1 M guanidine at 60°C for 12 or 24 hours increases the diffusion rate of α-crystallin (Figure-Fig. 4.5). This may either suggest that only smaller aggregates than those of native α-crystallin remain NMR-visible or that the aggregate anisotropy sensibly decreases by heating in the presence of 1 M guanidine. No effect from sample ageing is observed whereas the addition of L-carnosine, even at low concentration ratios, seems to rescue α-crystallin that in part recovers the diffusion coefficient of the unperturbed aggregate Fig. 4.5.

The results obtained by STD on the system α-crystallin/L-carnosine are consistent with the occurrence of a weak, but very efficient interaction that is related to the actual solution dynamics of the small heat shock protein. This interaction is unaffected by

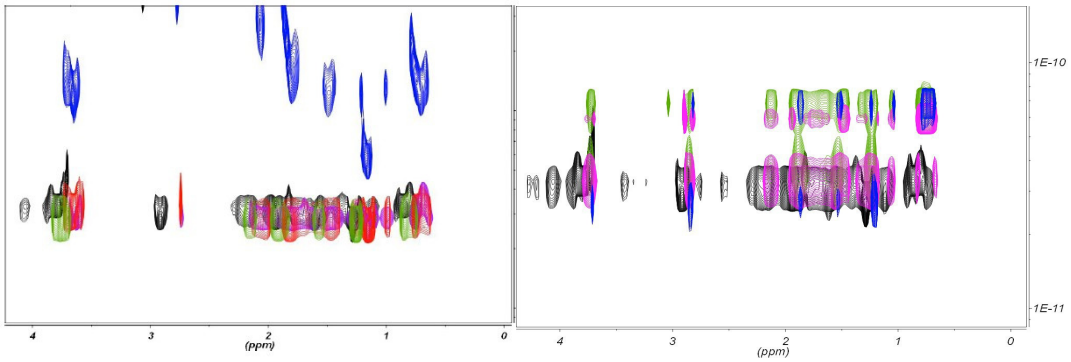


Fig. 4.5 Overlay of DOSY spectra - Left Panel: NMR DOSY at 25°C of aqueous α-crystallin (0.5 mM) alone (black contours) and inthe presence of L-carnosine 2.5 mM (purple contours), 11 mM (red contrours), 17 mM (red contours) and 17 mM with two months incubation (blue contours). Right Panel: NMR DOSY at 25°C of 0.5 mM aqueous α-crystallin alone (black contours) and after incubation with 1 M guanidine at 60°C for 12 hours, without sample ageing (green contours), with 1-month sample ageing (blue contours) and on addition of 1 mM L-carnosine to the freshly prepared sample (purple contours).

incubation with a chemical denaturant such as 1M guanidinium at ambient temperature. The remarkable extent of saturation transfer brought about by α-crystallin is apparently quite independent of the L-carnosine concentration. Control experiments with hen egg white lysozyme (Mw = 14.3 kDa) and βL-crystallin oligomers (Mw = 50-200 kDa) show quite limited (lysozyme) or extremely poor (βL-crystallin) saturation transfer to L-carnosine magnetization. Since similar concentration ratios and absolute values were tested with α-crystallin and in control experiments, the latter outcomes suggest that random encounters of large monomeric or oligomeric protein assemblies with L-carnosine cannot be responsible for the result observed with α-crystallin. Only under conditions leading to fibril formation by α-crystallin (75), i.e. heating at 60°C for 24 hours of the protein incubated with 1 M guanidinium in 20 mM phosphate (pH 7.2), the interaction with L-carnosine proves substantially weakened (albeit still stronger than with lysozyme and β L-crystallin), regardless whether L-carnosine is or not co-incubated with the denaturant and the protein prior to heating. This substantially compromised interaction capability of α-crystallin with L-carnsoine upon heating in the presence of denaturant suggests that the osmolyte is not capable of restoring the thermal stability of the native small heat shock domain. This may be related to an

inherent irreversibility of the amylodogenic transition of the protein that requires the simultaneous action of heat and guanidinium interaction. Guanidinium alone does not seem sufficient for this transition as indicated by the unchanged extent of L-carnosine STD measured by NMR in samples with guanidinium but no thermal pre-incubation with the denaturant. Although the osmolyte characteristics of L-carnosine are not strong enough to revert the transformation undergone by α-crystallin heated in presence of guanidinium, L-carnosine exerts some action on the pre-treated α-crystallin samples as is evident from the NMR DOSY results Fig. 4.5.

4.2 L-Carnosine either interacts transiently or does not interact with amyloidβ-peptide

Due to the poor solubility of the 1-42Aβ-peptide we have used the Aβ12-28 for all our NMR experiments. The results presented are based on two samples of the same in varying buffer conditions. The first sample was prepared in 80/20 water/d6-DMSO mixture with or without 50 mM d5-glycine at pH 3.35. All experiments with this sample were done at 21°C. To rule out any unfavourable effect of DMSO on interaction of carnosine with the peptide, a second sample starting from a basic solution of the peptide (pH >10), subsequently neutralized with phosphate was prepared. Experiments with this sample were initially performed at low temperatures (6.7-12°C) and later at higher temperatures (25-30°C).

4.2.1 Results from NOESY and TOCSY spectra

A four fold excess of Carnosine addition to a 1 mM Aβ12-28 sample brought about a systematic pH increase from 3.8 to 4.26. Therefore the chemical shift observed in the spectra on addition of Carnosine, at residues E22 and D23 side chains and K28 backbone (the C-terminal residue) can be attributed to the change in carboxyl-carboxylate ratio due to the increase in pH. Other limited chemical shift changes occurred at the backbone NH-αCH connectivity of A21 and S26. The most relevant changes were observed for the intensities of L17, V18, F19 and F20 spin systems that appeared significantly reduced, much like the reduction/disappearance seen with the backbone and side chain resonances of E22 and D23 (Fig. 4.6). Overall this pattern suggests that the inner stretch of Aβ12-28, i.e. L17-F20, undergoes slow conformational exchange related to slowed acid-base equilibria at E22 and D23 side chains that, in turn, may trigger intra and intermolecular hydrophobic interactions because of the overall net charge decrease when pH shifts towards the isoelectric point of the molecule which is 6.89 (190, 191).

To prevent unwanted effects from the increase in pH we prepared a sample with 20 mM d5-glycine, a condition that ensured a substantial pH stability up to a Aβ12-28/carnosine ratio of 1:2. A small pH increase, 0.1 unit, was observed when a 1:3 ratio was reached; the overall $\Delta\delta$ pattern is reported in Fig. 4.7. The small pH change also limited the resonance attenuation within the segment L17-F20 (compare Figures-Fig. 4.6 and Fig. 4.7), therefore confirming that the main effect on the NMR spectrum

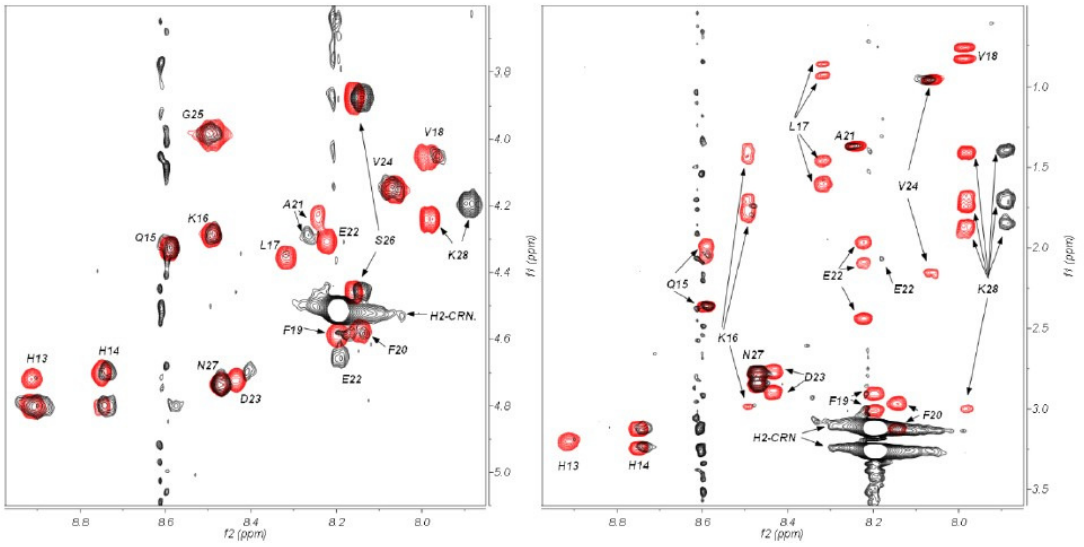


Fig. 4.6 ^{1}H2D TOCSY regions of 1.0 mM Aβ12-28 in H2O/d6-DMSO 80/20 at 21°C, in the absence (red contours) and in the presence (black contours) of 4 mM carnosine (CRN). - The contour plots show, on the left, the amide connectivities to the αCHs (and βCH2 for S26) and, on the right, the propagation to the other resonances of the spin systems, with the assignments indicated by the amino acid residue single letter code. Despite the same isotropic mixing time and power employed for data collection, the spin systems of the stretch L17-F20 prove attenuated in the presence of carnosine along with the patterns of E22 and D23. Similar features can be observed also in corresponding symmetric quadrants of the TOCSY spectrum. The apparent attenuation of H13 spin-system pattern is due, instead, to partial solvent suppression transfer to the αCH resonance (approaching the solvent frequency on carnosine addition).

of Aβ12-28 from carnosine addition mentioned in the above paragraph comes from carboxyl-carboxylate relative population changes. A further confirmation was obtained from experiments obtained under conditions ensuring a more stringent pH control by using 50 mM d5-glycine at pH 3.35 and by readjusting the pH to this value after addition of carnosine. The spectra thus acquired showed no chemical shift changes (Figures Fig. 4.8 and Fig. 4.9) thus confirming that carnosine does not affect the solution state of Aβ12-28, at least under the selected experimental conditions of both NOESY and TOCSY.

DMSO was used in the above two cases to prevent aggregation but the effects of DMSO on the possible interaction of carnosine with Aβpeptide cannot be ruled out

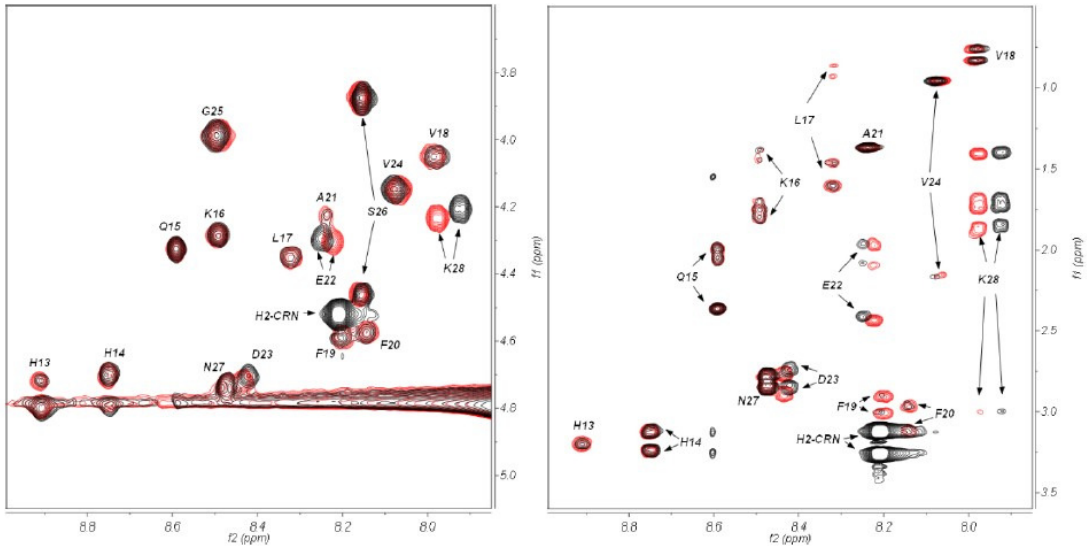


Fig. 4.7 ^{1}H2D TOCSY regions of 0.93 mM Aβ12-28 in H2O/d6 -DMSO 80/20 with 20 mM d5 -glycine at 21°C, in the absence (red contours) and in the presence (black contours) of 2.7 mM carnosine (CRN). - The contour plots show, on the left, the amide connectivities to the αCHs (and βCH2 for S26) and, on the right, the propagation to the other resonances of the spin systems, with the assignments indicated by the amino acid residue single letter code. Only chemical shift changes involving E22, D23 and K28 resonance patterns are clearly observed due to carnosine addition, without significant attenuation of the resonances in the stretch L17-F20.

completely. Thus we resorted to a protocol suggested elsewhere (192) to prevent aggregation of the full-length Aβ, i.e. starting from a basic solution of the peptide (pH >10), subsequently neutralized with phosphate (see Materials and Methods). To disfavour hydrophobic interactions most of the measurements were performed at low temperatures (6.7-12°C). Higher temperature (25-30°C) experiments were run only at a later stage. Figure-Fig. 4.10 depicts the overlay of the amide connectivity propagation from 2D TOCSY spectra of Aβ12-28 alone and in the presence of carnosine (1:2 peptide-to-carnosine). For each stoichiometric equivalent addition of carnosine the pH value increased progressively from the initial value of 7.36, reaching finally a value of 7.48, but no noteworthy shift was detected for the residues with carboxyl groups, i.e. E22, D23 and the C-terminus. This behaviour reflects the virtually complete deprotonation of all carboxylates at pH 7.36, a value more than 3 units above the typical side chain carboxyl group pKa. Chemical shift invariance with respect to carnosine addi-

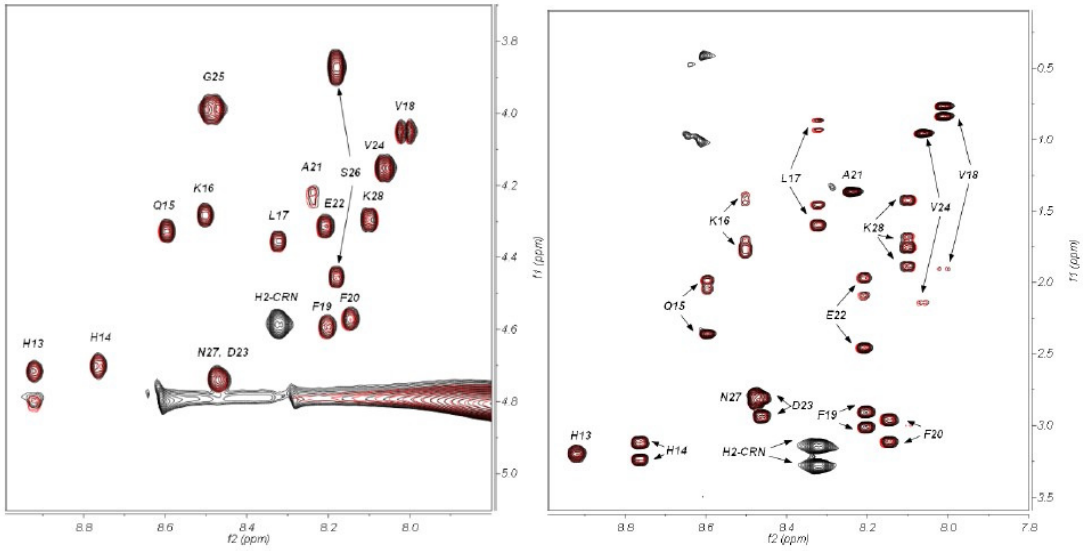


Fig. 4.8 ^{1}H2D TOCSY regions of 1.0 mM Aβ12-28 in H2O/d6 -DMSO 80/20 with 50 mM d5 -glycine at 21°C, in the absence (red contours) and in the presence (black contours) of 3.4 mM carnosine (CRN). - The contour plots show, on the left, the amide connectivities to the αCHs (and βCH2 for S26) and, on the right, the propagation to the other resonances of the spin systems, with the assignments indicated by the amino acid residue single letter code. Even under these conditions, the excess of carnosine was able to raise the pH from 3.35 to 3.65, and induce shifts of E22, D23 and K28 connectivities (Fig. S4). These shifts were readily corrected, as seen in the above overlays, by readjusting the pH value back to 3.31.

tions was also observed for all the other resonances of Aβ12-28, including those of the aromatic hydrogens, which suggests the lack of interaction between the two molecules or the existence of transient interactions not detectable within the timescale of NMR experiments.

The possible involvement of Aβ12-28 amide protons in H-bonds was ascertained by measuring the corresponding chemical shift thermal coefficients ($\Delta\delta/\Delta$T) over the range of temperature (6.7-12°C) in 2D TOCSY spectra. Table 4.1 lists the amide thermal coefficients with respect to carnosine presence and no difference is evident. The rather large $\Delta\delta/\Delta$T values, on average 7.4±1.4 and 7.6±1.3 ppb/degree, without and with cosolute, respectively, are consistent with the lack of stable H-bonds. The lowest $\Delta\delta/\Delta$T values, observed at segment S26-N27, could be interpreted to reflect a local trend towards H-bond formation, somehow reduced by carnosine. The Hα chemical

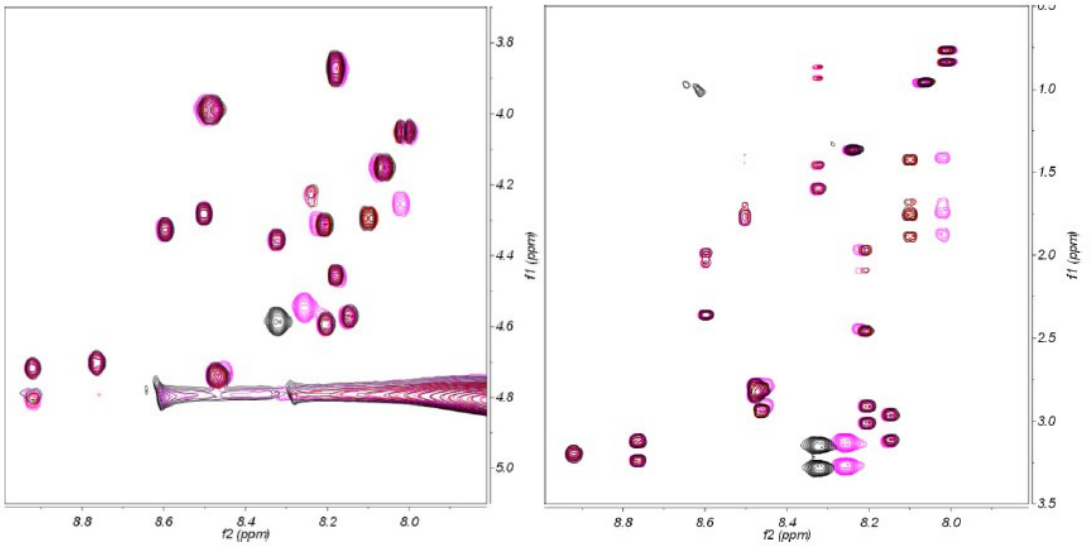


Fig. 4.9 ^{1}H2D TOCSY regions of 1.0 mM Aβ12-28 in H2O/d6 -DMSO 80/20 with 50 mM d5 -glycine at 21°C, in the absence, at pH 3.35 (red contours), and in the presence of 3.4 mM carnosine, at pH 3.65 obtained upon addition (magenta contours), and pH 3.31 obtained after readjustment (black contours). - The contour plots show the amide connectivities to the αCHs (and βCH2 for S26) , on the left, and the propagation to the other resonances of the spin systems, on the right, with the assignments indicated by the amino acid residue single letter code. Assignments are given in Fig. S3. The excess of carnosine was able to raise the pH and induce shifts of E22, D23, K28 and carnosine histidine connectivities.

shifts, however, including those of S26 and N27, are very close to the values observed for statistically disordered peptides,(193) except for A21 and E22 where upfield shifts by 0.10 and 0.15 ppm were measured. The H^{α} chemical shift deviations ($\Delta\delta H^{\alpha}$) from the standard values for random structures are diagnostic for secondary structure regions when not below 0.1 ppm and when occurring as consecutive groups. Based on the specific evidence from NOESY data (see Figure-Fig. 4.11), it is possible to attribute a structural meaning to the upfield deviations detected for A21 and E22 in terms of bending into a local α-like conformation that introduces a loose turn geometry. It seems likely that the loose turn arrangement involving the stretch V18-D23 previously reported,(161) has extended to L17-V24, in our case (in aqueous phosphate at low temperature and neutral pH).

Table 4.1 Thermal coefficients ($\Delta\delta$) in parts-per-billion, ppb, per degree) of Aβ12-28 amide hydrogen chemical shifts in the absence and presence of Carnosine, - measured over the range 6.7 12°C. The experimental error for all values is ±0.9 ppb/degree. Aβ12-28 0.46 mM in 92:8 HR2RO:DR2RO, 16 mM phosphate and 4 mM NaOH, at pH 7.36. On addition of 0.95 mM carnosine, pH increased to 7.48. For the unresolved H13/H14/D23 and Q15/K16 amide cross-peaks (Fig. 4.10), single $\Delta\delta$ values were determined.

Aβ12-28	**Aβ12-28**	**Aβ12-28 + Carnosine**
Residue	**$\Delta\delta$ (ppd/degree)**	**$\Delta\delta$ (ppd/degree)**
H13/H14	-6.9	-7.2
Q15/K16	-8.0	-8.9
L17	-9.1	-9.8
V18	-8.8	-8.8
F19	-9.4	-9.3
F20	-7.3	-7.4
A21	-6.6	-6.9
E22	-6.2	-6.5
D23	-6.9	-7.2
V24	-9.0	-8.8
G25	-7.0	-7.3
S26	-5.8	-6.1
N27	-4.8	-5.5
K28	-6.8	-6.8

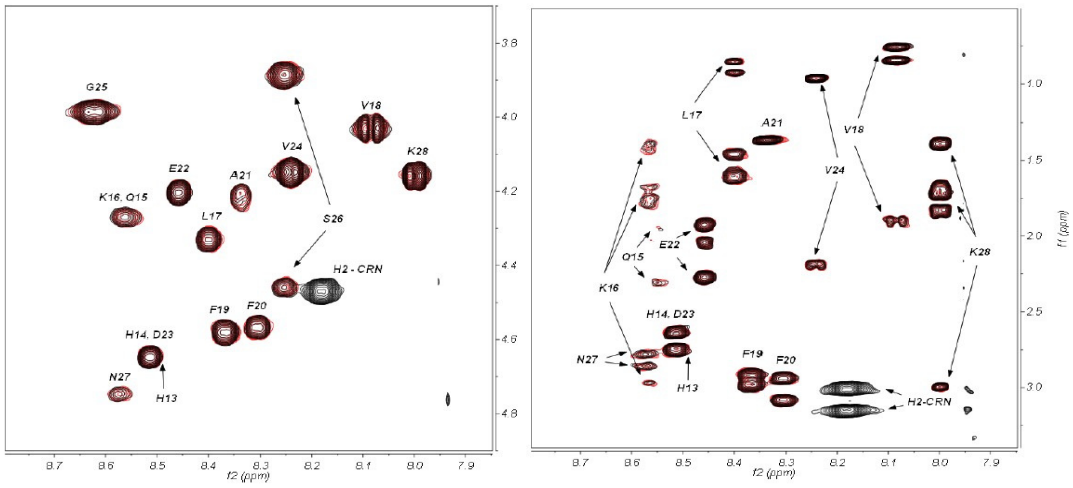


Fig. 4.10 ^{1}H2D TOCSY regions of 0.46 mM Aβ12-28 in H2O/D2O 92/8 with 4mM NaOH and 16 mM phosphate, pH 7.36, at 6.7°C, in the absence (red contours) and in the presence (black contours) of 0.95 mM carnosine (CRN). - The contour plots show, on the left panel, the amide connectivities to the αCHs (and βCH2 for S26) and, on the right panel, the propagation to the other resonances of the spin systems, with the assignments indicated by the amino acid residue single letter code. Under the mentioned conditions, the carnosine excess raised the pH to 7.48, but no shift was detected for E22, D23 and K28 connectivities.

4.2.2 Results from DOSY experiments

DOSY experiments were performed on a buffered sample (0.93 mM Aβ12-28 in H2O/d6 -DMSO 80/20 with 20 mM d5-glycine) in the presence and absence of an excess of carnosine. The relative diffusion coefficients (Figure-Fig. 4.12) are consistent with Aβ12-28 hydrodynamic radii of 1.17 and 1.13 nm, respectively, in the absence and presence of carnosine, at 2°C. According to previously reported estimates(190, 191) the hydrodynamic radii for monomeric and dimeric Aβ12-28 (calculated in water/D2O pH 5.0, at 25°C) should be 1.05 and 1.45nm respectively corresponding to a monomer-dimer ratio of 70/30 (in D2O at pH 5.0 and 25°C). According to the calculated hydrodynamic radii from our data there seems to be an increase in the monomer-dimer ratio from 70/30 to 80/20 (calculated according to a procedure reported elsewhere[61), due to the addition of carnosine. This could be due to the nonspecific cosolute effect on the monomer/dimer association equilibrium brought about by carnosine addition to the

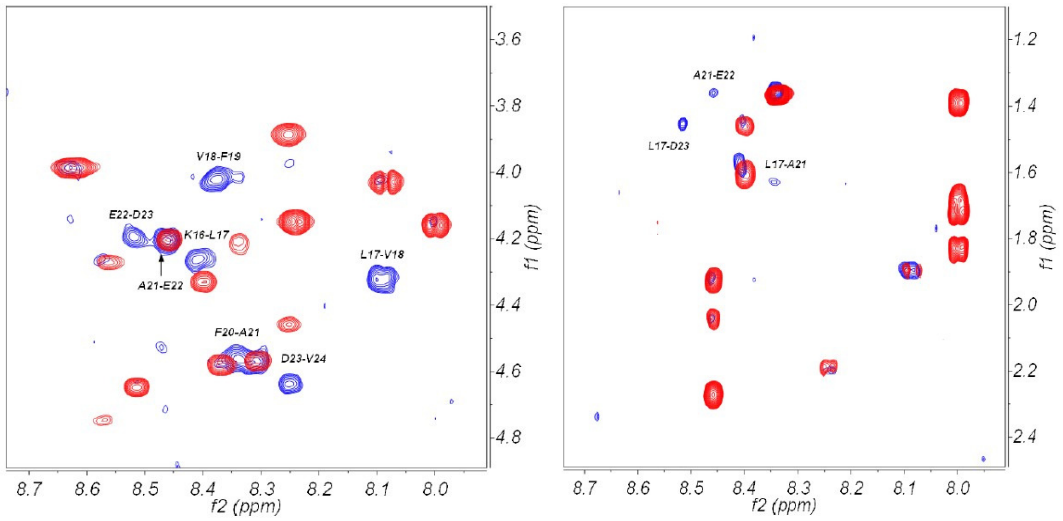


Fig. 4.11 Overlay of 2D TOCSY (red contours) and 2D NOESY (blue contours) regions of 0.46 mM Aβ12-28 in H2O/D2O 92/8 with 4mM NaOH and 16 mM phosphate, pH 7.36, at 6.7°C. - The intense Hα-HN conectivities are indicated in the left panel, whereas the right panel shows Hβ-HN connectivities. It is worth noting that all these connectivities occur within the segment L17-V24.

sample.

The results of the DOSY measurements (Figure-Fig. 4.13) in aqueous phosphate are reported in Table 4.2 in terms of hydrodynamic radii (Rh) of Aβ12-28 and carnosine (194). The results again confirm that Aβ12-28 is predominatly monomeric in aqueous phosphate at neutral pH and low temperature. The apparent Rh value of carnosine increases in the presence of Aβ12-28 because of the transient fast interaction. Although more marginally, also the Aβ12-28 diffusion coefficient and hence the apparent Rh value, is affected by the fast intermolecular interaction. Sample aging does not alter significantly the monomeric state of Aβ12-28, at least over three-four weeks, independent of the presence or absence of carnosine. In fact, the Rh values are constant. The monomeric state of Aβ12-28 appears to resist also temperature. No aggregation due to hydrophobic interactions can be inferred from the Rh values at higher temperatures. Rather, a decrease of those values is observed which could be because of the fast conformational dynamics resulting in changes of the average hydrodynamic dimensions of Aβ12-28. The dimerization processes inferred from in DMSO/water mentioned above do not appear to occur in aqueous phosphate at neutral pH, at least based on the

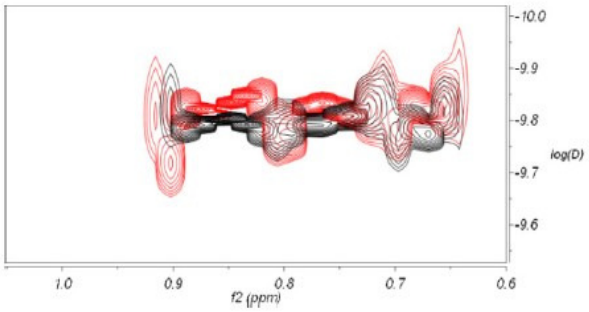


Fig. 4.12 1H 2D DOSY region of 0.93 mM Aβ12-28 in H2O/d6 -DMSO 80/20 with 20 mM d5-glycine at 21°C, in the absence (red contours) and in the presence (black contours) of 2.7 mM carnosine. - The ordinate reports the logarithm of diffusion coefficient.

observable NMR signals (Compare Figures Fig. 4.12 and Fig. 4.13. However, failure to detect dimers or other low-oligomeric species does not necessarily rule out their occurrence at low concentration along with highly-aggregated NMR-invisible states.

The failure to observe all of the amide resonances in the spectra acquired at T $\geq$ 25°C because of the fast solvent exchange expected above neutral pH is consistent with a statistical conformational averaging. This suggests that only at low temperature the aqueous phosphate at neutral pH supports a conformation similar to that observed in 20% DMSO, at moderately acidic pH and 21°C. The thermal stabilization observed in the latter solvent may arise from the association equilibria of Aβ12-28, primarily from dimerization. Surely dimerization is related to the side-chain charge states, that change at the experimental pH values, and unrelated to the N- and C- termini charge states, that do not change instead. The dimerization processes inferred in DMSO/water do not appear to occur in aqueous phosphate at neutral pH, at least based on the observable NMR signals. However, failure to detect dimers or other low-oligomeric species does not necessarily rule out their occurrence at low concentration along with highly-aggregated NMR-invisible states also in aqueous phosphate at neutral pH, similar to the NMR-invisible species that were shown to occur in solution at pH $\geq$5 (**?**). The highly-aggregated NMR-invisible states may be undergoing dissociation upon carnosine addition as is observed by NMR in water/DMSO.

Table 4.2 Hydrodynamic radii (R_h) of Aβ12-28 and carnosine measured as a function of temperature and sample age. Aβ12-28 0.46 mM in 92:8 H2O:D2O, 16 mM phosphate and 4 mM NaOH, at pH 7.36. When 0.95 mM carnosine was present, pH increased to 7.48. The listed R_h values were determined from the average diffusion coefficients determined by single exponential fitting of different Aβ12-28 methyl signals and the β-alanine methylene signals of carnosine. The reported R_h uncertainties are the standard deviations calculated from the dispersion of the diffusion coefficients. The values at 25 and 30°C were obtained from the diffusion coefficients estimated directly from the corresponding DOSY maps. In this case uncertainties are expected to be 5-10%, at most, as the R_h of TSP, i.e. the NMR chemical shift internal reference, proved to be 0.29 nm at both temperatures therefore ensuring the reliability of the determinations.

Temperature°C	**R_hAβ12-28(nm)**	**R_hCarnosine(nm)**	Sample age
6.7	1.04±0.09	-	1 day
6.7	1.04±0.09	0.38±0.02	1 day
6.7	-	0.34±0.01	1 day
6.7	0.97±0.05	-	7 days
6.7	0.98±0.06	0.40±0.03	7 days
6.7	0.99±0.03	-	20 days
6.7	1.01±0.03	0.40±0.03	20 days
12	0.96±0.01	-	21 days
12	0.99±0.02	0.37±0.01	21 days
25	0.92	-	28 days
25	0.92	0.37	28 days
30	0.93	-	28 days
30	0.93	0.37	28 days

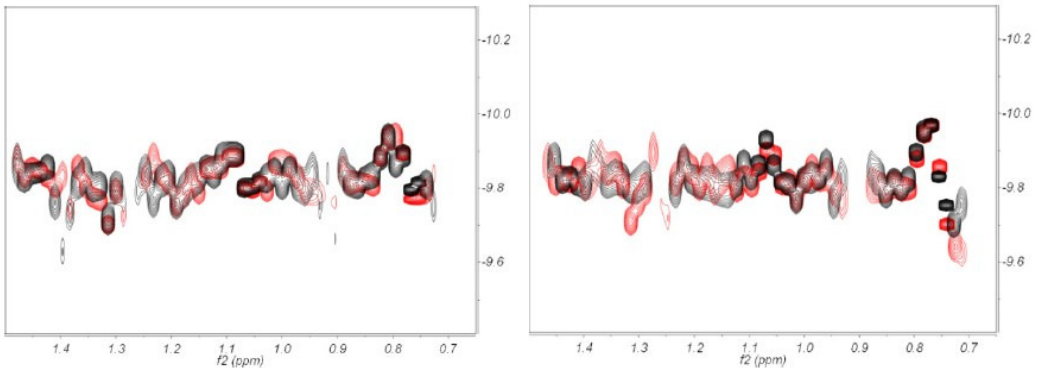


Fig. 4.13 1H 2D DOSY region of 0.46 mM Aβ12-28 in H2O/D2O 92/8 with 4mM NaOH and 16 mM phosphate, at 6.7°C, in the absence, pH 7.36 (red contours) and in the presence, pH 7.48 (black contours) of 0.95 mM carnosine. - The ordinate reports the logarithm of diffusion coefficient. Left: the sample at 1 day from preparation. Right: the same sample at 20 days from preparation.

4.3 L-Carnosine has no effect on the unfolding free energy landscape of β2microglobulin

Deuterium exchange was monitored by acquiring 1H,15N-SOFAST-HMQC experiments on each sample subjected to a temperature ramp from 293.15K to 313.13K over a time period of twenty nine hours. Fig. 4.14 shows the spectral plots of β2m with the exchange occurring over time during the temperature ramp. As more and more residues exchange, the corresponding peaks appear in the HMQC spectra. It is clear that there is no degradation of the sample after undergoing the thermal treatment.

The time evolution curves of the signals (S3, as described in materials and methods section), are plotted as decays. Fig. 4.15, Fig. 4.16 and Fig. 4.17 show the measured decays for the slowly exchanging residues of β2m samples in the blank and in the presence of carnosine at two different concentrations (b2m:carnosine (1:1) and β2m:carnosine (1:100)) respectively, fitted with the function (grey continuous line) N (t) = a[0]·exp(a[1]·exp(a[2]·t)) and fitted according to the selected thermodynamic model (red dashed). The exchange rates (k_{ex}) were calculated according to equations 3.13 and 3.14.

The data were analyzed in two ways, one by assuming a single EX2 process and second by dissecting the global and local processes. A program written in C was used for the first analysis. In this case the ΔG of unfolding of the well resolved amide residues were extracted by assuming a single EX2 model and is plotted in Fig. 4.18. No significant decrease in ΔG is noticed in the samples with carnosine when compared with the blank. Grouping the highest values of ΔG in the blank and the two samples leads to the identification of four residues Fig. 4.19 viz., Cys25, Tyr26, Cys80 and Arg81 (Cys80 couldn't be traced in the 1:100 sample as it was not well resolved). The average ΔG calculated for this group in blank is 8.96$\pm$0.42 kcal/mol, for the 1:1 sample is 8.86$\pm$0.46 kcal/mol and that for the 1:100 sample is 9.02$\pm$0.44. kcal/mol at 298K. This ΔG value is a measure of the protein stability with respect to global unfolding. The rise in this value from 8.86kcal/mol (1:1 sample) to 9.02kcal/mol (1:100) on increasing the concentration of carnosine with respect to β2m may suggest a slight stabilization of the protein brought about by carnsoine although this increment may not be significant enough to account for the 100 times increase in carnosine concentration. The fact that the ΔG of the blank is 0.1kcal/mol more than the 1:1 sample could be explained by

the fact that the batch of the protein used for the blank was different and hence the slightly different experimental conditions.

The decay analysis using the method described in the paper by Enrico et al (submitted) involves a clustering and selection step of the above mentioned decays to identify the local and global opening processes. Accordingly, amide sites Fig. 4.19. Cys25, Tyr26 and Arg81 have been identified as those exchanging through global unfolding with EXX kinetics in the two β2m-carnosine samples and Cys25, Arg81, Leu40 and Ala79 in the blank. The results of the analysis for the three samples are summarized in Fig. 4.20. The ΔG of unfolding in this case for the blanks is 7.60$\pm$0.099 kcal/mol, for the 1:1 sample is 8.21$\pm$0.14 kcal/mol and for the 1:100 sample is 8.23$\pm$0.13 kcal/mol. The comparison of the results between the blank and samples with carnosine may suggest a slight stabilization of the protein brought about by carnosine, however the increasing carnosine concentration in the two samples doesn't seem to have an effect on the unfolding free energy.

A summary of the partial unfolding parameters are given in Fig. 4.21. The residues 10, 28, 80, 82 and 83 in blank; 80 and 83 in the 1:1 sample and 28, 40 and 83 in the 1:100 sample show increase in their ΔG_{NP} values than their corresponding ΔG of global unfolding. The data for residue 80 shows some error because of its unresolved signal but in the case of the other residues, it may be reasoned to be arising because of the destabilization of the partially opened states that do not obey the typical EX2 behaviour but rather contributes to HX with an EX1 limit. As mentioned in the methods section, the pH of the 1:1 sample is 6.88, that of the 1:100 is 6.99 and that of the blank is 7.00. The relative k_{rc} changes with the slight increase in pH may be thought to explain for the relative increase in the ΔG_{NP} values of these residues that exchange with an EX1 limit. It may be postulated that these pH sensitive residues are exposed by a correlated sub-global opening leading to the disruption of the protein structure with the exception of the central core formed by the residues that exchange through global unfolding. The equilibria between the different global and sub-global unfolded states may vary with the pH or even higher concentrations of the osmolyte and could play an important role in determining the protein stability or its propensity to convert to amyloid fibrils.

The residues exhibiting the highest ΔG values can be thought of as structural determinants of the whole protein stability since their HX process would be exclusively

driven by global unfolding. The residues Fig. 4.19. identified for β2m both in the presence and absence of carnosine are Cys25, Tyr26, Arg81, and its neighbouring residues Cys80 (where ever it could be traced) and Ala79 and Leu40 (in the blank where the data was treated with the more rigorous approach by selecting and clustering the decays). In any case, these residues are the ones located in β strands B and F that encompass the disulphide-linked cysteines at positions 25 and 80 and to the adjacent C strand (Leu40) that altogether form the innermost structural core of the protein. Taken together, these results show that the β2m unfolding landscapes in the presence and absence of carnosine prove essentially constant. The presence of carnosine at both high (ratio of protein-ligand 1:100) and low (ratio of protein-ligand 1:1) concentrations do not cause any significant thermodynamic effects in the protein-osmolyte mixture. Overall, the results suggest that carnosine has no particular stabilizing effect on β2m.

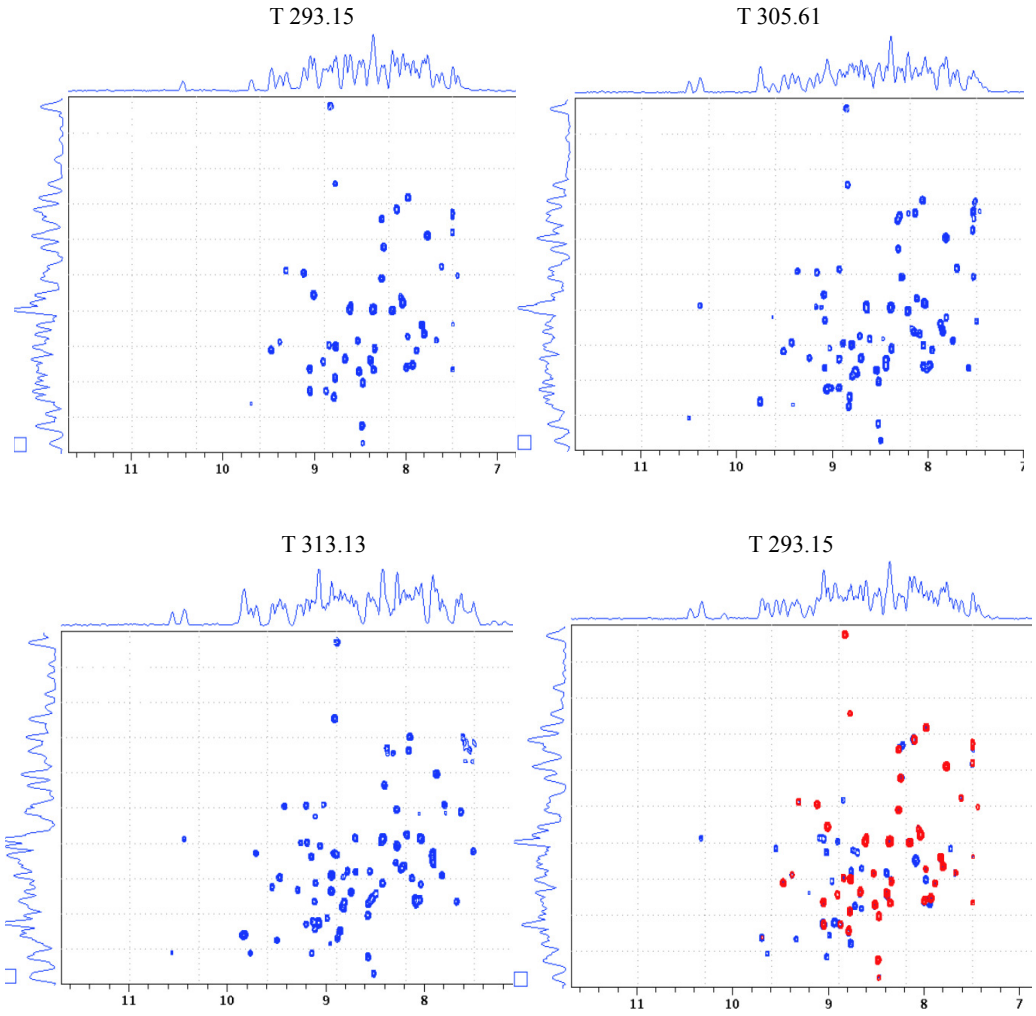


Fig. 4.14 1H,15N-SOFAST-HMQC spectra of β2m in the presence of carnosine (protein ligand ratio 1:1) at different temperatures (given at the top of each panel) during different time periods of the temperature ramp - The first spectrum was obtained at the start, the second at the mid and the third at the end of the temperature ramp. The last inlet shows the overlay of the spectra acquired at 293.15K at the start (red) and the end (blue) of the temperaute ramp. Red peaks are the fast exchanging residues, the rate of which could not be tracked by the exchange experiment protocol while the slowly exchanging residues (blue) that showed a gradual increase in intensity over time and without any peak-overlaps were used for the experimental analysis

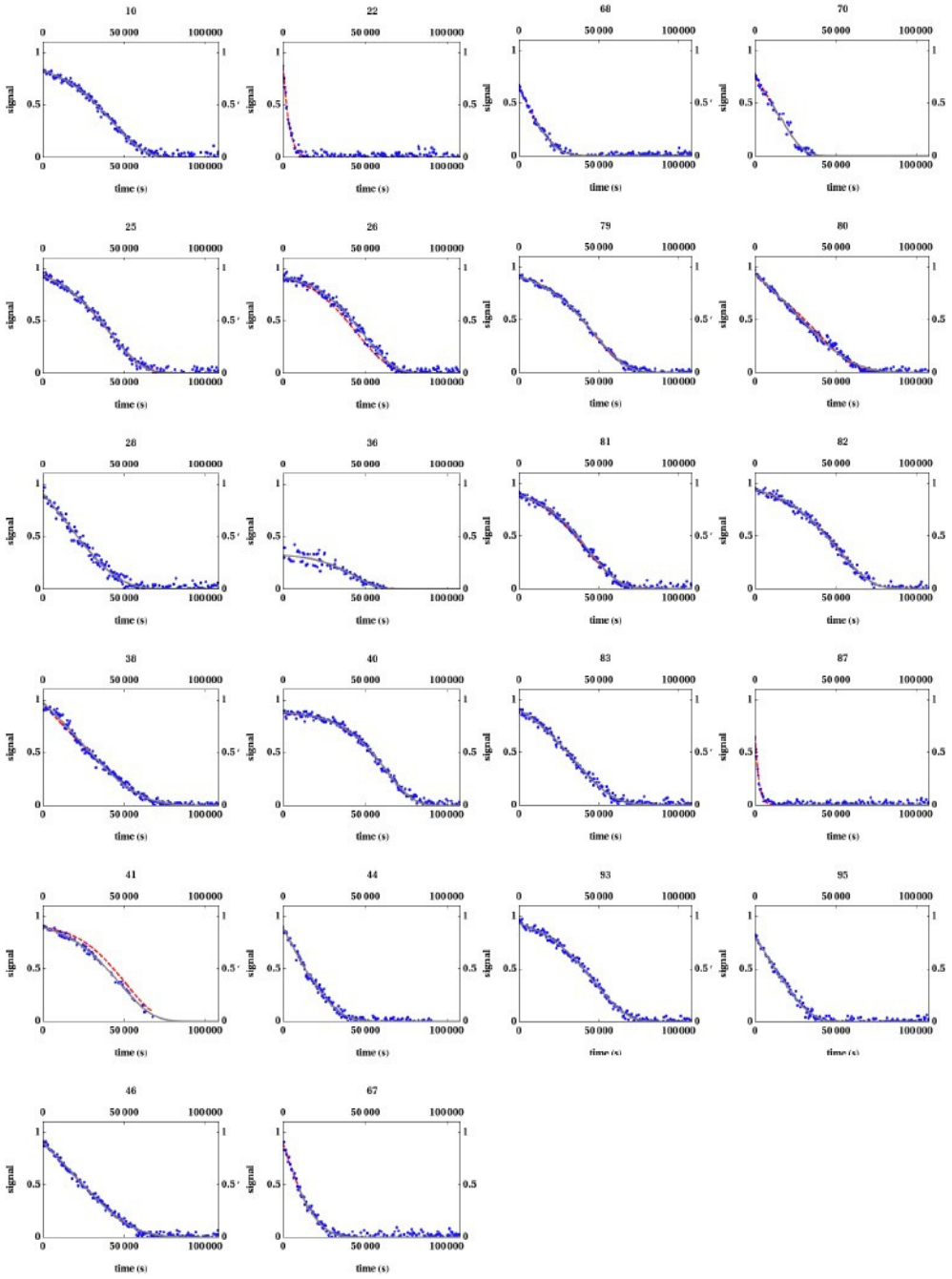


Fig. 4.15 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in phosphate buffer (final pH 7.00), together with the fit according to Eq. 3.14 (red lines) and the fit according to the chosen thermodynamic model (grey lines).

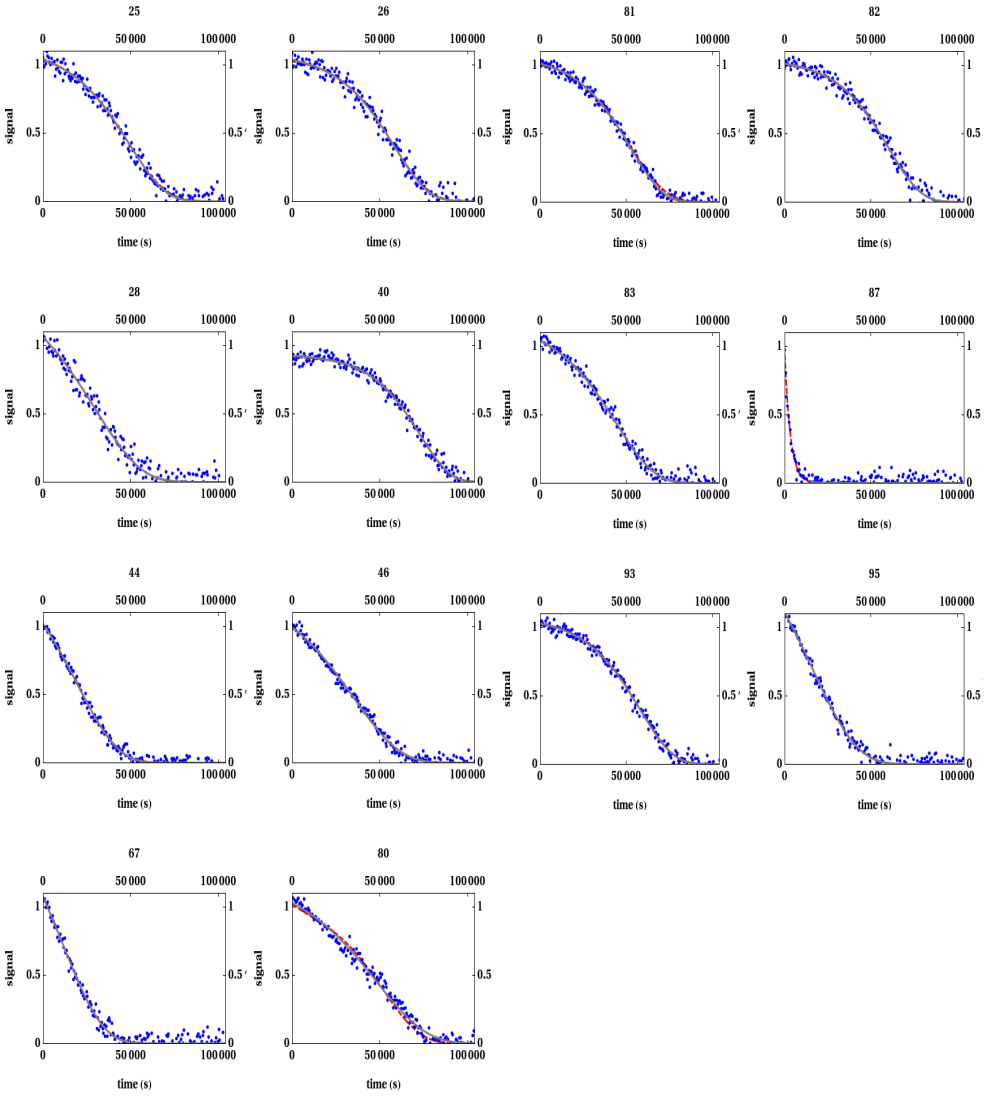


Fig. 4.16 Peak decays (blue circles) - for the slowly exchanging residues (residue numbers at top of each panel) of β2m in the presence of carnosine (at 1:1 ratio) in phosphate buffer (final pH 6.88), together with the fit according to Eq. 3.14 (red lines) and the fit according to the chosen thermodynamic model (grey lines)

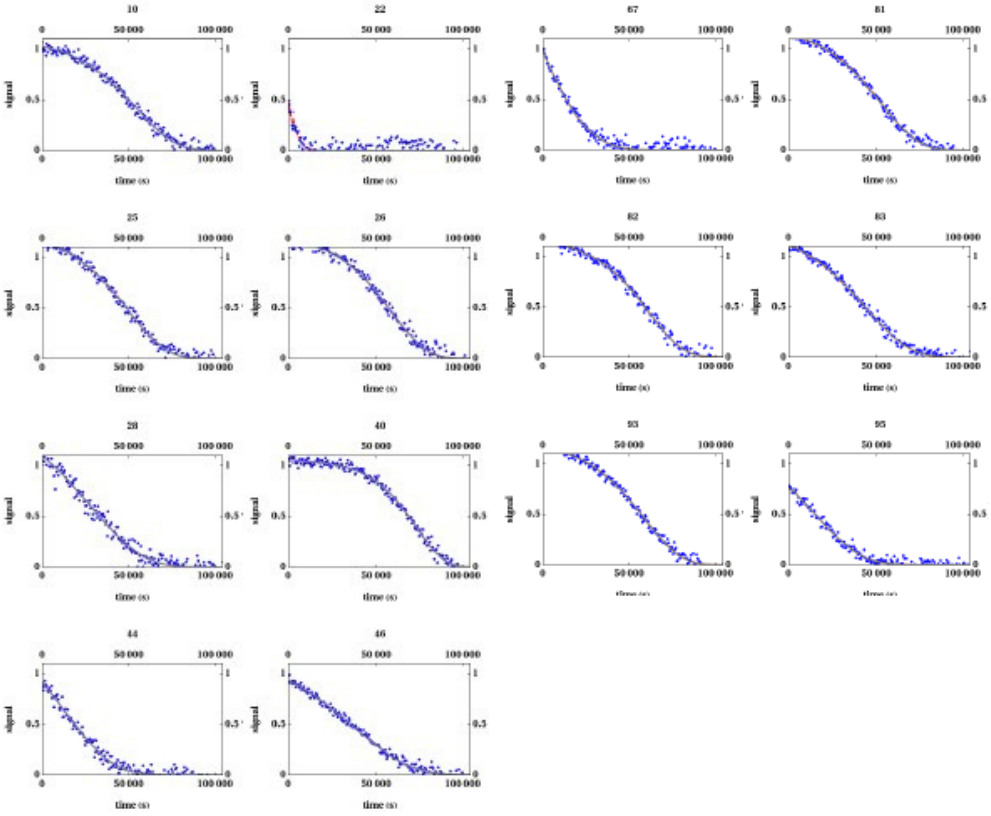


Fig. 4.17 Peak decays (blue circles) - for the slowly exchanging residues (residue numbers at top of each panel) of β2m in the presence of carnosine (at 1:100 ratio) in phosphate buffer (final pH 6.99), together with the fit according to Eq. 3.14 (red lines) and the fit according to the chosen thermodynamic model (grey lines)

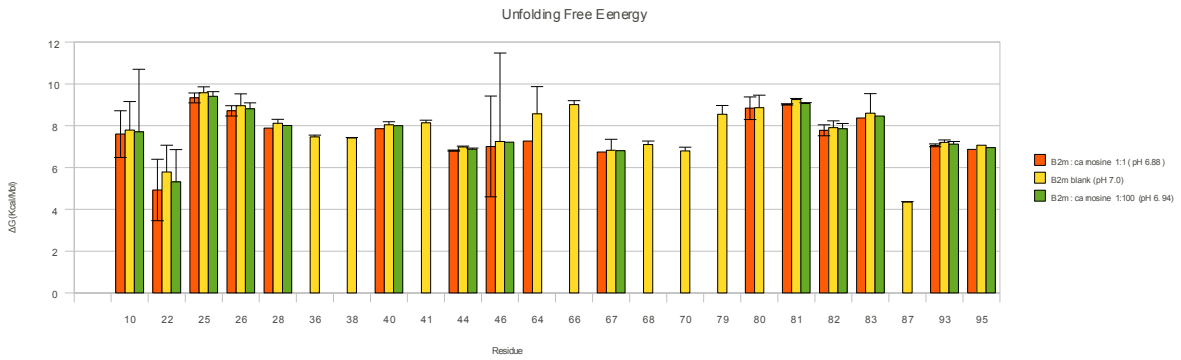


Fig. 4.18 Gibbs free energy values (kcal/mol) at 298 K for the backbone opening equilibrium (ΔG_{op}) - at specific β2m residues in the presence (orange-1:1 protein:osmolyte ratio, yellow-blank, green-1:100 protein:osmolyte ratio), as obtained after determining the individual $k_{ex}T$ functions from intensity data. Numerical values and errors (reported on the bars), are listed in the following page

Residue	β2m:Carnosine (1:1) ΔG$_{ND}$ (kcal/mol)			ΔH$_{ND}$ (kcal/mol)			TΔS$_{ND}$ (kcal/mol)			chi^2/N*pts*
10	7.6	+/-	0.01	25.05	+/-	0.83	17.45	+/-	0.83	17.44
22	4.93	+/-	1.12	101.96	+/-	97.32	97.03	+/-	98.2	3.53
25	9.33	+/-	1.47	27.65	+/-	119.44	18.32	+/-	119.92	1
26	8.71	+/-	0.24	29.2	+/-	19.85	20.49	+/-	19.93	0.95
40	7.85	+/-	0.03	35.86	+/-	2.41	28.01	+/-	2.42	0.39
44	6.81	+/-	0.07	21.66	+/-	6.02	14.85	+/-	6.05	0.71
46	7.01	+/-	0.03	17.19	+/-	2.45	10.18	+/-	2.46	1.14
64	7.27	+/-	2.41	49.72	+/-	105.7	42.45	+/-	103.52	1.65
67	6.74	+/-	0.12	23.28	+/-	10	16.54	+/-	10.05	1.46
80	8.84	+/-	0.53	14.57	+/-	41.78	5.72	+/-	41.9	1
81	9.02	+/-	0.54	31.2	+/-	44.08	22.18	+/-	44.26	0.56
82	7.78	+/-	0.03	31.67	+/-	2.53	23.89	+/-	2.54	1.49
83	8.37	+/-	0.26	23.6	+/-	20.94	15.22	+/-	21.03	0.95
93	7.06	+/-	0.01	29.77	+/-	1.01	22.71	+/-	1.01	0.77
95	6.87	+/-	0.07	17.24	+/-	5.58	10.36	+/-	5.61	1.21
Residue	**β2m:Carnosine (1:100) ΔG$_{ND}$ (kcal/mol)**			**ΔH$_{ND}$ (kcal/mol)**			**TΔS$_{ND}$ (kcal/mol)**			**chi^2/N*pts***
10	7.71	+/-	0.06	23.69	+/-	5.14	15.99	+/-	5.16	0.64
22	5.32	+/-	2.99	70.86	+/-	252.58	65.54	+/-	254.57	2.14
25	9.41	+/-	1.54	23.34	+/-	124.75	13.93	+/-	125.25	0.44
26	8.81	+/-	0.22	27.85	+/-	17.55	19.05	+/-	17.62	1.22
28	8.01	+/-	0.28	12.37	+/-	22.42	4.36	+/-	22.51	1.57
40	8	+/-	0.03	38.95	+/-	2.58	30.95	+/-	2.59	0.41
44	6.88	+/-	0.1	14.23	+/-	8.24	7.35	+/-	8.28	1.27
46	7.21	+/-	0.05	10.78	+/-	3.76	3.57	+/-	3.77	0.86
81	9.06	+/-	0.58	24.05	+/-	47.23	14.98	+/-	47.42	0.58
82	7.85	+/-	0.04	28.45	+/-	2.88	20.61	+/-	2.89	1.08
83	8.46	+/-	0.26	18.53	+/-	21.21	10.07	+/-	21.29	0.93
93	7.12	+/-	0.01	25.56	+/-	0.95	18.44	+/-	0.96	0.83
95	6.95	+/-	0.12	14.44	+/-	9.59	7.49	+/-	9.63	1.05
Residue	**β2m Blank ΔG$_{ND}$ (kcal/mol)**			**ΔH$_{ND}$ (kcal/mol)**			**TΔS$_{ND}$ (kcal/mol)**			**chi^2/N*pts***
10	7.79	+/-	0.09	28.13	+/-	7.31	20.34	+/-	7.3	0.61
22	5.79	+/-	1.36	85.57	+/-	148.13	79.78	+/-	148.54	1.51
25	9.58	+/-	1.28	27.42	+/-	108.12	17.83	+/-	107.97	0.86
26	8.96	+/-	0.28	31.19	+/-	23.49	22.24	+/-	23.45	1.22
28	8.12	+/-	0.56	15.98	+/-	46.85	7.86	+/-	46.77	1
36	7.47	+/-	0.19	43.19	+/-	19.14	35.71	+/-	19.11	0.98
38	7.41	+/-	0.08	8.11	+/-	6.21	0.7	+/-	6.2	0.48
40	8.04	+/-	0.02	36.16	+/-	1.93	28.13	+/-	1.93	0.82
41	8.14	+/-	0.15	26.71	+/-	17.9	18.56	+/-	17.92	0.45
44	6.98	+/-	0.12	18.5	+/-	10.76	11.51	+/-	10.75	0.86
46	7.24	+/-	0.05	9	+/-	3.87	1.76	+/-	3.86	0.82
64	8.57	+/-	47.24	44.19	+/-	1199.38	35.62	+/-	1153.6	2.08
66	9.01	+/-	1.3	48.21	+/-	43.81	39.2	+/-	42.6	0.8
67	6.82	+/-	0.19	19.69	+/-	16.66	12.87	+/-	16.64	1.98
68	7.1	+/-	0.53	26.39	+/-	47.83	19.29	+/-	47.82	0.58
70	6.79	+/-	0.17	41.75	+/-	22.34	34.96	+/-	22.42	1.89
79	8.55	+/-	0.18	27.22	+/-	15.28	18.68	+/-	15.27	0.42
80	8.86	+/-	0.42	7.88	+/-	33.08	-0.98	+/-	33	2.13
81	9.25	+/-	0.6	29.57	+/-	50.65	20.32	+/-	50.58	1.09
82	7.91	+/-	0.05	25.01	+/-	4.06	17.1	+/-	4.05	0.89
83	8.6	+/-	0.32	19.68	+/-	26.79	11.07	+/-	26.74	1.1
87	4.35	+/-	0.93	137.07	+/-	105.94	132.72	+/-	106.38	0.96
93	7.2	+/-	0.02	26.48	+/-	1.3	19.28	+/-	1.3	0.79
95	7.07	+/-	0.12	24.71	+/-	10.56	17.65	+/-	10.55	1.06

Fig. 4.18 Thermodynamic parameters (values ± standard deviations) for opening reactions of β2m at different experimental conditions - estimated from HX data and fitted with basic EX2 model

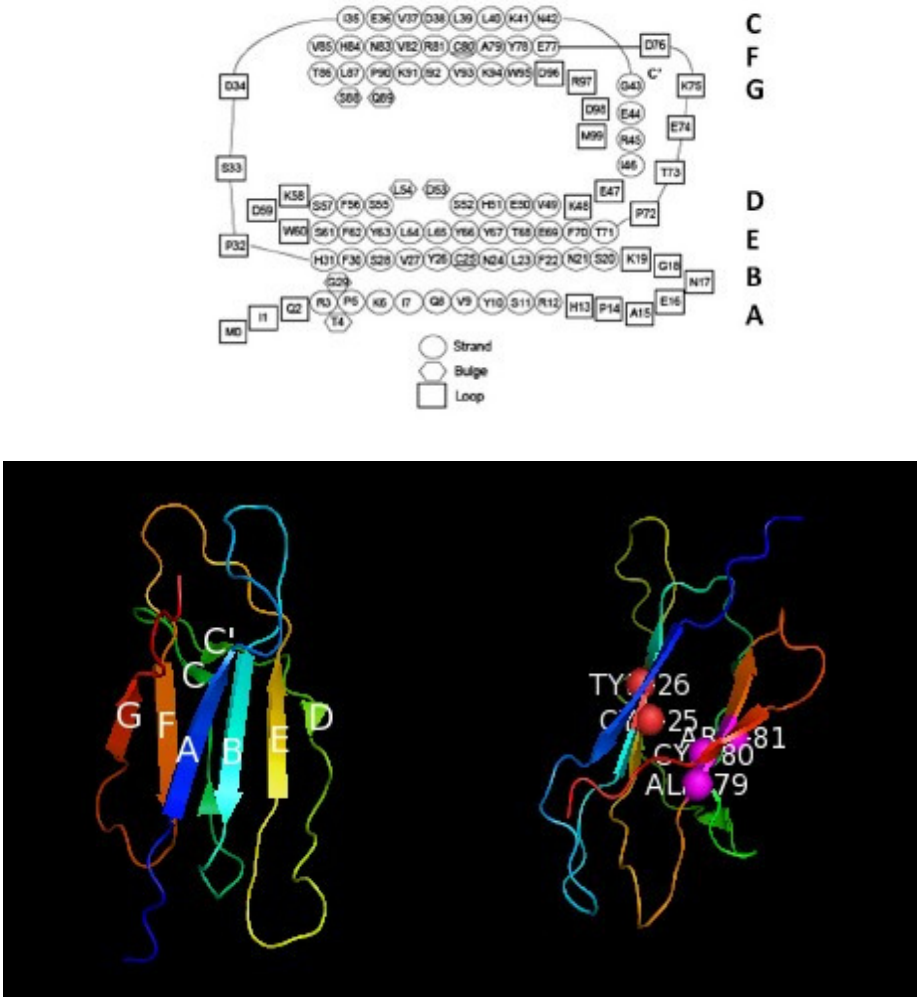


Fig. 4.19 Sequence (top panel) and secondary structure of β2m - The β strands are named in the cartoon representation of the protein on the left whereas the one on the right shows the residues involved in global unfolding process. The cartoon representations were prepared with PyMOL (DeLano Scientific LLC) (PDB code: 1JNJ (195))

β2m Blank

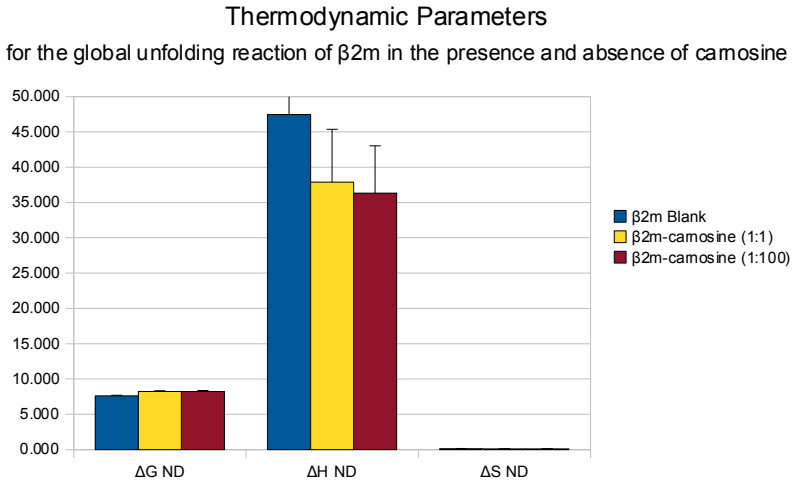
RESULTS	
Analyzed sites	10 22 25 26 28 36 38 40 41 44 46 67 68 70 79 80 81 82 83 87 93 95
Exchange	EXX
Global sites	25 40 79 81
Iperprotected sites	
Local sites	10 22 28 36 38 44 46 67 68 70 80 82 83 87 93 95
Excluded sites	26 41

β2m-Carnosine (1:1)

RESULTS	
Analyzed sites	25 26 28 40 44 46 67 80 81 82 83 87 93 95
Exchange	EXX
Global sites	25 26 81
Iperprotected sites	
Local sites	28 40 44 46 67 80 82 83 87 93 95
Excluded sites	

β2m-Carnosine (1:100)

RESULTS	
Analyzed sites	10 22 25 26 28 40 44 46 67 81 82 83 93 95
Exchange	EXX
Global sites	25 26 81
Iperprotected sites	
Local sites	10 22 28 40 44 46 67 82 83 93 95
Excluded sites	



	β2m Blank	Std. D	β2m-carnosine (1:1)	Std. D	β2m-carnosine (1:100)	Std. D
ΔG ND (kcal/mol)	7.600	0.099	8.210	0.140	8.230	0.140
ΔH ND (kcal/mol)	47.440	3.380	37.870	7.490	36.320	6.710
ΔS ND (kcal/K mol)	0.133	0.011	0.099	0.020	0.094	0.022

Fig. 4.20 A summary of the results obtained by selection and clustering of decays - to identify local and global processes of β2m in the presence and absence of carnosine. The histogram and the table at the bottom reports the thermodynamic parameters and relative standard deviations for the global unfolding reaction of β2m obtained by the exchange studies in the presence and absence of carnosine. Units are kcal/mol for ΔG and ΔH and kcal/Kmol for ΔS

β2m Blank

Site	ΔHNP (kcal/mol)	Std.er (kcal/mol)	ΔSNP (kcal/mol)	Std.er (kcal/mol)	ΔGNP (kcal/mol)	Std.er (kcal/mol)
10	20.8674	6.6127	0.043054	0.0219169	7.9512	0.0619068
22	18.1216	42.7225	0.0399871	0.144594	6.12545	0.66065
28	1.41239	5.8294	-0.0229214	0.0195073	8.2888	0.0622505
36	45.5066	14.2986	0.127276	0.047496	7.3238	0.121258
38	-17.2929	4.31142	-0.0829923	0.0144977	7.60482	0.0642714
44	15.4262	4.43042	0.0285509	0.0148217	6.86092	0.0364191
46	-7.59558	3.72244	-0.0497527	0.0124535	7.33022	0.0431878
67	17.3312	8.24396	0.0355626	0.0277388	6.66238	0.0946193
68	24.3166	10.0796	0.058039	0.0339155	6.90496	0.108326
70	41.7579	10.0118	0.117664	0.0336118	6.45858	0.0946005
80	-70.4782	16.4583	-0.268351	0.0560062	10.0272	3.11909
82	-21.3985	13.9685	-0.0995347	0.0487966	8.46194	3.70282
83	-4.27767	9.55206	-0.0449469	0.0319659	9.20639	0.10576
87	11.9139	69.8545	0.0220683	0.236636	5.29345	1.13879
93	19.1273	2.65157	0.0399935	0.00877878	7.12928	0.032374
95	22.0225	5.23179	0.0503846	0.0175142	6.90712	0.0409606

β2m-Carnosine (1:1)

Site	ΔHNP (kcal/mol)	Std.er (kcal/mol)	ΔSNP (kcal/mol)	Std.er (kcal/mol)	ΔGNP (kcal/mol)	Std.er (kcal/mol)
28	10.5091	5.22872	0.00813616	0.0175071	8.06829	0.0597282
40	38.0084	5.93446	0.0998407	0.0196563	8.05624	0.104902
44	19.4862	3.20782	0.0423328	0.0107451	6.78633	0.032439
46	14.0403	2.51368	0.023335	0.00837187	7.03982	0.0288379
67	20.3091	5.91531	0.0453474	0.019893	6.70482	0.0661595
80	-87.1435	42.0728	-0.327524	9477.77	11.1137	2.84336×10^6
82	30.5641	5.04746	0.0754751	0.016737	7.92162	0.0646854
83	15.8178	5.75025	0.0235041	0.0191601	8.76655	0.060576
87	-15.2462	40.8953	-0.0702018	0.139033	5.8143	0.818868
93	28.875	2.48282	0.0728316	0.00822015	7.02552	0.0318284
95	14.7387	3.80394	0.0261918	0.0127478	6.88114	0.0359155

β2m-Carnosine (1:100)

Site	ΔHNP (kcal/mol)	Std.er (kcal/mol)	ΔSNP (kcal/mol)	Std.er (kcal/mol)	ΔGNP (kcal/mol)	Std.er (kcal/mol)
10	18.9029	4.48899	0.0368719	0.0149148	7.84136	0.0495226
22	17.8097	129.861	0.039567	0.441556	5.93959	2.61272
28	3.27592	5.34529	-0.0165654	0.0178856	8.24553	0.055676
40	44.5283	4.25233	0.120869	0.0139177	8.26762	0.12703
44	11.8656	4.85268	0.0165412	0.0162585	6.90328	0.0473304
46	4.80021	2.86392	-0.00837207	0.00955849	7.31183	0.0390538
67	3.87744	7.91974	-0.00996775	0.0266404	6.86776	0.088523
82	25.841	4.8317	0.0593656	0.0160195	8.03127	0.0654662
83	3.45245	6.12609	-0.0182915	0.020492	8.93991	0.0690902
93	24.2184	2.26712	0.0569984	0.00749719	7.11886	0.0304913
95	11.8574	5.51847	0.0162684	0.0185053	6.97682	0.0522697

Fig. 4.21 A summary of the partial unfolding parameters of the locally exchanging residues - under the different experimental conditions

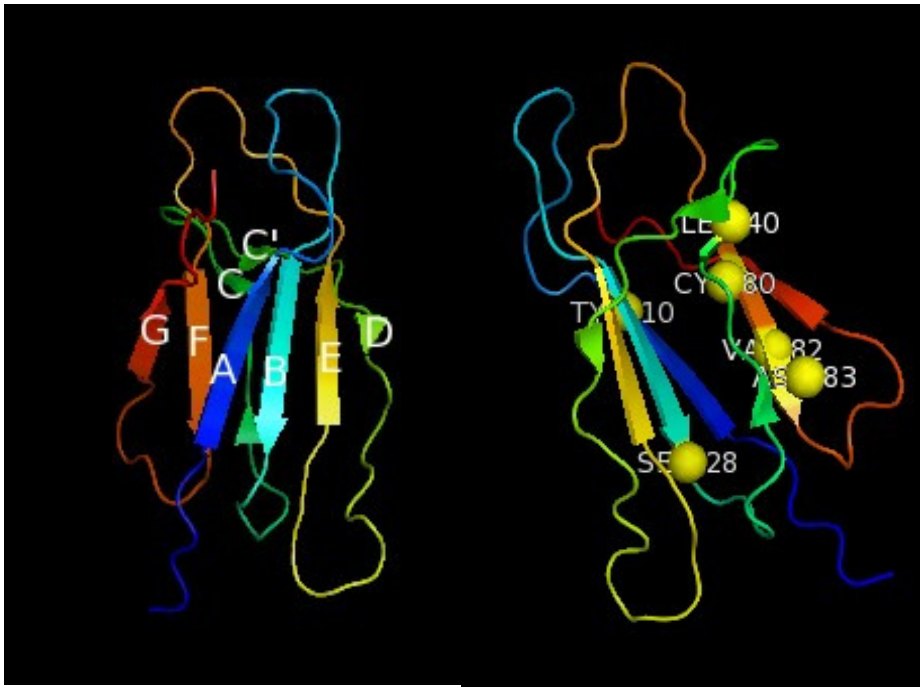


Fig. 4.22 Secondary structure of β2m showing the strands (left) and residues (right) involved in correlated local openings - expected to be exchanging in the EX1 limit. The cartoon representations were prepared with PyMOL (DeLano Scientific LLC) (PDB code: 1JNJ (195))

4.4 Heparan Sulfate (HS) exhibits destabilizing effect on β2m

As described in the previous chapter, deuterium exchange was monitored by acquiring 1H,15N-SOFAST-HMQC experiments on β2m samples in the presence and absence of HS at two different pH and subjected to a temperature ramp from 293.15K to 313.13K over a time period of twenty nine hours. There was no sample degradation or interaction between the protein and ligand observed during the time course of the experiments.

The decay maps depicting the the time-temperature dependence of a group of slowly exchanging amides of β2m along with the relative fitting to extract the $k_{ex}T$ curves according to equations 3.13 and 3.14 are given in Fig. 4.23, Fig. 4.24, Fig. 4.25 and Fig. 4.26. Data were analyzed using a C written routine that directly calculates the thermodynamic parameters starting from the experimental kinetic data in the EX2 limit. The slightly destabilizing environment made it difficult to analyze the data with the second rigorous method (of selecting and clustering the decays to identify the global and local residues, used in the previous section), therefore the results detailed below are from the analysis assuming a single EX2 process.

Fig. 4.27 illustrates the results obtained by the single EX2 model used for a subset of slowly exchanging and well resolved amide signal over the whole experiment. The ΔG of opening (unfolding free energy) for both the samples along with the corresponding blanks are plotted in the histogram and the whole list of data values are given in the following two pages. Grouping of the highest values of the reported ΔG values leads to a selection of β2m residues 25, 26, 28, 66, 79, 80, 81 and 83 in the pH 7.00 sample and residues 24, 25, 26, 66, 80 and 81 in the pH 6.44 sample. These residues exhibiting the highest values of unfolding free energy can be considered to be the structural determinants of the whole protein stability as their HX process is exclusively driven by the global unfolding.The average ΔG calculated for these groups are 8.87$\pm$0.62 kcal/mol for the blank at pH 7.00 and 8.68$\pm$0.92 kcal/mol for the corresponding sample at the same pH; 8.33$\pm$0.33 kcal/mol for the blank at pH 6.44 and 7.79$\pm$0.28 kcal/mol for the corresponding sample at pH 6.25. These ΔG values account for the protein stability with respect to the global unfolding of the protein and in both cases the addition of HS leads to a lowering of the ΔG values implying a destabilization of the protein brought about by HS addition. Also, from the ΔG values and the corresponding histograms

it is clear that the destabilizing effect exerted by HS on β2m is more pronounced at pH 6.25 than at pH 7.0. These results further confirm that HS functions as a destabilizing osmolyte and may in part contribute to the initiation or the acceleration of the amyloidogenesis process of β2m.

It may be noted that, as in the case with carnosine the residues involved in global unfolding of β2m Fig. 4.28 are again the ones located in the β strands B, C, E and F which form the structural core of the protein. Previous studies (47, 48, 49, 50) have shown that HS is a universal component of amyloid and is perhaps playing an important role in amyloid generation by promoting fibrillogenesis. Despite the large body of evidence supporting the importance of HS in amyloidogenesis, little is known about the precise mechanism of interaction between this highly charged GAG and the proteins involved in amyloidogenesis. It is at this juncture that this NMR study on the model protein β2m, proves useful. In accordance with the previous studies, our results confirm the destabilization effect of HS on β2m. Additionally, the data identifies the residues of β2m that are the major contributors of the global unfolding process. It can be postulated that HS exerts its destabilizing effects on β2m through these residues bringing about protein unfolding and perhaps a gradual propensity to convert into amyloid fibrils.

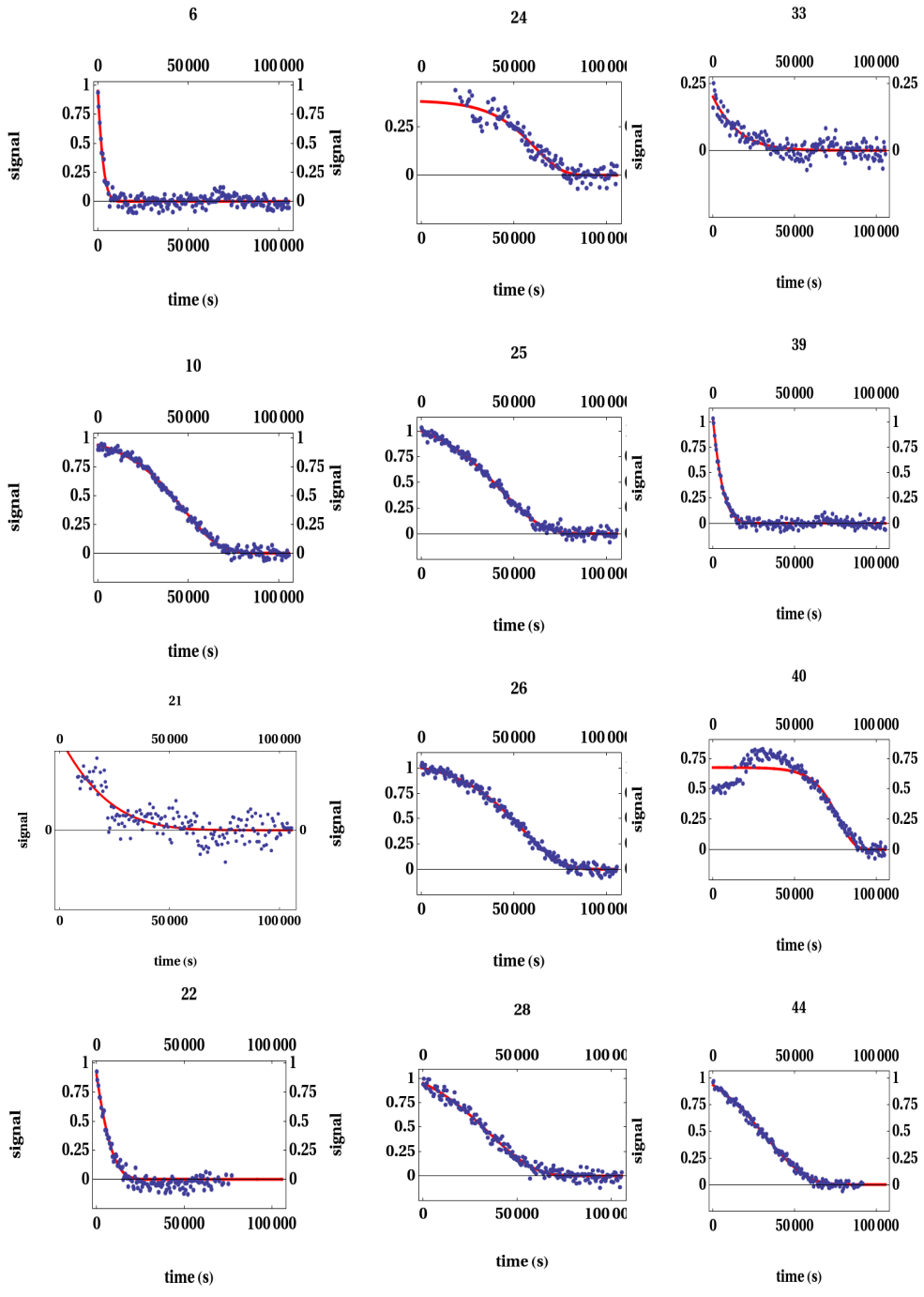


Fig. 4.23 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in phosphate buffer (final pH 6.44), together with the fit according to Eq. 3.14 (red lines).

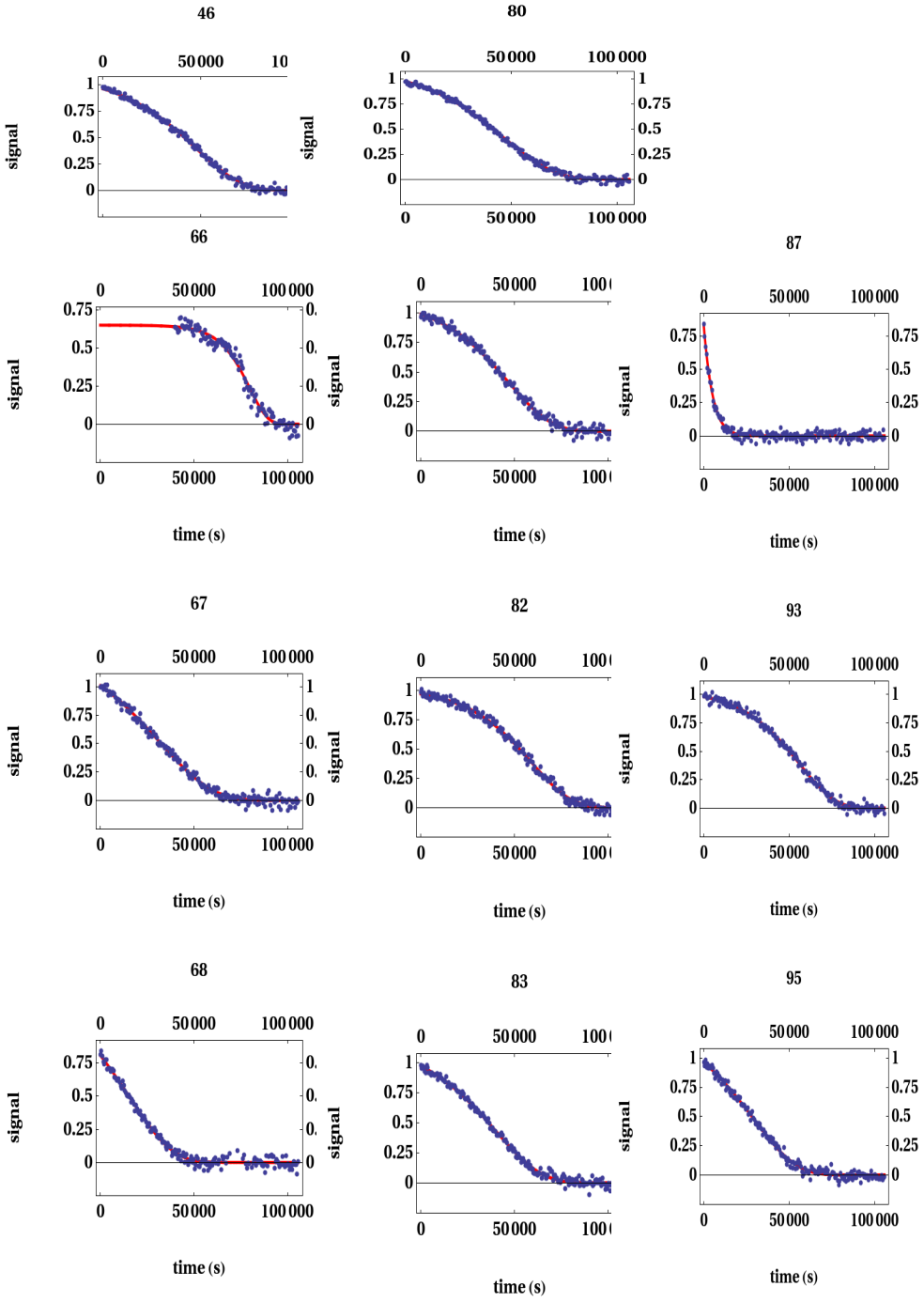


Fig. 4.23 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in phosphate buffer (final pH 6.44), together with the fit according to Eq. 3.14 (red lines).

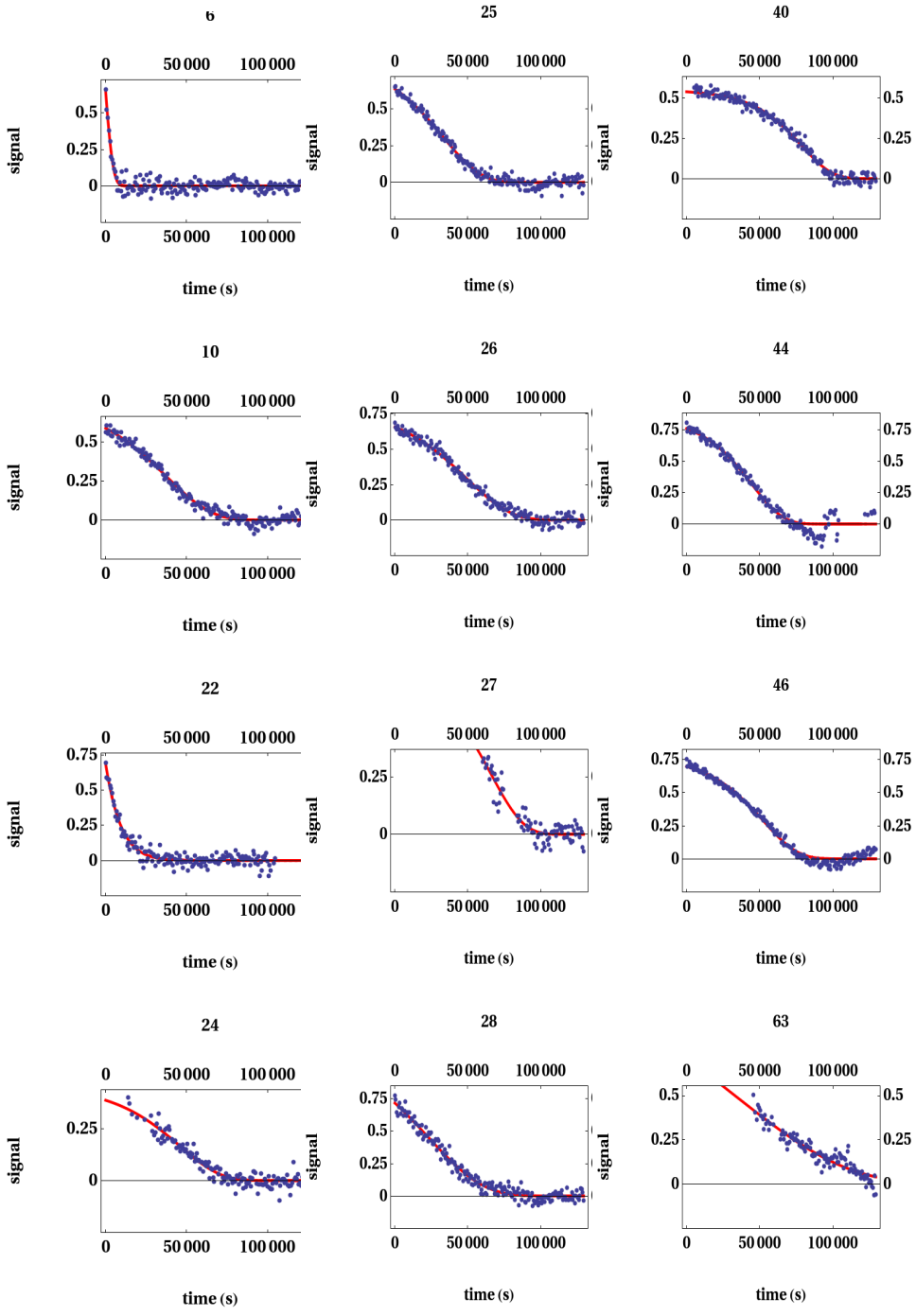


Fig. 4.24 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in the presence of 10%HS in phosphate buffer (final pH 6.25), together with the fit according to Eq. 3.14 (red lines).

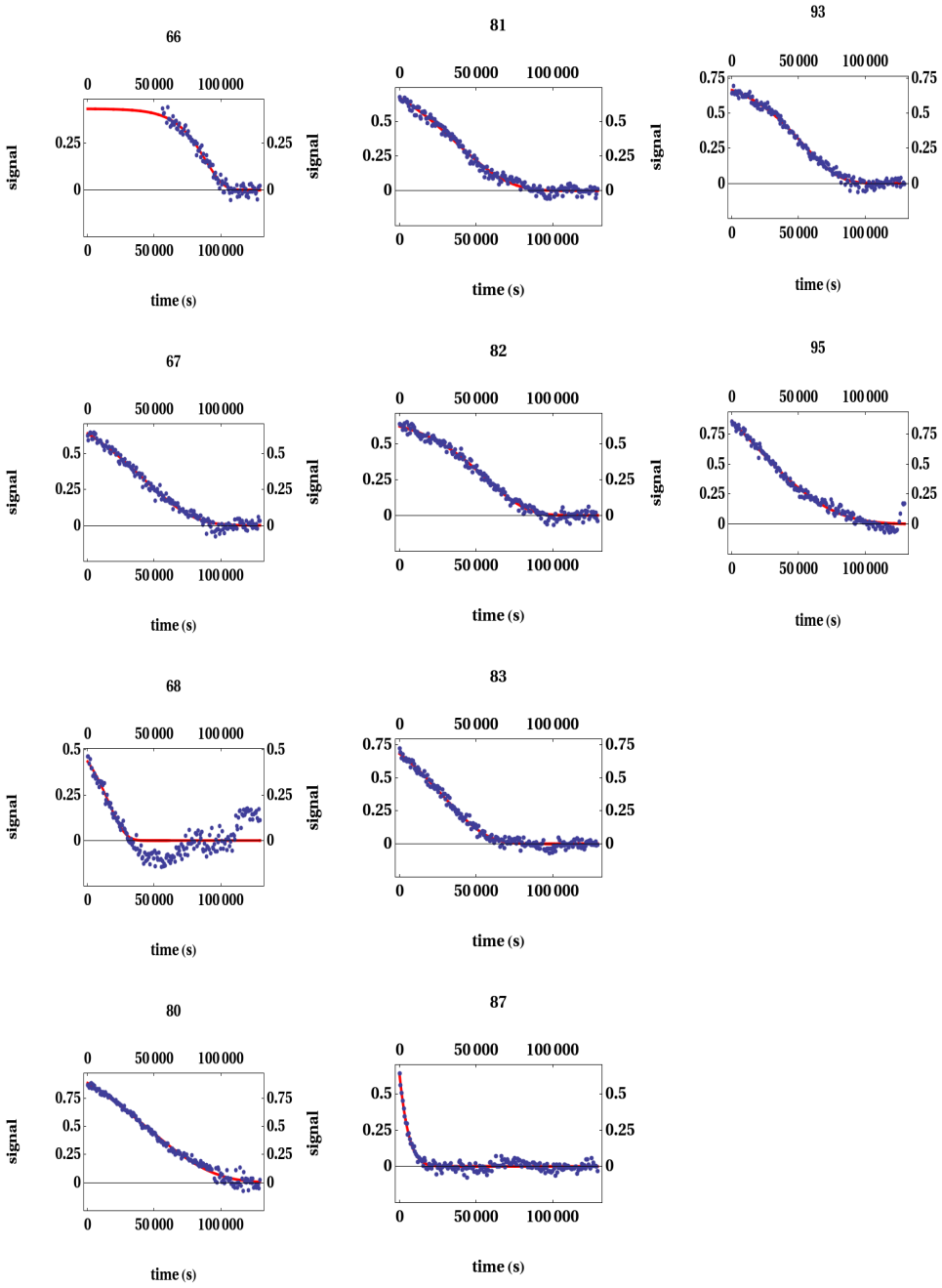


Fig. 4.24 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in the presence of 10%HS in phosphate buffer (final pH 6.25), together with the fit according to Eq. 3.14 (red lines).

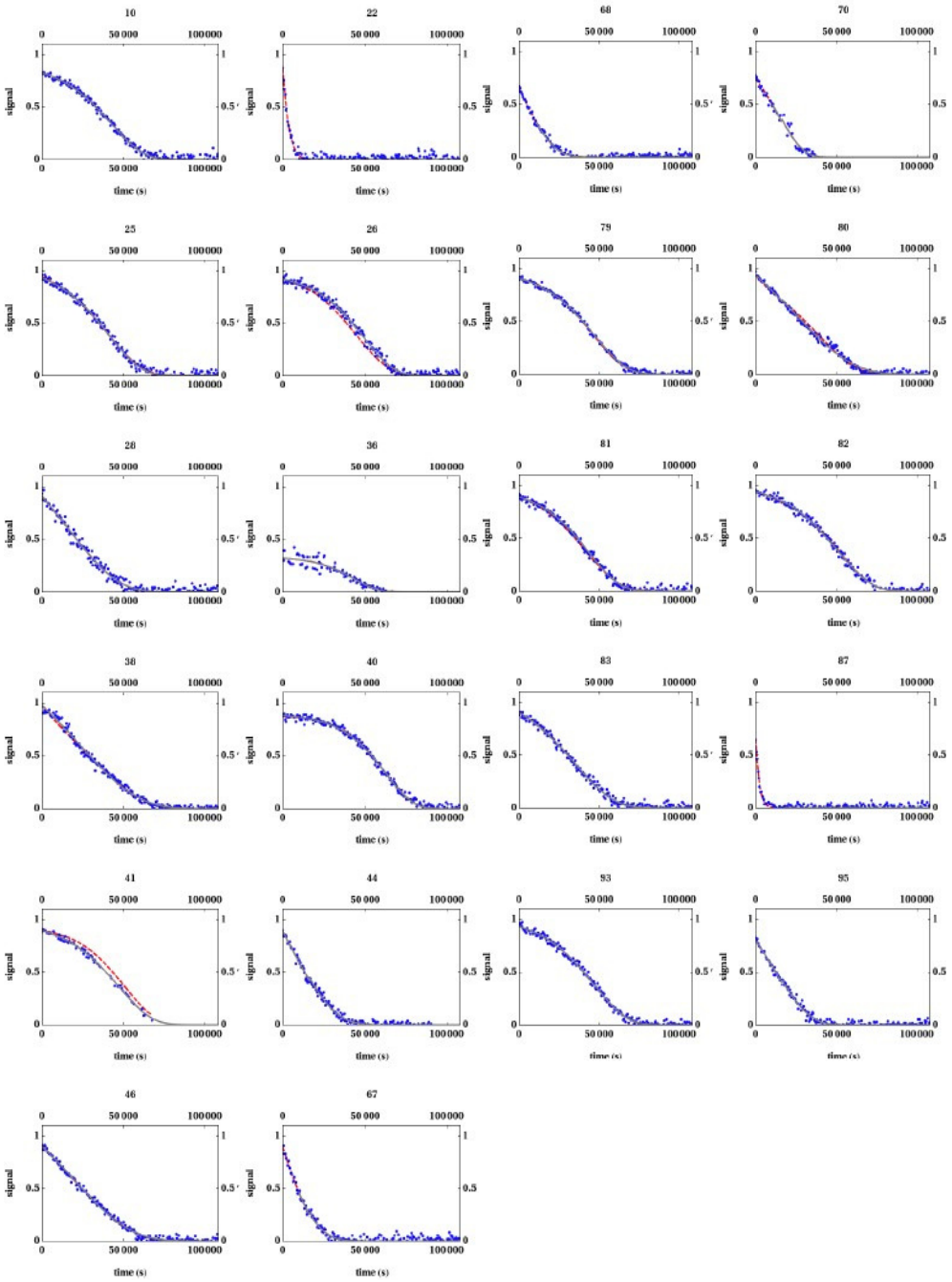


Fig. 4.25 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in phosphate buffer (final pH 7.00), together with the fit according to Eq. 3.14 (red lines).

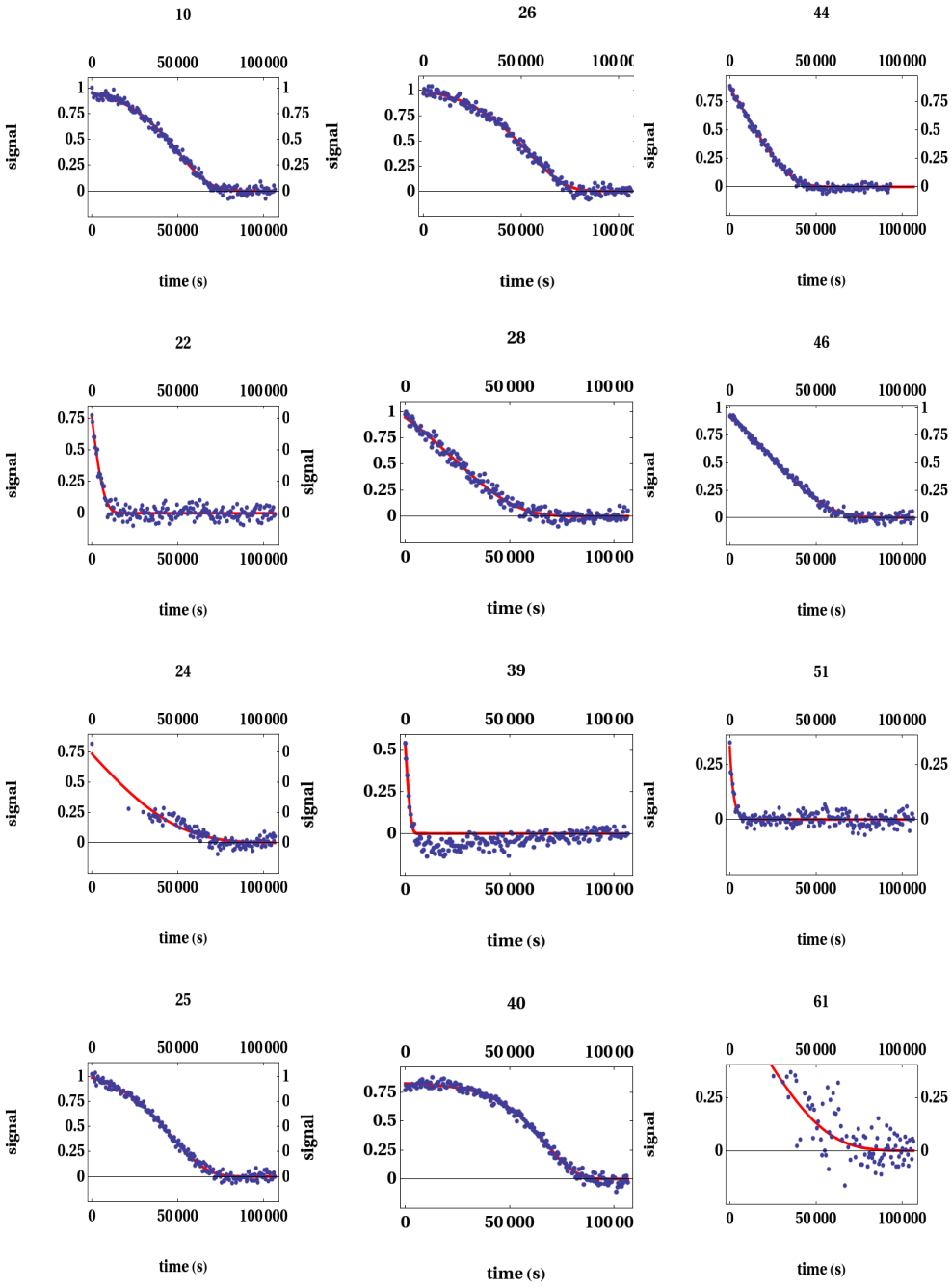


Fig. 4.26 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in the presence of 10%HS in phosphate buffer (final pH 7.00), together with the fit according to Eq. 3.14 (red lines).

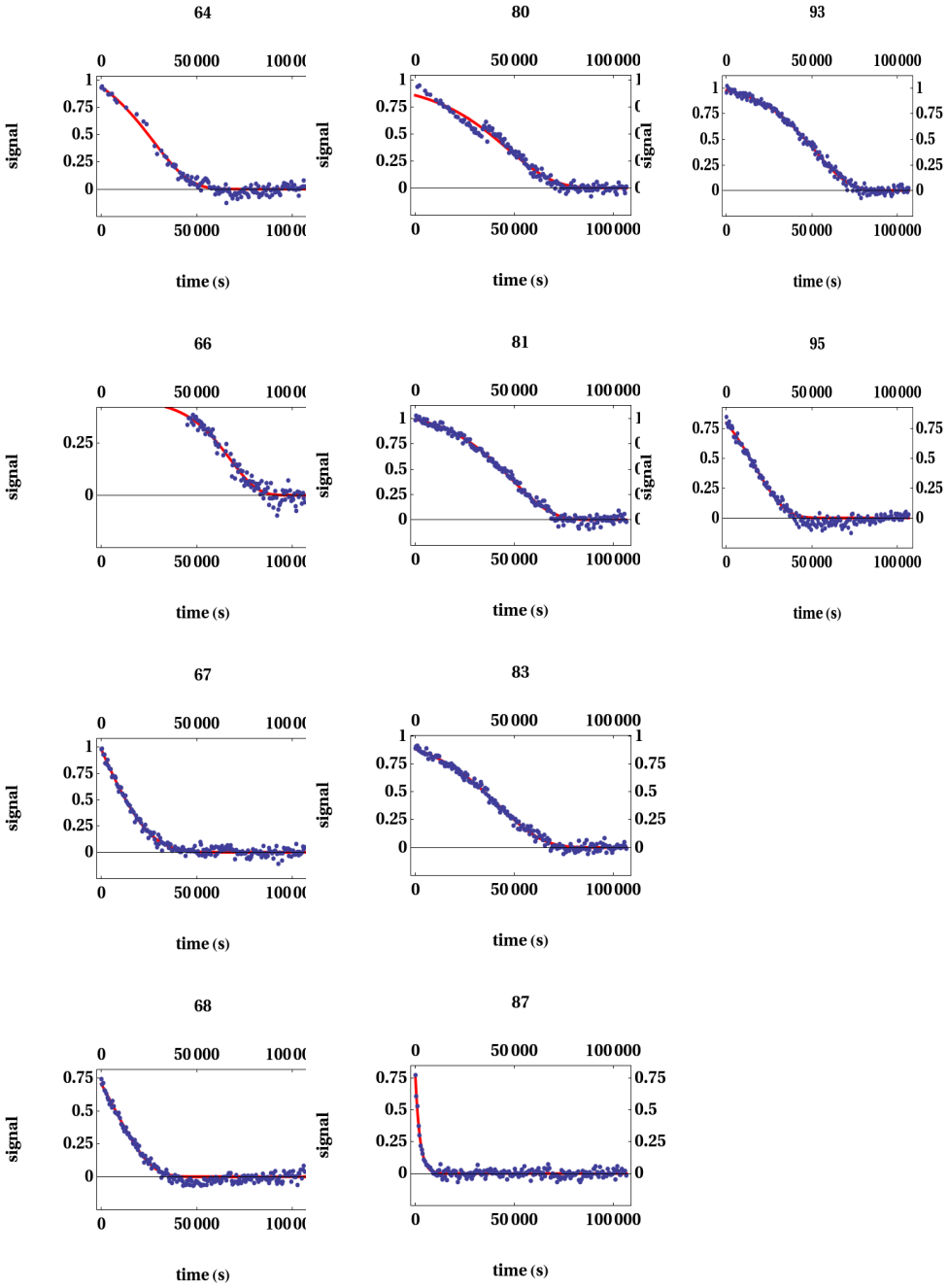


Fig. 4.26 Peak decays (blue circles) - for the slowly exchanging residues of β2m (residue numbers at top of each panel) in the presence of 10%HS in phosphate buffer (final pH 7.00), together with the fit according to Eq. 3.14 (red lines).

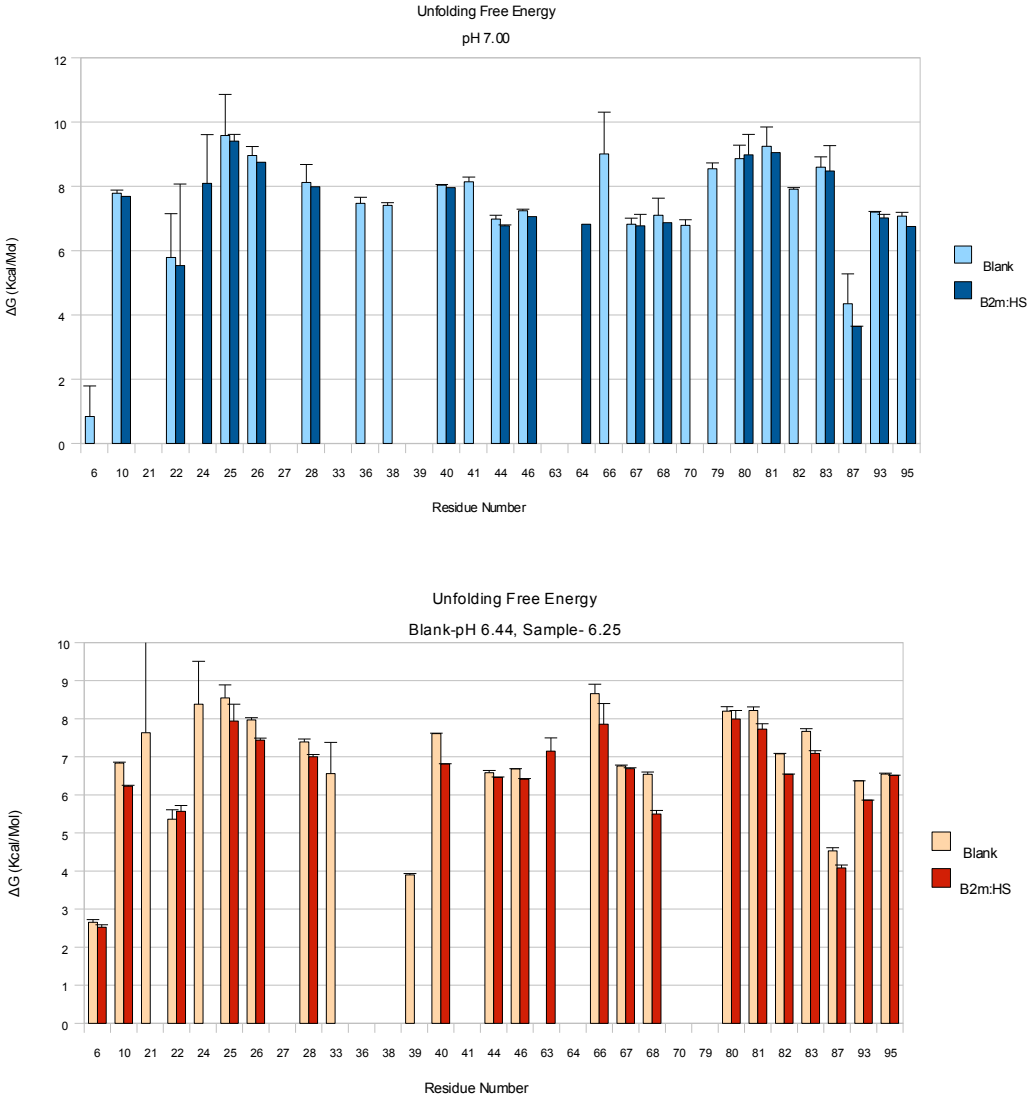


Fig. 4.27 Gibbs free energy values (kcal/mol) at 298 K for the backbone opening equilibrium (ΔG_{op}) and the thermodynamic parameters (values $\pm$ standard deviations) - at specific β2m residues, as obtained after determining the individual k_{ex}T functions from intensity data. Histogram on top shows the plot for the blank and sample at pH 7.00 and the one at the bottom shows the plot of the blank at pH 6.44 and the sample at pH 6.25. Numerical values and errors (reported on the bars), are listed in the following two pages.

Residue	B2m Blank (pH 7.00) ΔG_{ND} (kcal/mol)			ΔH_{ND} (kcal/mol)			$T\Delta S_{ND}$ (kcal/mol)			chi²/N*pts*
6	0.84	+/-	0.95	447.99	+/-	107.73	447.15	+/-	108.48	13.95
10	7.79	+/-	0.09	28.13	+/-	7.31	20.34	+/-	7.3	0.61
22	5.79	+/-	1.36	85.57	+/-	148.13	79.78	+/-	148.54	1.51
25	9.58	+/-	1.28	27.42	+/-	108.12	17.83	+/-	107.97	0.86
26	8.96	+/-	0.28	31.19	+/-	23.49	22.24	+/-	23.45	1.22
28	8.12	+/-	0.56	15.98	+/-	46.85	7.86	+/-	46.77	1
36	7.47	+/-	0.19	43.19	+/-	19.14	35.71	+/-	19.11	0.98
38	7.41	+/-	0.08	8.11	+/-	6.21	0.7	+/-	6.2	0.48
40	8.04	+/-	0.02	36.16	+/-	1.93	28.13	+/-	1.93	0.82
41	8.14	+/-	0.15	26.71	+/-	17.9	18.56	+/-	17.92	0.45
44	6.98	+/-	0.12	18.5	+/-	10.76	11.51	+/-	10.75	0.86
46	7.24	+/-	0.05	9	+/-	3.87	1.76	+/-	3.86	0.82
66	9.01	+/-	1.3	48.21	+/-	43.81	39.2	+/-	42.6	0.8
67	6.82	+/-	0.19	19.69	+/-	16.66	12.87	+/-	16.64	1.98
68	7.1	+/-	0.53	26.39	+/-	47.83	19.29	+/-	47.82	0.58
70	6.79	+/-	0.17	41.75	+/-	22.34	34.96	+/-	22.42	1.89
79	8.55	+/-	0.18	27.22	+/-	15.28	18.68	+/-	15.27	0.42
80	8.86	+/-	0.42	7.88	+/-	33.08	-0.98	+/-	33	2.13
81	9.25	+/-	0.6	29.57	+/-	50.65	20.32	+/-	50.58	1.09
82	7.91	+/-	0.05	25.01	+/-	4.06	17.1	+/-	4.05	0.89
83	8.6	+/-	0.32	19.68	+/-	26.79	11.07	+/-	26.74	1.1
87	4.35	+/-	0.93	137.07	+/-	105.94	132.72	+/-	106.38	0.96
93	7.2	+/-	0.02	26.48	+/-	1.3	19.28	+/-	1.3	0.79
95	7.07	+/-	0.12	24.71	+/-	10.56	17.65	+/-	10.55	1.06
Residue	**B2m-HS (pH 7.00) ΔG_{ND} (kcal/mol)**			**ΔH_{ND} (kcal/mol)**			**$T\Delta S_{ND}$ (kcal/mol)**			**chi²/N*pts***
10	7.69	+/-	0.02	30.71	+/-	1.25	23.02	+/-	1.25	9.75
22	5.54	+/-	1.22	70.6	+/-	103.43	65.07	+/-	104.24	1.88
24	8.09	+/-	2.53	-1.27	+/-	118.57	-9.37	+/-	116.92	1.14
25	9.41	+/-	1.52	26.87	+/-	123.56	17.46	+/-	124.08	0.46
26	8.75	+/-	0.21	30.96	+/-	17.3	22.21	+/-	17.37	1.41
28	7.99	+/-	0.36	14.03	+/-	29.44	6.04	+/-	29.55	1.09
40	7.96	+/-	0.04	43.69	+/-	2.84	35.73	+/-	2.85	0.46
44	6.77	+/-	0.1	21.64	+/-	8.59	14.87	+/-	8.63	0.65
46	7.06	+/-	0.03	12.24	+/-	2.77	5.18	+/-	2.78	1.1
67	6.77	+/-	0.15	13.37	+/-	12.42	6.59	+/-	12.48	1.23
68	6.87	+/-	0.36	29.83	+/-	29.82	22.96	+/-	29.98	0.89
80	8.98	+/-	0.88	21.81	+/-	63.94	12.83	+/-	63.85	1.69
81	9.05	+/-	0.64	27.5	+/-	51.64	18.45	+/-	51.86	0.69
83	8.48	+/-	0.3	23.17	+/-	24.55	14.68	+/-	24.65	0.78
87	3.64	+/-	0.79	124.13	+/-	65.11	120.49	+/-	65.72	0.94
93	7.02	+/-	0.01	28.21	+/-	0.95	21.19	+/-	0.95	1.07
95	6.75	+/-	0.11	33.55	+/-	9.41	26.8	+/-	9.46	1.68

Fig. 4.27 Thermodynamic parameters (values ± standard deviations) for opening reactions of β2m at different experimental conditions - estimated from HX data and fitted with basic EX2 model

Residue	B2m Blank (pH 6.44) ΔG_{ND} (kcal/mol)			ΔH_{ND} (kcal/mol)			$T\Delta S_{ND}$ (kcal/mol)			chi^2/N_{pts}
6	2.65	+/-	0.07	156.62	+/-	5.57	153.97	+/-	5.62	20.06
10	6.84	+/-	0.02	28.56	+/-	1.56	21.71	+/-	1.57	0.4
21	7.63	+/-	3.05	2.05	+/-	244.34	-5.57	+/-	243.65	1.68
22	5.36	+/-	0.25	48.92	+/-	22.49	43.56	+/-	22.66	1.85
24	8.38	+/-	1.13	42.49	+/-	67.07	34.11	+/-	66.27	0.6
25	8.55	+/-	0.34	23.01	+/-	27.44	14.47	+/-	27.55	0.82
26	7.97	+/-	0.06	26.32	+/-	4.82	18.34	+/-	4.84	1.1
28	7.39	+/-	0.08	20.11	+/-	6.5	12.72	+/-	6.52	1.34
33	6.56	+/-	0.82	2.68	+/-	65.07	-3.88	+/-	65.3	0.38
39	3.9	+/-	0.03	61.96	+/-	2.27	58.06	+/-	2.29	1.75
40	7.61	+/-	0.01	42.61	+/-	0.73	35	+/-	0.73	7.37
44	6.58	+/-	0.06	19.04	+/-	4.76	12.46	+/-	4.79	0.06
46	6.68	+/-	0.01	21.13	+/-	0.62	14.45	+/-	0.63	0.62
66	8.66	+/-	0.25	65.5	+/-	10.23	56.84	+/-	10	2.35
67	6.76	+/-	0.02	15.09	+/-	1.96	8.33	+/-	1.97	1.27
68	6.54	+/-	0.06	21.95	+/-	4.99	15.41	+/-	5.02	0.88
80	8.2	+/-	0.12	20.19	+/-	9.66	11.98	+/-	9.7	0.49
81	8.22	+/-	0.09	22.23	+/-	6.98	14.01	+/-	7.01	1.85
82	7.08	+/-	0.01	26.13	+/-	0.71	19.05	+/-	0.71	1
83	7.67	+/-	0.07	19.66	+/-	5.47	11.99	+/-	5.49	0.78
87	4.53	+/-	0.08	49.97	+/-	6.81	45.45	+/-	6.86	1.06
93	6.37	+/-	0	26.38	+/-	0.23	20.01	+/-	0.24	1
95	6.55	+/-	0.02	18.55	+/-	1.34	12	+/-	1.34	1.21
Residue	B2m Blank (pH 6.25) ΔG_{ND} (kcal/mol)			ΔH_{ND} (kcal/mol)			$T\Delta S_{ND}$ (kcal/mol)			chi^2/N_{pts}
6	2.52	+/-	0.07	150.57	+/-	5.62	148.05	+/-	5.67	11.8
10	6.23	+/-	0.02	23.07	+/-	1.45	16.84	+/-	1.46	0.66
22	5.57	+/-	0.15	28.29	+/-	13.2	22.73	+/-	13.28	1.88
25	7.94	+/-	0.44	26.33	+/-	36.61	18.39	+/-	36.78	0.61
26	7.44	+/-	0.05	20.29	+/-	4.31	12.85	+/-	4.32	1.13
28	7	+/-	0.06	12.89	+/-	4.73	5.89	+/-	4.75	1.74
40	6.81	+/-	0.01	37.95	+/-	0.75	31.14	+/-	0.75	0.39
44	6.46	+/-	0.01	35.7	+/-	1.05	29.25	+/-	1.06	4.69
46	6.41	+/-	0.02	31.18	+/-	1.34	24.76	+/-	1.34	0.62
63	7.15	+/-	0.35	4.29	+/-	14.22	-2.86	+/-	13.91	1.52
66	7.86	+/-	0.54	56.39	+/-	19.83	48.53	+/-	19.32	0.39
67	6.69	+/-	0.02	11.88	+/-	1.38	5.19	+/-	1.38	0.85
68	5.5	+/-	0.09	70.95	+/-	7.64	65.46	+/-	7.69	9.06
80	7.99	+/-	0.23	5.68	+/-	18.06	-2.3	+/-	18.11	0.15
81	7.73	+/-	0.14	15.96	+/-	11.49	8.23	+/-	11.53	0.72
82	6.54	+/-	0.01	25.21	+/-	0.74	18.67	+/-	0.74	0.84
83	7.09	+/-	0.07	22.65	+/-	5.68	15.57	+/-	5.7	1.08
87	4.08	+/-	0.08	64.53	+/-	6.55	60.45	+/-	6.6	1.09
93	5.86	+/-	0	24.39	+/-	0.25	18.54	+/-	0.25	0.74
95	6.51	+/-	0.01	1.48	+/-	0.82	-5.03	+/-	0.82	2.79

Fig. 4.27 Thermodynamic parameters (values ± standard deviations) for opening reactions of β2m at different experimental conditions - estimated from HX data and fitted with basic EX2 model

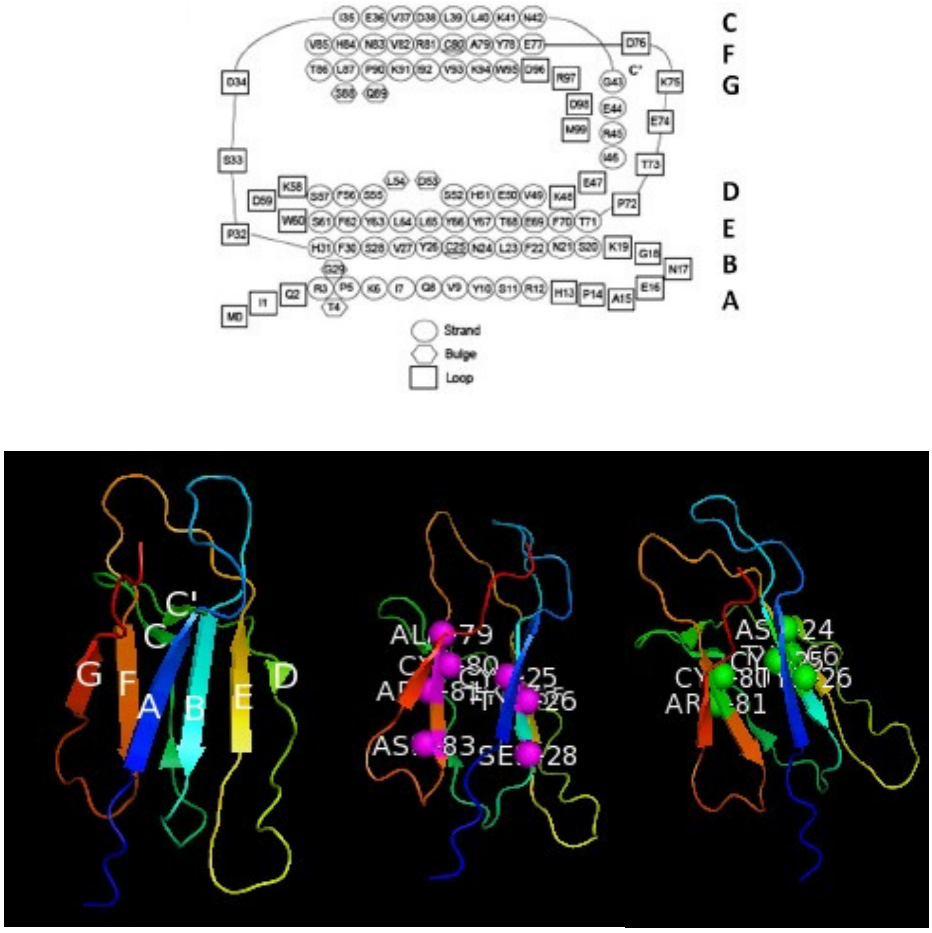


Fig. 4.28 Sequence (top panel) and secondary structure of β2m showing the β strands (bottom left) and residues involved in global unfolding process - in the presence of 10%HS in phosphate buffer at pH 7.00 (bottom middle) and at pH 6.25 (bottom right). The cartoon representations were prepared with PyMOL (DeLano Scientific LLC) (PDB code: 1JNJ (195))

4.5 Osmolytes and amyloid forming proteins

Osmolytes are important intracellular cosolutes that are assumed to have evolved to stabilize proteins against adverse environmental conditions. Several studies have shown the importance of osmolytes in protein folding and functionality. It has been thought that osmolytes exert their action on proteins by affecting the solvent properties of water as related to protein polarity and protein diffusion and that the ability of the osmolytes to stabilize the proteins against denaturing stress originates from the unfavourable interaction of the osmolyte with the peptide backbone. The osmolyte-induced protein folding is a very weak protein:osmolyte interaction and do not occupy any significant fraction of the backbone allowing no real competition for backbone binding sites. Osmolytes within a cell function in a concerted way and the concentrations of different osmolytes in different cells depend merely on the site of action and the cell-types involved. Since osmolytes are universally recognized as important players in protein folding/functionality it is worth thinking that these organic cosolutes may have important roles in regulating/controlling protein misfolding and may be important determinants of the amylodogenic pathway. Therefore, this study was designed to understand the role of osmolytes in stabilizing/destabilizing proteins involved in amyloidoses. This is the first experimental study of this kind and the results discussed here are preliminary and sets a background for further studies on protein-osmolyte mixtures in the context of amyloidogenesis.

We have chosen to study three amyloidogenic systems viz.; the eye-lens protein α-crystallin which upon aggregation and precipitation gives rise to cataract, a fragment from the Amyloidβ-peptide (Aβ) a major constituent of the Alzheimers plaque and β2-microglobulin (β2m) the protein responsible for dialysis related amyloidosis. Our results essentially prove that protein-osmolyte interaction (in cases where it exists) is a weak and transient interactive force but extremely specific to proteins. The same osmolyte may or may not exert the same kind of action on different proteins. The protective effects of osmolytes on proteins depends on its site of action, the protein involved, the intracellular environment (pH playing an important role) and perhaps its co-existence with other osmolytes and cosolutes within a cell.

4.5.1 Is carnosine a protective osmolyte?

Carnosine has been described as an important protective agent in long-lived tissues like the human brain, kidneys, innervated tissues and eye-lens at concentrations as high as 20mM. Its a molecule that has been widely studied in the realm of its antioxidant activities and biological buffering capacity, preventing damage to many biological structures and protecting cells against osmotic shock, oxidative stress or toxic e?ects of metabolites and free radicals. In this thesis work we have studied the type of interaction between carnosine and two proteins viz., α-crystallin and a fragment from the Amyloidβ -peptide. Then we have studied the unfolding landscape of the model protein β2-microglobulin in the presence and absence of carnosine.

Our results from the STD experiments on the αcrystallin-carnosine system show that an efficient yet weak interaction occurs occurs between α-crystallin and carnosine but how these interactions effect the stability of the protein has not yet been clear. Magnetization transfers occurring at high concentrations of the osmolyte (upto a protein:ligand ratio of 1:1000) shows the efficacy of the interaction at ratios that may be normal within cells. Limitations of the method stopped us from tracing the interaction any further than this ratio. Our studies on the control protein βL-crystallin and lysozyme that show no efficient/or little magnetic transfer further strengthens the applicability of this method in elucidating from a semi-quantitative view point, the existence of interactions and points out that the results not due to any random collisions occurring because of the the co-existence of the protein and osmolyte in solution at such high concentrations. Our studies on Amyloidβ-peptide also points out to a weak and transient interaction that may be existing in solution but not as efficient as the one with α-crystallin. Exchange studies were conducted on the β2m system to check for the unfolding free energy landscape of the protein at large and to see the residue-level changes in unfolding free energy that carnosine may bring about in the protein. Our results show no particular effect of carnosine on the protein at concentrations that were 100 times the concentration of the protein in solution. We noticed a slight increase in ΔG of unfolding but this increase was not sufficient to say if carnosine exerts a protective effect on β2m or that it stabilizes the protein. Perhaps, at higher concentrations the effect could be more pronounced but to get this information at residue-level specificity at such high concentrations is difficult with the attempted method.

At this point, we can conclude that carnosine is an osmolyte that acts in a variety of ways on different proteins but our results are not sufficient to state if the effects are always protective. It is however clear that carnosine exerts no harmful effects on the proteins studied and therefore do not exert a destabilizing effect.

4.5.2 Is Heparan Sulfate (HS) a destabilizing osmolyte?

HS is an important molecule for pharmaceutical research, given its ubiquitous presence on cell surfaces and its roles in development, homeostasis and pathological processes. Studies have shown that HS directly interacts with amyloid proteins in vitro and investigations of in vivo amyloid deposits have shown that HS is a co-deposit in all types of amyloids. Despite the large body of evidence supporting the importance of HS in amyloidogenesis, little is known about the precise mechanism of interaction between this highly charged GAG and the proteins involved in amyloidogenesis. Our deuterium-exchange studies on the β2m-HS system shows that HS exerts a destabilizing effect on β2m and this effect is more pronounced at pH 6.25 than at pH 7.00. We have also ellucidated the residues involved in this destabilization process and conclude by stating that the osmolyte HS exerts its destabilizing effects through these residues that may in part contribute to the initiation or the acceleration of the amyloidogenesis process of β2m.

However, as stated earlier the effects of the osmolyte may be specific for one protein or cell type. The results discussed here may not be extended to other proteins. Although the effects of HS on β2m are destabilizing, this result is not enough to categorize HS as a destabilizing osmolyte. More research in this direction involving many other proteins need to be done before tagging HS as a destabilizing osmolyte.

5

Appendix

5.1 Other projects in which significant contributions were made

5.1.1 Structural Polymorphism in the HRAS Promoter and Involvement of G4-DNA in Transcription Regulation

The PhD activity started in the light of working with the project"Structural polymorphism in the HRAS promoter and involvement of G-4 DNA in transcription regulation". The overall goal of the project was to understand the structural features of the human HRAS oncogene and to provide models for developing anti cancer drugs against the same. The results of CD, FRET and PCR stop assay that were already carried out suggested the formation of quadruplex structures at the promoter regions of the two genes (HRAS-1 and HRAS-2) under study. This G-4 DNA seemed to be directly involved in regulating the expression of these genes. Therefore, we attempted to determine the structure of these quadruplex forming sequences by molecular modeling followed by NMR.

Our first step was to create working models for understanding the possible conformations. Modeling programs like the Mc-sym were initially used to extract necessary fragments from already studied models which were then used in a progarm to create the final structures by superimposing the proper residues by calculating the RMSD.However these attempts did not yield expected results for the HRAS-2 sequence which was thought to form a parallel quadruplex structure from the CD experiments.

Therefore a program to extract all possible nucleotide fragments of 1-4 nucleotide length from already published structures/models was written in house. 1-4 residue length nucleotide stretches formed the loops connecting the G- stretches in the quadruplex backbone. These fragments were again used in a program to superimpose these fragments on the G-4 parallel and antiparallel stacks already created. The residues in the loops were then mutated and renumbered to get the final HRAS-1 (antiparallel) and HRAS-2 (parallel) quadruplex structure. These were further subjected to energy minimization by using the NAMD program using CHARMM forcefield.

NMR studies were taken up for the experimental determination of these structures. 2D NMR experiments were carried out and data acquisition and assignment of resonances to determine the conformation present was done first for HRAs-1. We could conclude that HRAS-1 adopted an antiparallel quadruplex. NMR suggested a coexistence of multiple conformers in the sample. This was particularly true for the HRAS-2. The extensive overlap of the NMR signals and the inability to obtain a single stable conformer in solution hindered any further interpretation of NMR data.

In the meanwhile, general protocols for molecular dynamics simulations and docking studies for quadruplex structures were developed and a manuscript was written describing the same. The article is attached in the following pages.

BMC Structural Biology

BioMed Central

Research article

Open Access

Molecular models for intrastrand DNA G-quadruplexes

Federico Fogolari*[1,2], Haritha Haridas[1], Alessandra Corazza[1,2], Paolo Viglino[1,2], Davide Corà[3], Michele Caselle[3], Gennaro Esposito[1,2] and Luigi E Xodo[1]

Address: [1]Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, Piazzale Kolbe 4 - 33100 Udine, Italy, [2]Istituto Nazionale Biostrutture e Biosistemi, Viale Medaglie d'Oro 305, 00136 Roma, Italy and [3]Dipartimento di Fisica Teorica Università di Torino, Via P. Giuria 1 10125 Torino, Italy

Email: Federico Fogolari* - federico.fogolari@uniud.it; Haritha Haridas - haritha.haridas@uniud.it; Alessandra Corazza - alessandra.corazza@uniud.it; Paolo Viglino - paolo.viglino@uniud.it; Davide Corà - cora@to.infn.it; Michele Caselle - caselle@to.infn.it; Gennaro Esposito - gennaro.esposito@uniud.it; Luigi E Xodo - luigi.xodo@uniud.it

* Corresponding author

Published: 7 October 2009

BMC Structural Biology 2009, **9**:64 doi:10.1186/1472-6807-9-64

Received: 8 June 2009
Accepted: 7 October 2009

This article is available from: http://www.biomedcentral.com/1472-6807/9/64

Abstract

Background: Independent surveys of human gene promoter regions have demonstrated an overrepresentation of $G_3X_{n1}G3X_{n2}G_3X_{n3}G_3$ motifs which are known to be capable of forming intrastrand quadruple helix structures. In spite of the widely recognized importance of G-quadruplex structures in gene regulation and growing interest around this unusual DNA structure, there are at present only few such structures available in the Nucleic Acid Database. In the present work we generate by molecular modeling feasible G-quadruplex structures which may be useful for interpretation of experimental data.

Results: We have used all quadruplex DNA structures deposited in the Nucleic Acid Database in order to select a list of fragments entailing a strand of three adjacent G's paired with another strand of three adjacent G's separated by a loop of one to four residues. These fragments were further clustered and representative fragments were finally selected. Further fragments were generated by assemblying the two strands of each fragment with loops from different fragments whenever the anchor G's were superimposable. The fragments were used to assemble G quadruplex based on a superimposability criterion.

Conclusion: Molecular models have been generated for a large number of $G_3X_{n1}G_3X_{n2}G3X_{n3}G_3$ sequences. For a given sequence not all topologies are possible with the available repertoire of fragments due to steric hindrance and low superimposability. Since all molecular models are generated by fragments coming from observed quadruplex structures, molecular models are in principle reliable and may be used for interpretation of experimental data. Some examples of applications are given.

Background

It is generally recognized that in addition to the canonical Watson-Crick double-stranded conformation, DNA can assume a variety of secondary structures including triplex [1-3], cruciform [4], quadruplex [5-7] and Z-DNA [8]. Quadruplex DNA, also called G4-DNA, is stabilized by G-

5. APPENDIX

quartets, planar arrays of four guanines paired by Hoogsteen hydrogen bonding, and monovalent alkali cation, K+ or Na+, located in the central cavity of the structure. G-quartets can stabilize a variety of quadruplex structures which can be intermolecular or intramolecular, in which single-stranded DNA is folded to provide the four strands of the guanine scaffold. In the human genome the sites that can potentially form G4-DNA are estimated to be more than 300.000. They are not randomly distributed, but located preferentially in repetitive genomic sequences such as the telomeres, ribosomal DNA and the immunoglobulin heavy-chain switch regions [7]. Moreover, G-rich sequences have been found with a high frequency in the control regions of proto-oncogenes, either upstream or downstream the transcription start site (TSS) [9]. While the formation of G4-DNA structures in the 5' overhang of the telomeres has the function of reducing the effect of endogenous nucleases and stabilizing the chromosomes, the possible role of G4-DNA in the promoter of proto-oncogenes is still a matter of debate. The observation that some common transcription factors including SP1 (binding site: RGGCGKR), KLF (binding site: GGGGTGGGG), and MAZ (binding site: GGGAGGG), recognize regions composed by runs of guanines, potentially capable to extrude G4-DNA, raises the hypothesis that this unusual structure may be somehow involved in transcription regulation. Hurley and co-workers reported that a G-rich element (-142 to -115 bp) upstream of the major P1 promoter folds into a stable G-quadruplex [10]. As G > A point mutations abrogating the capacity of the promoter to form a quadruplex enhance transcription, while porphyrinic ligands that stabilize G4-DNA reduce transcription, it was concluded that quadruplex DNA should behave as a repressor. Such mechanism has been hypothesized also for other proto-oncogenes including KRAS [11-13], CKIT [14], VEGF [15], CMYB [16], Rb [17] and BCL-2 [18,19]. Nucleic acids structures are difficult to probe in vivo, and the main evidence that G4-DNA exists in cells is that antibodies raised against G-quadruplex DNA label the macronuclei of a ciliate [20]. Furthermore, the observation that several prokaryotic and eukaryotic proteins recognize and bind to quadruplex DNA [21] also supports indirectly that it exists in vivo. Some of these proteins, hnRNP A1 [22], POT-1 [23] and human Werner syndrome helicase [24] have also resolvase activity against this structure.

Given its biological importance, G-quadruplex structures have become target for several drug design studies (see e.g. [6,25-28]). Many efforts have been made to resolve by crystallography or NMR the structure of quadruplex DNA. However, so far a limited number of structures has been resolved, mainly because G-rich sequences at high concentrations tend to assume a variety of inter-molecular and intra-molecular structures. So, molecular modeling can be very helpful to get insight into putative G4-DNA structures formed by biological relevant sites.

In particular, there is a widespread interest in sequences possessing the motif $G_{3+}X_{n1}G_{3+}X_{n2}G_{3+}X_{n3}G_{3+}$, where G_{3+} indicates 3 or more G's and $n1$, $n2$ and $n3$ are numbers greater than one. These sequences have been demonstrated to be able to form intrastrand G-quadruplexes [5,25,29-39].

Structure determination of intrastrand G-quadruplex has been elusive, because of the observed conformational equilibria which are detrimental for both NMR and X-ray crystallographic studies. Indeed, base modifications have been used to stabilize a particular conformation and more in general it has been reported that only one out of several tens of starting G-quadruplex putative sequences are amenable to structural study [5]. To the best of our knowledge there are only thirteen intrastrand G-quadruplex structures solved which do not contain modified bases.

When this figure is compared with the number of potential G-quadruplexes identified around the TSS of genes and involved in gene regulation by independent studies [40-48] the enormous gap between sequence and structure studies is apparent.

Besides the possibility that the same sequence could adopt more conformations, which could prevent structure resolution, the high concentration typically required for structural methods could favor intermolecular assembly over intramolecular formation of G-quadruplexes. Intermolecular G-quadruplexes (dimers or tetramers) are roughly ten times more represented in the Nucleic Acid Database (NDB) [49] or Protein Data Bank (PDB) [50] than intramolecular G-quadruplexes.

However, the 3D structure of nucleic acids can be inferred from sequence and indeed a pipeline of RNA secondary structure prediction and structure reconstruction has been recently shown to predict RNA structures with high accuracy [51-53]. The quality of the putative models relies on the quality of RNA secondary structure prediction.

For G-quadruplexes the complexity of possible topologies and the limited repertoire of structures solved makes this task much more difficult. The MC-Fold and MC-sym prediction pipeline proceeds from a single sequence to a single structural model determined according to restraints derived from structural prediction [51].

In this work we proceed in a different way, i.e. we simply explore what conformations could be assembled by the repertoire of observed fragments in a dataset of quadruplex structures. The rationale behind this study is that the

latter dataset entails the most stable structural features of G-quadruplexes. It is reasonable to expect that a predictive model incorporating features found in this dataset should be stable. We assemble novel quadruplex structures by assembling combinatorially all fragments encoding for strands participating in the G-quadruplex stems and loops connecting two strands of the G-quadruplex. The set of predictive models is instructive in that it highlights those topologies and loop lengths which can be combined to assemble a model together with their frequencies.

The method is inspired by the program MC-Sym [51-53] which, combined with the secondary stucture prediction program MC-Fold was able to accurately predict RNA structure starting from a dataset of fragments. The program assembles the fragments in a hierarchical manner, subject to constraints and retaining all or only the best fragments generated at each step [51]. The program has many options to control the number of fragments kept at each step of the building procedure and is designed to achieve accuracy and efficiency.

No energy or scoring function is used on the contrary here because the constraints imposed by the quadruplex structure are sufficient to efficiently counterbalance the number of conformations assembled combinatorially from the starting fragments.

We determine a library of 4418 structures (and sequences), further refined by energy minimization, which cover more than half of the possible topologies. The structures are grouped together according to unique glycosidic bond conformation, topology and loop length and for each group the most representative structure is chosen. This clustering procedure results in a set of 116 representative G-quadruplex structures which can be used, in the absence of other structural information, to interpret data like those coming from UV, CD or FRET experiments which provide only partial structural information. Examples of possible applications are given.

Results and Discussion

G-quadruplex model generation

Quadruplex structure selection

The search in the Nucleic Acid Database (NDB) [49] for quadruplex DNA structures returned 101 entries. Unfortunately this list did not include all G-quadruplex containing structures. The Protein Data Bank (PDB) [50] was searched for entries containing the words "tetraplex" or "quadruplex" and the list was filtered by visual inspection. The sequences extracted for each chain in the corresponding PDB files were searched for a $G_3X_{n1}G_3X_{n2}G_3X_{n3}G_3$ motif. Only 14 such sequences were found that were corresponding to intrastrand G-quadruplexes (PDB ids. 143D, 186D, 1KF1, 1XAV, 230D, 2F8U, 2GKU, 2HY9, 2JPZ, 2JSL, 2JSM, 2O3M, 201D, 3CDM), including 230D which contains the nucleotides uridine and inosine-phosphate. A literature survey was also done independently to retrieve the released intramolecular quadruplex structures. The search query ((quadruplex OR tetrad OR tetraplex OR G-4 OR tetramer) AND (intramolecular OR unimolecular OR monomolecular)) in Pubmed resulted in 344 hits. Scanning the abstracts manually resulted in 86 articles relevant to structural studies of quadruplexes. The author names from these articles were collected and searched for individually in Nucleic Acid Database for any deposited quadruple helix structures. No novel intrastrand structure was found in this procedure and thus we trust the set of 14 structures to be complete.

Such paucity of intrastrand G-quadruplex structure may be related to the well known polymorphism of poly-dG [54,55] and the difficulty in obtaining crystals for longer DNA sequences or obtaining single solution forms for NMR studies [5].

Assembly of DNA G-quadruplex stems from fragments

The selection of fragments from the available structures produced, after clustering and selection of representatives for similar conformations, 58 stem fragments and 65 loop fragments, each representing different features, with respect to diversity in sequence, parallel or antiparallel arrangement, loop length and base pairings.

We use here the term "base pairing" as possible participation in the same G-tetrad. The base pairings of the first base in the fragment may involve the edge of the base involved in Watson-Crick base pairing (entailing atoms N1 and N2) or the edge of the base which is involved in Hoogsteen base pairing (entailing atom O6 and N7). These base pairs are hereafter named edge-wise. Alternatively hydrogen bonds may be missing altogether when the second stem is located at the opposite corner of the tetrad. These base pairs will be hereafter named tip-wise. We refer to the three possibilities mentioned above as WH, HW or X (cross) pairing, respectively, or for the sake of notation 0, 2 and 1, respectively.

Loops connecting edge-wise and tip-wise antiparallel strands correspond to lateral and diagonal loops, respectively according to Webba da Silva [56]. Edge-wise loops connecting parallel strands correspond to propeller loops, according to the same author.

The distribution in loop lengths is uneven, with just two loops of length 2 and seven loops of length 1. These short loops are found exclusively in a parallel arrangement. 3 and 4 nucleotide loops are found 35 and 22 times, respectively. Longer loops are found both parallel and antiparallel. Interestingly, a loop connecting two parallel strands at

5. APPENDIX

the opposite corners of a tetrad is also present. Bases in this loop, however, participate the G-tetrads and therefore will be discarded, for steric reasons, in the following assembly of G-quadruplexes.

The features of the selected fragments are reported in Table 1.

The stems of the fragments were used to build up the four-strand G-quadruplex stems. With the loose requirements of no more than 0.8 Å RMSD between the superimposing fragments and no overlap below 0.5 times the sum of van der Waals radii (see Methods) 646 G-quadruplexes were built whose tetrad planes were rebuilt using the frame provided by the first three G strand in the sequence. In this step it was checked that the model G-tetrad could be well placed on the C1' anchor points. Models which exhibited an RMSD larger than 3.0 Å were discarded, leaving a set of 509 G-quadruplex stem models.

Table 1: Non-redundant features of the fragments selected from the database.

syn/anti	a/p	loop length	pairing	counts
a s a	a	4	111	13
a a a	p	3	000	10
s a a	a	3	200	7
s s a	a	3	002	6
s a a	p	3	200	6
a a a	p	1	000	6
a s a	a	4	020	5
s s a	a	3	220	2
a s a	a	3	020	2
a a a	p	4	111	2
s s a	a	4	111	1
s a s	a	3	111	1
s a a	p	2	200	1
s a a	p	1	200	1
s a a	a	4	200	1
s a a	a	4	111	1
a s a	a	4	202	1
a s a	a	3	202	1
a a a	p	2	000	1

syn/anti indicates the conformation at the glycosidic bond of the first three G's, a/p indicates antiparallel/parallel arrangement, 0, 1 and 2 pairings are described in the text and the number of fragments different in sequence or conformation is given in the column "counts".

Rebuilding the G-tetrad was necessary because, due to the tolerant cutoff used for fragment assembly, base pairing was not always consistent with the hydrogen bonding pattern of a G-tetrad. For this reason the four G's constituting the G-tetrad were replaced by a standard G-tetrad by first superimposing the first G (numbered 1 in Figure 1) on the G of the first strand in the molecule in order to determine the orientation of the G-tetrad and then superimposing the C1' atoms of the tetrad with those of the G-quadruplex.

Assembly of DNA G-quadruplex from G-quadruplex stems and loops

The strands of the 509 G-quadruplex stems determined as described above were connected using the loops of the fragments selected from the NDB and PDB quadruplex dataset. 65 non-redundant loops were used resulting in 509 × 65 × 65 × 65 possible combinations. Many of these were ruled out by steric hindrance or poor superposition of the anchor G's preceding and following the loop. Nevertheless 4418 molecular models have been generated reflecting a variety of parallel/antiparallel dispositions, loop lengths, syn/anti glycosidic bond angles. Many of these models still suffered from long bonds resulting from merging fragments and steric hindrances and for this reason they were refined by energy minimization.

Molecular mechanics refinement

All 4418 model were subjected to 300 steps of molecular mechanics minimization keeping the G-tetrads fixed. At the end the energy distribution of the models was quite homogeneous with energies ranging mostly between 700 and 1500 kcal/mol. Only four models were clearly separated from the remaining ones at much higher energy (two at ca. 13000 kcal/mol and two at ca. 63000 kcal/mol) pointing out serious steric hindrance. Visual examination shows that the rebuilt G-tetrads are too close for these four models. The latter models have not been considered in the following clustering procedure.

Clustering of structural models

All energy minimized models are available, together with sequences and a table of energies and topologies, from the authors. However, for more convenient usage, the models were clustered according to unique glycosidic bond conformation, topology and loop length. The models sharing the same glycosidic bond conformations, topology and

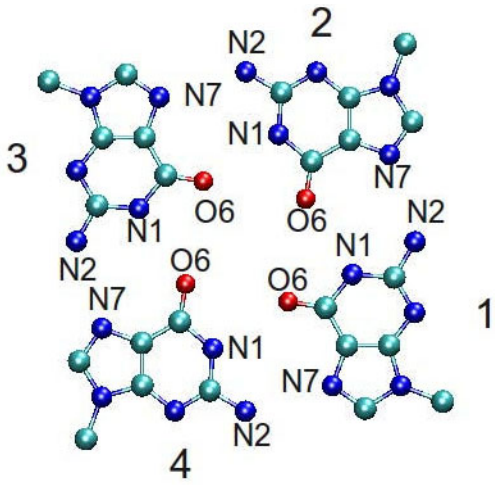


Figure 1
The standard G-tetrad with reference numbering. When the glycosidic bond angle is *anti* the chain progresses over the page, when it is syn the chain progresses below the page. According to Webba da Silva [56] the sign of the loop topology is positive when the first stem is progressing towards the viewer and the second stem is found rotating clockwise, and negative when it is found rotating anti-clockwise. E. g., when the glycosidic bond angle is *anti* the topology of a loop connecting the stem of base 1 to the stem of base 2 would be marked with - sign, and the topology of a loop connecting the stem of base 1 to the stem of base 4 would be marked with + sign.

loop lengths were pairwise compared and for each model a threshold RMSD was chosen and a weight was assigned based on the RMSD with all other structures. The model with largest weight was chosen as representative of all models with RMSD lower than threshold. The procedure was repeated, increasing the threshold RMSD, until a single model was left. The energies of the resulting models range between 690 and 1656 kcal/mol, a range comparable with that found for experimental structures subjected to the same minimization procedure (590 to 913 kcal/mol), taking into account that experimental structures are typically already refined. The most representative structure for each of the 116 clusters (see Table 2) is provided [see Additional file 1].

Analysis of G-quadruplex models

Comparison with experimental structures

An obvious test for the methodology is to check whether it is able to recover the observed intrastrand G-quadruplexes from the fragments which are not taken from that intrastrand G-quadruplexes. Due to sequence diversity it will be in general hard to recover exactly the same sequence. For instance, for loops of length 2 there are only two fragments with different sequence. When one of the two loops is excluded from the list of fragments there will be no possibility to obtain a loop of length 2 with the same sequence. Nevertheless we will consider here the topologies which are generated from assembly of fragments with the same loop lengths.

In order to test the overall reliability of the method we considered the set of intrastrand structures with three G stem strands with loops of length 3. In the following we indicate the overall topology of a model by noting the sequence of loops as lateral (l), propeller (p) or diagonal (d). The clockwise (+) or anti-clockwise (-) rotation of lateral and propeller loops is with respect to a common frame of reference (see [56]). It was not possible to extend this analysis to the other structures because they contain loops of length 2 for which only two fragments are present in the dataset.

The structures with pdb id. 1KF1, 2GKU, 2HY9, 2JPZ, 2JSL, 2JSM and 3CDM all contain the core sequence GGGTTAGGGTTAGGGTTAGGG and adopt three different topologies: namely -p-p-p, -l-l-p, -p-l-l.

We considered for each topology all built models which do not contain any fragment derived from structures with the same topology. Moreover, due to the fragment clustering procedure adopted, no fragment is present in the dataset closely related to those present in those structures. All three topologies are actually represented several times in the built models with RMSD over sugars and phosphates from the original PDB structures between 2.0 and 3.0 Å. An example with the real structure (pdb id: 2hy9) and the model assembled from fragments is reported in Figure 2. Although not all loops are similar to the real ones, by construction, the topology and overall conformation is reproduced well by the model. The RMSD computed on all backbone atoms is 2.2 Å.

It must be however noted that not only the observed topologies for the sequence $G_3N_3G_3N_3G_3N_3G_3$ are found in the models but also others, although the observed topology is among the most represented ones.

Topologies and loop lengths

Another test for the methodology is to check whether the relationship between loop lengths and topology matches the available experimental evidences.

Overall, the topologies of the models generated using all the available fragments are not evenly distributed (Table 3). Only 14 out of 26 possible looping topologies are

5. APPENDIX

found. There is a clear predominance of the -p-p-p topology that is found in ca. one third of all models.

This finding is consistent with the observation that the number of parallel fragments which are actually used for assemblying the stems of G-quadruplex is larger than the number of antiparallel fragments, although the starting fragments (i.e. strand-loop-strand fragments) do not show such parallel predominance. This might reflect a general more regular arrangement of parallel versus antiparallel strands, at least in the selected dataset. The adoption of a -p-p-p topology leads to right-handedness of the polynucleotide chain in the G-quadruplex.

Other well represented topologies are the mixed topologies +l+p+p, -l-l-p, -p-p-l, -pd+p, +l+p+l d+pd and the all-antiparallel +l+l+l.

Some care must be taken when considering potential G-quadruplexes involving loops of 4 residues because no propeller-like loop is found in the starting dataset of fragments. There are therefore no all parallel topologies involving loops of 4 residues, although it has been shown that the G-rich sequence in the human VEGF gene promoter adopts an all-parallel structure involving a loop of 4 residues [15,57]. Most frequently diagonal fragments are found.

Only 47 possible combinations of loop lengths are found out of 64 possible. Of these some are more largely represented as a consequence of the uneven distribution of the number of starting loops. In general there is no direct relationship between loop lengths and topology, although sequences with loops as short as 1 or 2 nucleotides are, as expected from the starting fragments, found in all-parallel topology. For longer loop lengths typically many different topologies are found.

It is interesting to note that some of the combination of loop lengths are found with unique topology, among these the most widely represented are 4, 3 and 4 (122 models), 1, 3 and 1 (58 models), 1, 1 and 3 (53 models) (Table 2).

The effect of loop length on G-quadruplex topology and/or stability has been studied by many authors [34,35,37-39,58-61] under different conditions. Not all studies however address the formation of intramolecular G-quadruplexes. The two recent papers by Bugaut and Balasubramanian [37] and Smargiasso et al. [38] investigate systematically the effect of loop length, with randomized sequences, on G-quadruplex stability and topology. In both studies intramolecular vs. intermolecular G-quadruplex formation is experimentally addressed. Notwithstanding different experimental conditions, these studies provide, among other results, a general conclusion which is well in line with previous evidences: in general short loops (and in particular the presence of loops of length one) strongly favor parallel arrangement of the strands while for longer loops antiparallel and mixed arrangements are observed. The topology of the models built here appears consistent with experimental evidences.

Possible applications

The present study constitutes a proof of principle, obviously physical or statistical effective energy functions should more accurately measure the stability of the predictive models. Moreover the limited diversity in sequence does not allow to build models for all possible sequences. A third limitation of the present approach is that no consideration of flanking residues which are known to be important for the stability of G-quadruplex is taken into account. In addition to these problems the starting fragments are in limited number as exemplified by the lack of parallel propeller loops of length 4.

It is worth however to explore how structural predictions could complement experimental and bioinformatics approaches.

It must be clear that the actual structure adopted by a DNA sequence depends on many factors including flanking and loop sequences and environmental conditions. The models built from experimental fragments constitute however a set of structures whose features are consistent with experimental structures. Due to the limited number of structures solved so far, the set is not expected to cover all possible structures. However, even in the presence of polymorphism the models proposed here constitute structural working hypotheses that can complement experimental techniques.

The aim of the following subsection is to show, by selecting a few possible applications that inferences based on the built models are consistent with experimental evidence and thus provide an overall test of reliability for the proposed models.

It is well known that potential G-quadruplex sequences play a regulatory role but the nature of such role is different according to the position of the sequence with respect to the TSS and the strand where it is found [43].

The models provided by the present study could be used straightforwardly as starting models for molecular dynamics simulations or docking studies. Another possibility is to use the topology information provided here to complement other studies. The same topology could be required by different DNA quadruplex sharing a common

Table 2: Features of modeled intrastrand G-quadruplexes.

syn/anti	loop topology	strand polarity	loop 1	loop 2	loop 3	counts
a a a	-p-p-p	ppp	1	1	1	10
a a a	-p-p-p	ppp	1	1	2	6
a a a	-p-p-p	ppp	1	1	3	53
a a a	-p-p-p	ppp	1	2	1	8
a a a	-p-p-p	ppp	1	2	2	2
a a a	-p-p-p	ppp	1	2	3	20
a a a	-p-p-l	ppa	1	2	3	6
a a a	-p-p-l	ppa	1	2	4	2
a a a	-p-p-p	ppp	1	3	1	58
a a a	-p-p-p	ppp	1	3	2	8
a a a	-p-p-p	ppp	1	3	3	237
a a a	-p-p-l	ppa	1	3	3	34
a a a	-p-l-l	pap	1	3	3	3
a a a	-p-p-l	ppa	1	3	4	12
a a a	-pd+p	paa	1	4	1	3
a a a	-pd+p	paa	1	4	2	10
a a a	-pd+l	ppa	1	4	3	10
a a a	-p-l-l	pap	1	4	3	1
a a a	-pd+p	paa	1	4	3	35
a a a	-p-p-p	ppp	2	1	1	4
a a a	-p-p-p	ppp	2	1	2	2
a a a	-p-p-p	ppp	2	1	3	21
a a a	-p-p-p	ppp	2	2	1	3
a a a	-p-p-p	ppp	2	2	2	1
a a a	-p-p-p	ppp	2	2	3	11
a a a	-p-p-l	ppa	2	2	3	3
a a a	-p-p-l	ppa	2	2	4	1
a a a	-p-p-p	ppp	2	3	1	23
a a a	-p-p-p	ppp	2	3	2	3

5. APPENDIX

Table 2: Features of modeled intrastrand G-quadruplexes. *(Continued)*

a a a	-p-p-l	ppa	2	3	3	17
a a a	-p-l-l	pap	2	3	3	2
a a a	-p-p-p	ppp	2	3	3	74
a a a	-p-p-l	ppa	2	3	4	5
a a a	-pd+p	paa	2	4	1	13
a a a	-pd+p	paa	2	4	2	3
a a a	-pd+p	paa	2	4	3	16
a a a	-pd+l	ppa	2	4	3	1
a a a	-p-l-l	pap	2	4	3	1
a a a	-p-p-p	ppp	3	1	1	28
s a s	-p-p-p	ppp	3	1	1	2
s s a	+l+p+p	aaa	3	1	1	34
s a a	-p-p-p	ppp	3	1	1	6
a a a	-p-p-p	ppp	3	1	2	10
s a a	-p-p-p	ppp	3	1	2	6
s s a	+l+p+p	aaa	3	1	2	7
s s a	+l+p+p	aaa	3	1	3	120
s a a	-p-p-l	ppa	3	1	3	12
a a a	-p-p-p	ppp	3	1	3	131
s s a	+l+p+l	paa	3	1	3	27
s a a	-p-p-p	ppp	3	1	3	49
a a a	-p-p-l	ppa	3	1	3	4
a a a	-p-p-l	ppa	3	1	4	1
s a a	-p-p-l	ppa	3	1	4	3
s s a	+l+p+l	paa	3	1	4	6
s s a	+l+p+p	aaa	3	2	1	10
s a a	-p-p-p	ppp	3	2	1	6
a a a	-p-p-p	ppp	3	2	1	9
s a a	-p-p-p	ppp	3	2	2	2
a a a	-p-p-p	ppp	3	2	2	3

Table 2: Features of modeled intrastrand G-quadruplexes. *(Continued)*

s s a	+l+p+p	aaa	3	2	2	3
s a a	-p-p-p	ppp	3	2	3	22
a a a	-p-p-p	ppp	3	2	3	33
s s a	+l+p+p	aaa	3	2	3	46
s a a	-p-p-l	ppa	3	2	3	6
a a a	-p-p-l	ppa	3	2	3	9
s s a	+l+p+l	paa	3	2	3	9
s a a	-p-p-l	ppa	3	2	4	2
a a a	-p-p-l	ppa	3	2	4	3
s s a	+l+p+l	paa	3	2	4	3
s a a	-l-l-p	app	3	3	1	102
s s a	+l+p+p	aaa	3	3	1	102
a a a	-p-p-p	ppp	3	3	1	125
s a a	-p-p-p	ppp	3	3	1	80
a s a	-l-l-p	app	3	3	2	2
s a a	-l-l-p	app	3	3	2	70
s a a	-p-p-p	ppp	3	3	2	8
a a a	-p-p-p	ppp	3	3	2	9
s s a	+l+p+p	aaa	3	3	2	9
s a a	-l-l-l	apa	3	3	3	10
s s a	+l+l+l	apa	3	3	3	167
s a a	-p-p-p	ppp	3	3	3	241
a a a	-p-l-l	pap	3	3	3	30
s s a	+l+p+p	aaa	3	3	3	367
s a a	-l-l-p	app	3	3	3	385
a s a	-ld+l	aap	3	3	3	3
a a a	-p-p-p	ppp	3	3	3	450
s a a	-p-p-l	ppa	3	3	3	66
a a a	-p-p-l	ppa	3	3	3	68
a s a	-l-l-p	app	3	3	3	8

5. APPENDIX

Table 2: Features of modeled intrastrand G-quadruplexes. *(Continued)*

s a a	-p-l-l	pap	3	3	3	91
s s a	+l+p+l	paa	3	3	3	95
s a a	-p-p-l	ppa	3	3	4	12
s s a	+l+p+l	paa	3	3	4	15
a a a	-p-p-l	ppa	3	3	4	19
s a a	-l-l-l	apa	3	3	4	19
s s a	+ld-p	ppa	3	4	1	24
s a a	-pd+p	paa	3	4	1	6
a a a	-pd+p	paa	3	4	1	9
a a a	-pd+p	paa	3	4	2	15
s s a	+ld-p	ppa	3	4	2	3
s a a	-pd+p	paa	3	4	2	5
s a a	-pd+l	ppa	3	4	3	12
a a a	-pd+l	ppa	3	4	3	13
s a a	-p-l-l	pap	3	4	3	14
a a a	-p-l-l	pap	3	4	3	3
s s a	+ld-l	paa	3	4	3	3
s a a	-pd+p	paa	3	4	3	44
s s a	+ld-p	ppa	3	4	3	51
a a a	-pd+p	paa	3	4	3	80
s s a	+l+l+l	apa	3	4	3	82
s a a	-l-l-p	app	4	3	1	2
s a a	-l-l-p	app	4	3	2	8
a s a	-ld+l	aap	4	3	3	2
s a a	-l-l-p	app	4	3	3	33
a s a	d+pd	aap	4	3	4	114
s a a	d+pd	aap	4	3	4	8

The notation here follows Webba da Silva [56]. syn/anti indicates the conformation at the glycosidic bond of the first three G's. The loop topology is indicated by letters p (parallel), l (lateral) and d (diagonal) preceded by + or - sign to indicate clockwise or anti clockwise rotation when the first strand is progressing towards the viewer (see Figure 1). Similarly, the parallel or antiparallel (a/p) strand polarity in column 2 is with reference to the first strand and the order is according to the position in the quadruplex (rotating anti-clockwise with the first strand progressing towards the viewer), and in general not according to sequence order. The next three fields indicate loop lengths and the last field indicate the number of built models found with these features.

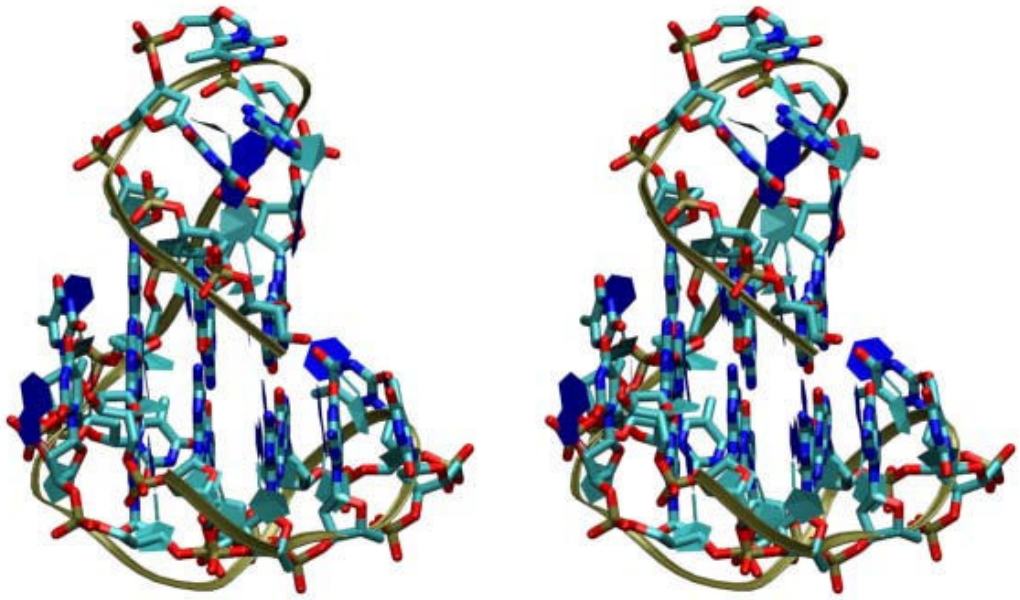


Figure 2
Model for a human telomeric DNA G-quadruplex structure (pdb id: 2HY9). In the stereoview the experimental structure is displayed as a ribbon with sugars and bases schematic representations and the model is displayed as solid bonds. The RMSD computed on all backbone atoms is 2.2 Å.

mechanism of gene regulation. We consider here that among loop length combination showing a unique (all-parallel -p-p-p) conformation we find loop length combination of (1,1,1), (1,2,1) and (1,3,1). The second one has been described experimentally as an all-parallel G-quadruplex [62], while the other two have not been solved experimentally. For the first one the all parallel topology should be strongly favored by the presence of all single-nucleotide loops [36-38]. For the last combination of loop lengths also a mixed topology is in principle possible, but it is not found among our models, notwithstanding the large number of loops of length 3 available among the starting fragments.

A (1,3,1) loop length combination has been found in the promoter of the oncogene RET and its topology was described as all-parallel consistent with our predictive model [63].

A word of caution is due here: although intramolecular G-quadruplex formation has been observed for this loop length combinationation [37], a study by Vorlickova [35] and colleagues tested under different conditions sequences $(G_3X_n)_3G_3$ with n = 1, 2, 3 and they found that these sequences formed mostly intermolecular G-quadruplexes. Only in ethanol solutions the same sequences adopted intramolecular parallel conformation. In this work we do not consider intermolecular, but only intramolecular G-quadruplexes. Moreover, in view of the known polymorphism of G-quadruplexes, our models suggest which conformation could be attained by a given sequence, compatible with structural observations. These conformations could be adopted only under peculiar environmental conditions, e.g. such as those described by Vorlickova and coworkers [35] or Bugaut and Balasubramanian [37].

In the following we will focus on the G-quadruplex forming sequence found in the RET promoter, whose loop length combination is associated with a unique all-parallel (-p-p-p) topology, and with all loop length combinations sharing the latter feature in our models.

Molecular dynamics simulations

An obvious application of structural models is computer simulations of their molecular dynamics. A necessary,

5. APPENDIX

Table 3: Topology distribution of model G-quadruplexes.

topology	counts
-p-p-p	1764
+l+p+p	698
-l-l-p	610
-p-p-l	285
+l+l+l	249
-pd+p	239
+l+p+l	155
-p-l-l	145
d+pd	122
+ld-p	78
-pd+l	36
-l-l-l	29
-ld+l	5
+ld-l	3
total number of topologies	total number of models
14	4418

The distribution of topologies (independent of glycosidic bond conformation and loop lengths) of all 4418 models is reported. The notation here follows Webba da Silva [56]. p, l and d stand for propeller-like, lateral or diagonal loop. Theand + signs refer to anti-clockwise or clockwise rotation of the loop around the G-quadruplex stem, respectively, when the first strand is progressing towards the viewer (see Figure 1).

albeit not sufficient, condition for a model to be accurate is that the structure is stable during a molecular dynamics simulation for a time sufficient in principle to develop major conformational rearrangements. The benefits and limits of molecular dynamics simulations of G-quadruplexes have been reviewed by Spooner and Spackova [64]. In the study by Hazel et al. [58] molecular dynamics simulations complemented experiments and model building was performed in order to provide starting models. We consider here as an example the sequence GGGCGGGGCGGGCGGG that is found in the promoter of the oncogene RET, which adopts an all-parallel topology [63].

The most representative predictive model for the unique all-parallel (-p-p-p) topology for loop length combination (1,3,1) was taken and the sequence was mutated to the target sequence. Two potassium ions were added at the centre of the O6 atoms of adjacent tetrads and counterions were further added to make the system neutral. The system was solvated in a box of water extending at least 12 Å away from each heavy atom of the solute. The preparation of the system is essentially as previously described for a different system [65]. The forcefield employed is CHARMM version 31 [66,67].

Molecular dynamics simulation was run for 20 ns in order to check for any major conformational change which could indicate bad quality of the starting model or wrong topology [58,64].

After few hundred picoseconds one of the two potassium ions at the centre of adjacent tetrads goes in solution while the other is firmly retained. Loss of ions from the central channel has been observed before in molecular dynamics simulations and it has been ascribed to forcefield inaccuracies [64]. During the simulation the G-quadruplex structure is mantained. The average RMSD from the starting structure is 1.0 Å. Larger fluctuations are observed at the three residue loop both for the backbone and for the base moieties similar to other molecular dynamics simulation studies [58,64]. No loop residue is involved in intramolecular hydrogen bonds. This example proves (at least on the timescale of 20 ns) that the model quality is suitable for molecular dynamics simulations because otherwise large changes in the G-quadruplex structure would be expected [58,64].

Docking studies

Predictive models of G4 may be employed for docking studies (see e.g. [68,69]). As an example we considered the model for the sequence GGGCGGGGCGGGGCGGG that is found in the promoter of the oncogene RET, which adopts an all-parallel topology [63], as in the previous section.

This sequence has been shown to be stabilized by the cationic porphyrin TMPyP4 (5,10,15,20-tetrakis(1-methylpyridin-1-ium-4-yl)-21,22-dihydroporphyrin) and it was suggested that the binding involves stacking rather than intercalation [63].

Two G-quadruplex-TmPyP4 complexes have been structurally characterised by NMR (pdb id. 2A5R, [70]) and by X-ray crystallography (pdb id. 2HRI, [71]). The two complexes show remarkable differences. In the NMR structure the porphyrin is stacking over the first tetrad and is covered by the two residues 5' to the G-quadruplex. In the crystal structure one porphyrin is stacked over a base pair over the tetrads and the other is contacting a grove with electrostatic interaction with a phosphate and a stacking interaction with a base in the loop.

The ligand structure was taken from the Hic-Up server http://xray.bmc.uu.se/hicup/[72]. Based on the SMILES representation of the compound available from the database PubChem (CID: 4234) [73] partial charges have been assigned by the program Babel [74] implementing the Gasteiger and Marsili method [75]. The structure of the model G-quadruplex with partial charges has been obtained as decribed in the previous section. The program Dock6.3 [76] has been used for generating, scoring and clustering TMPyP4-G-quadruplex complexes following a standard protocol and using the AMBER forcefield for estimating the energy of van der Waals intermolecular contacts. 20000 poses were generated and after clustering at 2.0 Å RMSD the best 10 were retained, showing all large negative interaction energies. Consistent with the above cited previous studies, in nine out of ten complexes the arrangement of the prophyrin is parallel and stacking onto the tetrads, although stacking involves only half of the tetrad (Figure 3). In the remaining complex the porphyrin is contacting the G-quadruplex in the groove and displays electrostatic interaction between the pyrimidinium and the phosphate.

Overall these results are consistent with what could be expected based on previous structural characterization and thus show that the models can provide a starting target for use in docking studies.

Cancer genes

Potential G-quadruplex forming sequences have been found in a number of cancer genes [33] mostly sharing the first and last loops of length 1.

Consistent with earlier studies [33,37,38], sequences having the second loop of length 1 to 3 can adopt in our models only an all-parallel -p-p-p topology, while any other topology would be not consistent with the available set of experimental structures. We further explore whether the presence of sequences for which a unique all-parallel topology is found could be a distinctive feature of cancer genes. Following previous analyses we have limited our search to a putative regulatory region of -200 bp, 0 bp around the TSS of all genes in the Ensembl database. We first aligned the sequences:

$G_3N_1G_3N_1G_3N_1G_3$,

$G_3N_1G_3N_2G_3N_1G_3$,

$G_3N_1G_3N_3G_3N_1G_3$

on the putative regulatory regions and the search returned 728 unique genes containing at least one of the sequences.

Before any further consideration it should be considered that the pattern $G_3N_1G_3$ is shared with consensus motif G_3CG_3 of SP1 binding site, which is rather common in human genes in the region within 200 bp upstream of the TSS [77]. Inclusion of the G_3CG_3 consensus sequence for SP1 binding site in many of the potential G-quadruplex sequences is likely to add noise to any statistical analysis and to reduce the calculated significance values.

The Ensembl gene names were translated, where possible, onto HNGC gene names and the overlap between the set of the resulting 686 genes and the Census set of 385 cancer genes available at the http://www.sanger.ac.uk/genetics/CGP/Census/[78,79] was determined. The overlap set contained 23 genes, a number higher than that expected by chance, i.e. 14.5.

The probability that 686 genes chosen randomly out of 19589 genes with HGNC name could have an overlap of 23 or more with the Census set (385 genes) was calculated using the hypergeometric distribution and the resulting p-value was 0.01. This result, based on putative adoption of a common structure, suggests that G-quadruplex gene regulation may be a common feature of cancer genes.

The above set of loop length combinations is however only a restricted set of the larger set of all loop length combinations which are associated in the predicted models to a unique all-parallel (-p-p-p) topology (see Table 2).

The same analysis has been repeated considering the latter set. If the topology is an important feature shared by G-quadruplex sequences found in many oncogenes we would expect also for the larger set of loop combinations a higher number of hits in oncongenes than expected by chance. Indeed this is the case. The overlap between the 1607 genes, containing at least one of the selected loop combinations, and the Census set consists of 47 genes (Table 4), higher than the expected 32, and corresponding to a p-value of 0.003. These results are consistent with the known importance of all-parallel topology for G-quadruplex forming sequences in the regulation of proto-oncogenes [33].

Developmental genes

The restricted set of genes that contain a potential all-parallel quadruplex helix has been screened for overrepresentation in Gene Ontology annotation. The general terms "developmental process", "system development", "anatomical structure development", "multicellular organismal development" are found with the lowest p-values (less than 10^{-10}). The same analysis on the larger set of genes containing a potential all-parallel quadruplex helix gives essentially the same results with even lower p-values (ranging from 10^{-12} to 10^{-17}).

5. APPENDIX

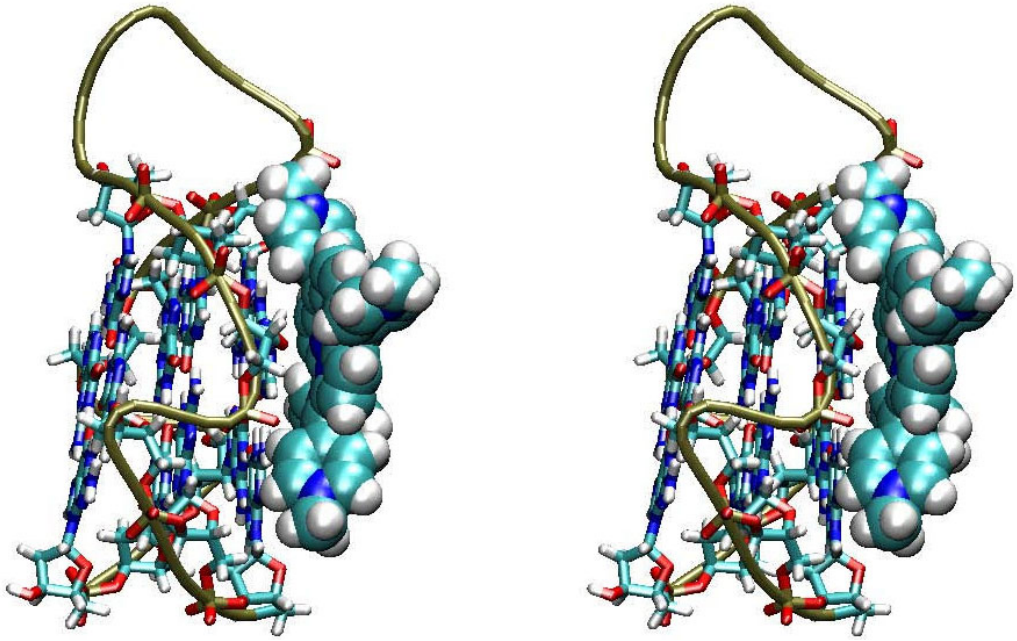


Figure 3
Model of TMPyP4 docking on the model for RET promoter G-quadruplex structure. In the stereoview TMPyP4 atoms are shown as Van der Waals spheres and DNA backbone is shown as a tube. The bonds of residues of the G-quadruplex tetrads are shown.

The same search for the all parallel motifs discussed above on the restricted set was performed on putative regulatory regions of -200 bp, 0 around the TSS of all mouse genes in the Ensembl database returning 841 genes. Remarkably enough, looking for overrepresented Gene Ontology terms in this set of genes we found exactly the same terms already found in the human case, albeit with slightly higher p-values (less than 10^{-8}). Comparing the 728 human genes and the 841 mouse genes using the list of 21605 orthologous genes between human and mouse contained in the Ensembl database we found an intersection of 104 genes. The p-value of the overlap, computed under the assumption of no significant conservation in putative gene regulatory regions is as low as 10^{-31}. However, since some conservation in the putative regulatory regions is expected and is indeed found the latter p-value should be regarded with some caution.

Conclusion

A large number of molecular models has been generated for intrastrand G-quadruplex formed by $G_3X_{n1}G_3X_{n2}G_3X_{n3}G_3$ sequences. For a given sequence not all topologies are possible with the available repertoire of fragments due to steric hindrance and low superimposability. Since all molecular models are generated by fragments coming from observed quadruplex structures, molecular models are in principle reliable and may be used for interpretation of experimental data. Molecular models for different loop length combinations suggest that the all-parallel topology is strongly favored. Notwithstanding the limitations of the approach, the models could be useful for molecular modeling and docking studies, and in general to complement other laboratory and bioinformatics methods.

Methods

DNA fragment generation

Quadruplex structure selection

Structures for DNA (or DNA/RNA or RNA) quadruplexes were selected using the search tools available at the nucleic acid databank http://ndbserver.rutgers.edu/[49] requiring "quadruple helix" as structural feature.

The search resulted in 101 entries. Of these 91 are containing only DNA quadruplexes. The PDB code for the 101

entries was used for retrieving the relevant structures from the Protein Data Bank http://www.rcsb.org/[50].

This search was apparently missing some of the G-quadruplexes in the PDB. For this reasons we selected all structures in the PDB containg anywhere the words quadruplex or tetraplex and we hand filtered those that could contain a genuine DNA G-quadruplex. The latter step retrieved additional 51 structures. The selected structures were searched for the presence of strands with three adjacent G's paired with another strand with three adjacent G's with either parallel or antiparallel linear arrangement, and with loop connection of one to four nucleotides. The loop lenghts considered here are somewhat shorter than the limit of seven used e.g. by Chowdhury and coworkers [47]

Since the quadruplex are assembled from these fragments we found useful to reference these fragments and their pairs of bases defining the G-tetrad plane to a standard G-tetrad. The parallel or antiparallel orientation of the two G-strand is an obviously important feature of the fragment. Referencing serves the purpose of detecting and storing fragments that can be used for building the G-quadruplex and is not meant as a definition for classification. A standard for classification and notation of G-quadruplexes has been proposed by Webba da Silva [56]. We conform here to that proposed standard, although we report also a local description of the structure (vide infra)

A standard G-tetrad was generated by rotation and translation of a G base (taken from the standard fragments of the X3DNA program [80]). The best hydrogen bonding geometry was obtained by rotating repeatedly the G base of 90 degrees and traslating by -.70 and 7.10 Å along the x and y axis respectively, with reference to the coordinates used in the X3DNA base coordinates.

The first G base of each fragment with sequence $G_3X_{n1}G_3$ was superimposed to base 1 in the model G-tetrad. Then the first (last) G base of the second run of three G's in the fragment was superimposed in turn onto the other bases in the G-tetrad in order to find the first base pair of the parallel (antiparallel) three G's pair. The same procedure with due modifications was repeated for the second and third G pairs. The fragment was accepted as good if the RMSD in all three superpositions was less than 1.0 Å. The tolerant threshold was dictated by the large conformational heterogeneity observed in G-tetrads.

The first (last) G base could be in anyone of the three other positions of the G-tetrad (Figure 1). The second and third G's could be found over or below the plane defined by the G-tetrad used for superposition of the first G depending on the torsion angle at the glycosidic bond.

For each fragment the loop and the "stem" constituted by two strands of 3 adjacent G's were stored. The 147 fragments (with the 147 loops and 147 stems) obtained in this way are redundant because the same structure may have been resolved by different groups and techniques and because a single PDB entry may contain the same structure more than once.

In order to remove redundancy we performed clustering. All fragments identical in sequence were compared and representatives were selected in such a way that none chosen conformation has less than 0.8 Å RMSD on all heavy atoms. This led to 68 unique complete fragments, 67 loops and 57 stems.

DNA G-quadruplex assembly from fragments

The 57 stems were used to assemble the quadruple helix. Three "stems" were assembled together by superimposing the second three G's with the first three G's of the next stem with less than 0.8 Å RMSD and with no overlap of heavy atoms at more than 0.5 times the sum of their van der Waals radii. This procedure led to 646 quadruplex stem models. The stems were therefore further modified by substituting the model G-tetrad for each G-tetrad. The model G-tetrad was set in place by superimposing first the G of the first strand and then superimposing the four C1' atoms to the closest ones in the G-tetrad. If the RMSD was larger than 3.0 Å the model was not taken into account. At the end of this step 509 models were retained. The conformation of the glycosidic bond angle is thus determined by the first strand of the quadruplex.

Finally loops were added to the quadruplexes whenever the superposition of the sugars linked to the G's preceding and following the loop gave an RMSD less than 0.8 Å and with no overlap of heavy atoms at more than 0.5 times the sum of their van der Waals radii. Although the combinatorial number of possible models is extremely large, in practice this computation may be performed on a PC. This last step generated 4418 models.

Molecular mechanics refinement

Due to the rather tolerant cutoff on RMSD's the bonds connecting stems and loops were in many instances large. The refinement was performed by first substituting the G-tetrads with the regular G-tetrad generated by optimal rototraslation of G, as described above, and then keeping the G-tetrads fixed and performing energy minimization. For this purpose the program NAMD [81] was used employing a dielectric constant of 10 and the forcefield CHARMM version 31 [66,67]. 300 steps of conjugate gradients minimization were performed keeping the base atoms of the tetrads fixed.

5. APPENDIX

Table 4: Human cancer genes containing potential all-parallel G-quadruplexes.

Gene symbol	Gene name
AKT1	v-akt murine thymoma viral oncogene homolog 1
ASPSCR1	alveolar soft part sarcoma chromosome region, candidate 1
ATF1	activating transcription factor 1
BCL3	B-cell CLL/lymphoma 3
BRCA2	familial breast/ovarian cancer gene 2
CARD11	caspase recruitment domain family, member 11
CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)
CLTCL1	clathrin, heavy polypeptide-like 1
ELN	elastin
EPS15	epidermal growth factor receptor pathway substrate 15 (AF1p)
ERCC2	excision repair cross-complementing rodent repair deficiency complementation group 2 (xeroderma pigmentosum D)
ETV6	ets variant gene 6 (TEL oncogene)
FGFR3	fibroblast growth factor receptor 3
FNBP1	formin binding protein 1 (FBP17)
FOXP1	forkhead box P1
FSTL3	follistatin-like 3 (secreted glycoprotein)
GATA1	GATA binding protein 1 (globin transcription factor 1)
HIP1	huntingtin interacting protein 1
HOXA11	homeo box A11
HOXA13	homeo box A13
HOXA9	homeo box A9
IGK@	immunoglobulin kappa locus
IRF4	interferon regulatory factor 4
JAZF1	juxtaposed with another zinc finger gene 1
LHFP	lipoma HMGIC fusion partner
MLLT6	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 6 (AF17)
MSI2	musashi homolog 2 (Drosophila)
MSN	moesin

Table 4: Human cancer genes containing potential all-parallel G-quadruplexes. *(Continued)*

MUC1	mucin 1, transmembrane
MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
PIM1	pim-1 oncogene
POU2AF1	POU domain, class 2, associating factor 1 (OBF1)
PTEN	phosphatase and tensin homolog gene
RANBP17	RAN binding protein 17
RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1
RET	ret proto-oncogene
SEPT6	septin 6
SFRS3	splicing factor, arginine/serine-rich 3
SS18L1	synovial sarcoma translocation gene on chromosome 18-like 1
TAF15	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68 kDa
TCF12	transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)
TMPRSS2	transmembrane protease, serine 2
TRIM33	tripartite motif-containing 33 (PTC7, TIF1G)
TSHR	thyroid stimulating hormone receptor
ZNFN1A1	zinc finger protein, subfamily 1A, 1 (Ikaros)

Clustering of structural models

All energy minimized models sharing the same glycosidic bond conformations, topology and loop lengths were clustered in separate groups. All models within a single group were pairwise compared. A threshold RMSD t was chosen and a weight w_i was assigned to each model i based on the RMSDs lower than t with all other structures:

$$w_i = \sum_j \cos\left(\frac{\pi}{2}\frac{rmsd_{ij}}{t}\right)$$

The model with the largest weight was chosen as representative of all models with RMSD lower than threshold. The procedure was repeated doubling progressively the threshold starting from 0.4 Å until a single model was left.

Genomic searches and analysis

All regions 200 bp upstream the ranscription start site (TSS) of all human genes for all transcripts have been downloaded from Biomart site http://www.biomart.org/. The database and the dataset were ENSEMBL 53 GENES and NCBI36i respectively. The search for potential G-quadruplex sequences with proper loop lengths was performed using the program glsearch in the fasta35.1 software package ftp://ftp.ebi.ac.uk/pub/software/unix/fasta/fasta3/. The same analysis was repeated for all mouse genes using the same database and the dataset NCBI37.

The list of orthologous genes was obtained from the Biomart site selecting only protein-coding genes.

The Census set of 385 cancer genes was downloaded from the http://www.sanger.ac.uk/genetics/CGP/Census/ [78,79].

5. APPENDIX

In order to evaluate the significance of the overlap of k genes between two given sets of n and m genes both taken from the same set of N genes we estimated the probability (p-value) that an equal or larger overlap set could be obtained by chance.

This probability is computed using the hypergeometric distribution:

$$p(k) = \sum_{l=k,\min(n,m)} f(N,n,m,k) = \sum_{l=k,\min(n,m)} \frac{\binom{m}{l}\binom{N-m}{n-l}}{\binom{N}{n}}$$

Authors' contributions

FF and HH carried out the implementation and computational analysis. AC, GE, PV carried out structural analysis. MC and DC wrote most programs for genomic analysis and participated designing genomic analysis. LEX participated in designing the study and preparing the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Models for intrastrand G-quadruplexes*. The name of the file contains all topology information. Each field is separated by the underscore character. The notation here follows Webba da Silva [56]. The first field indicates the glycosidic bond conformation in the first G-quadruplex strand a stands for* anti *and s stands for* syn*. The second field indicates the loop topology by letters p (parallel), l (lateral) and d (diagonal) preceded by + or - sign to indicate clockwise or anti clockwise rotation when the first strand is progressing towards the viewer. Similarly, the third field indicates the parallel or antiparallel (a/p) strand polarity with reference to the first strand. The order is according to the position in the quadruplex (rotating anti-clockwise with the first strand progressing towards the viewer), and in general not according to sequence order. The next three fields indicate loop lengths. The .nrg files contain the energy as ouput by the program NAMD [81]. The total energy is reported in the twelfth field. The .fas files contain the sequence of the representative model.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1472-6807-9-64-S1.TGZ]

Acknowledgements

This work has been supported by Ministero dell'Universita' e della Ricerca (FIRB RBNE03B8KK (FF), FIRB RBRN07BMCT (GE), PRIN 2007M3E2T2 003 (FF) and Borse giovani ricercatori indiani (HH)).

References

1. Moser H, Dervan P: **Sequence-specific cleavage of double helical DNA by triple helix formation.** *Science* 1987, **238:**456-450.
2. Cooney M, Czernuszewicz G, Postel E, Flint S, Hogan M: **Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro.** *Science* 1988, **241:**456-459.
3. Paramasivam M, Cogoi S, Filichev V, Bomholt N, Pedersen E, Xodo L: **Purine twisted-intercalating nucleic acids: a new class of antigene molecules resistant to potassium-induced aggregation.** *Nucleic Acids Res* 2008, **36:**3494-3507.
4. Timsit Y, Moras D: **Cruciform structures and functions.** *Q Rev Biophys* 1996, **29:**279-307.
5. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S: **Quadruplex DNA: sequence, topology and structure.** *Nucl. Acids Res* 2006, **34:**5402-5415.
6. Han H, Hurley L: **G-quadruplex DNA: a potential target for anti-cancer drug design.** *Trends Pharmacol Sci* 2000, **21:**136-142.
7. Eddy J, Maizels N: **Gene function correlates with potential for G4 DNA formation in the human genome.** *Nucleic Acids Res* 2006, **34:**3887-3896.
8. Wang G, Vasquez K: **Z-DNA, an active element in the genome.** *Front Biosci* 2007, **12:**4424-4438.
9. Maizels N: **Dynamic roles for G4 DNA in the biology of eukaryotic cells.** *Nat Struct Mol Biol* 2006, **13:**1055-1059.
10. Siddiqui-Jain A, Grand C, Bearss D, Hurley L: **Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription.** *Proc Natl Acad Sci USA* 2002, **99:**11593-11598.
11. Cogoi S, Xodo L: **G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription.** *Nucleic Acids Res* 2006, **34:**2536-2549.
12. Cogoi S, Paramasivam M, Spolaore B, Xodo L: **Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins.** *Nucleic Acids Res* 2008, **36:**3765-3680.
13. Rankin S, Reszka A, Huppert J, Zloh M, Parkinson G, Todd A, Ladame S, Balasubramanian S, Neidle S: **Putative DNA quadruplex formation within the human c-kit oncogene.** *J Am Chem Soc* 2005, **127:**10584-10589.
14. Bejugam M, Sewitz S, Shirude P, Rodriguez R, Shahid R, Balasubramanian S: **Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression.** *J Am Chem Soc* 2007, **129:**12926-12927.
15. Sun D, Guo K, Rusche J, Hurley L: **Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents.** *Nucleic Acids Res* 2005, **33:**6070-6080.
16. Palumbo S, Memmott R, Uribe D, Krotova-Khan Y, Hurley L, Ebbinghaus S: **A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity.** *Nucleic Acids Res* 2008, **36:**1755-1769.
17. Xu Y, Sugiyama H: **Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb).** *Nucleic Acids Res* 2006, **34:**949-954.
18. Dexheimer T, Sun D, Hurley L: **Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter.** *J Am Chem Soc* 2006, **128:**5404-5415.
19. Dai J, Dexheimer T, Chen D, Carver M, Ambrus A, Jones R, Yang D: **An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution.** *J Am Chem Soc* 2006, **128:**1096-1098.
20. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps H, Pluckthun A: **In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei.** *Proc Natl Acad Sci USA* 2001, **98:**8572-8577.
21. Fry M: **Tetraplex DNA and its interacting proteins.** *Front Biosci* 2007, **12:**4336-4351.
22. Paramasivam M, Membrino A, Cogoi S, Fukuda H, Nakagama H, Xodo L: **Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription.** *Nucleic Acids Res* 2009, **37:**2841-2853.
23. Zaug A, Podell E, Cech T: **Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro.** *Proc Natl Acad Sci USA* 2005, **102:**10864-10869.
24. Fry M, Loeb L: **Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n.** *J Biol Chem* 1999, **274:**12797-12802.
25. Neidle S, Parkinson GN: **Quadruplex DNA crystal structures and drug design.** *Biochimie* 2008, **90:**1184-1196.

26. Franceschin M: **G-quadruplex DNA structures and organic chemistry: more than one connection.** *Eur J Org Chem* 2009, **14:**2225-2238.
27. Monchaud D, Teulade-Fichou MP: **A hitchhiker's guide to G-quadruplex ligands.** *Org Biomol Chem* 2008, **6:**627-636.
28. De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou JF, Mergny JL: **Targeting telomeres and telomerase.** *Biochimie* 2008, **90:**131-155.
29. Gilbert DE, Feigon J: **Multistranded DNA structures.** *Curr Opin Struct Biol* 1999, **9:**305.
30. Simonsson T: **G-quadruplex DNA structures-variations on a theme.** *Biol Chem* 2001, **382:**621-628.
31. Parkinson GN: **Fundamentals of quadruplex structures.** In *Quadruplex nucleic acids* Edited by: Neidle S, Balasubramanian S. Cambridge, UK: RSC Publishing; 2006:1-30.
32. Patel DJ, Phan AT, Kuryavyi V: **Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics.** *Nucl Acids Res* 2007, **35:**7429-7455.
33. Qin Y, Hurley LH: **Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions.** *Biochimie* 2008, **90:**1149-1171.
34. Risitano A, Fox KR: **Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexes.** *Biochemistry* 2003, **42:**6507-6513.
35. Vorlickova M, Bednarova K, Kejnovska I, Kypr J: **Intramolecular and intermolecular guanine quadruplexes of DNA in aqueous salt and ethanol solutions.** *Biopolymers* 2007, **86:**1-10.
36. Rachwal PA, Brown T, Fox KR: **Sequence effects of single base loops in intramolecular DNA quadruplex DNA.** *FEBS Lett* 2007, **581:**1657-1660.
37. Bugaut A, Balasubramanian S: **A sequence-independent study of the in uence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes.** *Biochemistry* 2008, **47:**689-697.
38. Smargiasso N, Rosu F, Hsia W, Colson P, Baker ES, Bowers MT, De Pauw E, Gabelica V: **G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation.** *J Am Chem Soc* 2008, **130:**10208-10216.
39. Kumar N, Maiti S: **A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes.** *Nucleic Acids Res* 2008, **36:**5610-5622.
40. Huppert JL, Balasubramanian S: **Prevalence of quadruplexes in the human genome.** *Nucl Acids Res* 2005, **33:**2908-2916.
41. Todd AK, Johnston M, Neidle S: **Highly prevalent putative quadruplex sequence motifs in human DNA.** *Nucl Acids Res* 2005, **28:**2901-2907.
42. Rawal P, Kummarasetti VBR, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S: **Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation.** *Genome Res* 2006, **16:**644-655.
43. Du Z, Kong P, Gao Y, Li N: **Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription.** *Genome Res* 2008, **18:**233-241.
44. Huppert JL, Balasubramanian S: **G-quadruplexes in promoters throughout the human genome.** *Nucl Acids Res* 2007, **35:**406-413.
45. Eddy J, Maizels N: **Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes.** *Nucl Acids Res* 2008, **36:**1321-1333.
46. Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, Johnson FB: **Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae.** *Nucl Acids Res* 2008, **36:**144-146.
47. Yadav VK, Kappukalayil A, Mani P, Kulshrestha R, Chowdhury S: **QuadBase: genome-wide database of G4 DNA-occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes.** *Nucl Acids Res* 2008, **36:**381-385.
48. Zhang R, Lin Y, Zhang CT: **Greglist: a database listing potential G-quadruplex regulated genes.** *Nucl Acids Res* 2008, **36:**D372-D376.
49. Berman HM, Olson WK, Beveridge D, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B: **The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids.** *Biophys J* 1992, **63:**751-759.
50. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov S, Bourne PE: **The Protein Data Bank.** *Nucl Acids Res* 2000, **28:**235-242.
51. Parisien M, Major F: **The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.** *Nature* 2008, **452:**51-55.
52. Major F, Turcotte F, Gutheret D, Lapalme G, Fillion E, Cedergren R: **The combination of symbolic and numerical computation for three-dimensional modeling of RNA.** *Science* 1991, **253:**1255-1260.
53. Major F: **Building three-dimensional ribonucleic acid structures.** *Comput Sci Eng* 2003, **5:**44-53.
54. Guschlbauer W, Chantot JF, Thiele D: **Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA.** *J Biomol Struct Dyn* 1990, **8:**491-511.
55. Davies JT: **G-Quartets 40 Years Later: From 5 -GMP to Molecular Biology and Supramolecular Chemistry.** *Angew Chem* 2004, **43:**668-698.
56. Webba da Silva M: **Geometric formalism for DNA quadruplex folding.** *Chemistry* 2007, **13:**9738-9745.
57. Sun D, Guo K, Rusche J, Hurley L: **The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents.** *Mol Cancer Ther* 2008, **7:**880-889.
58. Hazel P, Huppert JL, Balasubramanian S, Neidle S: **Loop-length dependent folding of G-quadruplexes.** *J Am Chem Soc* 2004, **126:**16405-16415.
59. Rachwal PA, Findlow IS, Werner JM, Brown T, Fox KR: **Intramolecular DNA quadruplexes with different arrangements of short and long loops.** *Nucl Acids Res* 2007, **35:**4214-4222.
60. Kumar N, Sahoo B, Maiti S, Maiti S: **Effect of loop length variation on quadruplex-Watson Crick duplex competition.** *Nucleic Acids Res* 2008, **36:**4433-4442.
61. Arora A, Maiti S: **Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents.** *J Phys Chem B* 2009, **113:**8784-8792.
62. Ambrus A, Chen D, Dai J, Jones RA, Yang D: **Solution structure of the biologically relevant G-quadruplex element in the human c-MYK promoter. Implications for G-quadruplex stabilization.** *Biochemistry* 2005, **44:**2048-2058.
63. Guo K, Pourpak A, Beetz-Rogers K, Gokhale V, Sun D, Hurley LH: **Formation of pseudo-symmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene.** *J Am Chem Soc* 2007, **129:**10220-10228.
64. Spooner J, Spackova N: **Molecular dynamics simulations and their application to four-stranded DNA.** *Methods* 2007, **43:**278-290.
65. Fogolari F, Corazza A, Viglino P, Zuccato P, Pieri L, Faccioli P, Bellotti V, Esposito G: **Molecular dynamics simulation suggests possible interaction patterns at early steps of 2-microglobulin aggregation.** *Biophys J* 2007, **92:**1673-1681.
66. Mackerell AD, Banavali N: **All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution.** *J Comp Chem* 2000, **21:**105-120.
67. Foloppe N, Mackerell AD: **All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data.** *J Comp Chem* 2000, **21:**86-104.
68. Redman JE, Granadino-Roldán JM, Schouten JA, S L, Reszka AP, Neidle S, Balasubramanian S: **Recognition and discrimination of DNA quadruplexes by acridine-peptide conjugates.** *Org Biomol Chem* 2009, **7:**76-84.
69. Foloppe N, Mackerell AD: **Molecular docking study of binding of TMPyP4 to a bimolecular human telomeric G-quadruplex.** *Nucleic Acids Symp Ser (Oxf)* 2008, **52:**173-174.
70. Phan AT, Kuryavyi V, Gaw HY, Patel DJ: **Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter.** *Nat Chem Biol* 2005, **1:**167-173.
71. Parkinson GN, Ghosh R, Neidle S: **Structural basis for binding of porphyrin to human telomeres.** *Biochemistry* 2007, **46:**2390-2397.
72. Kleywegt GJ: **Crystallographic refinement of ligand complexes.** *Acta Cryst D* 2000, **63:**94-100.
73. Sayers E, Barrett T, Benson D, Bryant S, Canese K, Chetvernin V, Church D, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer L, Helm-

5. APPENDIX

berg W, Kapustin Y, Landsman D, Lipman D, Madden T, Maglott D, Miller V, Mizrachi I, Ostell J, Pruitt K, Schuler G, Sequeira E, Sherry S, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova T, Wagner L, Yaschenko E, Ye J: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2009, **37:**D5-D15.

74. Guha R, Howard M, Hutchison G, Murray-Rust P, Rzepa H, Steinbeck C, Wegner J, EL W: **The Blue Obelisk - interoperability in chemical informatics.** *J Chem Inf Model* 2006, **46:**991-998.
75. Gasteiger J, Marsili M: **A new model for calculating atomic charges in molecules.** *Tetrahedron Lett* 1978, **34:**3181-3184.
76. Lang P, Brozell S, Mukherjee S, Pettersen E, Meng E, Thomas V, Rizzo R, Case D, James T, Kuntz I: **DOCK 6: Combining Techniques to Model RNA-Small Molecule Complexes.** *RNA* 2009, **15:**991-998.
77. Todd AK, Neidle S: **The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements.** *Nucl Acids Res* 2008, **36:**2700-2704.
78. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of the human cancer genes.** *Nature Rev Cancer* 2004, **4:**177-183.
79. Stratton MR, J CP, Futreal PA: **The cancer genome.** *Nature* 2009, **458:**719-724.
80. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolberger C, Berman HM: **A standard reference frame for the description of nucleic acid base-pair geometry.** *J Mol Biol* 2001, **313:**229-237.
81. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K: **NAMD2: greater scalability for parallel molecular dynamics.** *J Comp Phys* 1999, **151:**283-312.

5.1.2 NMR Studies of the APE/REF-1 N-terminus Interactions

Another project in which significant contributions were made but not mentioned in the main text of this thesis was "NMR studies of the APE1 N-terminus interactions". An overview of the work done is described below.

Human apurinic (apyrimidinic) endonuclease/redoxfactor1 (APE/Ref-1) is a ubiquitously expressed multifunctional protein (196, 197). It plays a central role in base excision repair (BER) pathways of DNA lesions and also functions as a redox transcriptional coactivator. These two biologcal activities of the molecule are located in two functionally distinct protein domains, the N-terminal region principally denoted by redox activity and the C-terminal excising enzymatic activity on abasic sites of DNA (198, 199).

Structural studies conducted on this 318aa, 35.6 kDa molecule have excluded the first 38 residues because the initial investigations on the full length protein by both crystallography (200) and NMR (201) showed a highly disordered conformation in the 1-44 region. A recent proteomic study (**?**) to identify and characterize the APE1 interactome stated a stable interaction of the first 33 N-terminal residues with nucleophosmin and rRNA. The study suggested an important role of this segment in RNA quality control process and in fine-tuning different functions of the full length APE1. We attempted to characterize these interactions by using nuclear magnetic resonance spectroscopy (NMR).

Standard NMR experiments were conducted on the unlabelled 33-residue N-terminal peptide of APE1. Although a complete assignment of the peptide was not possible, our investigation suggested that the peptide forms an α-Helix in the G8-E20 region. However for the confirmation and completion of the assignments we required a single (15N) or double labelled (13C, 15N) product. In this view, the expression and purification of the labelled peptide was taken up.

The expression and subsequent purification was hampered by difficulties such as unfavourable protease digestion of the product, low expression in labelled media and lack of a direct purification strategy. We therefore optimized a protocol (by optimizing parameters like incubation temperature, induction OD, concentration of IPTG and incubation time) for the expression and purification of N-15 labelled 48 residue N-terminal peptide in Minimal Media. A final yield of 10-12mg/Litre in labelled media

was obtained. The quality of the purified peptide was checked with mass spectrometry and NMR and it was found to be amenable for further NMR analyses. The optimized and followed protocol for the expression and purification of the N-15 peptide is described below.

1) Transform *E.coli* (1-2.5 μl in 50 μL of Bl21DE3). Incubate on ice for 30 minutes followed by incubation at 42°C for 45 seconds. Add 1ml SOC medium and incubate at 37°C for one hour.

2) Inoculate 700 μL on a prepared agar plate with 100μg Ampicillin. Overnight incubation at 37°C.

3) Transfer the plate to °C in the morning; inoculate four colonies into 15 ml of Luria-Bertani (LB) broth with 1% glucose and 100 μg/ml Ampicilin (in evening), leave for overnight incubation at 37°C

4) The next morning take 10 ml of inoculum and inoculate in 1L LB plus 10 g of Glucose and 100μg/ml ampicilin. Divide in to two conical flasks of 500 mL. Leave for 2 hours inside a shaker incubator set at 230 rpm at 37°C.

5) When the OD (optical density) at 600nm reaches 0.8 or 0.9 take the solution and centrifuge with (GS3 rotor used) with 4 bottles at 3000 rpm for 20 minutes (10-20°C).

6) Resuspend the pellet with pre-warmed minimal media (Minimal media constituents- 100 ml M9 medium (10x), 4 g glucose, 1 ml 1 M $MgSO_4$, 0.3 ml 1 M $CaCl_2$, 1 ml Thiamin (1 mg/ml), 50 μg/ml ampicilin).

7) Wait 30 minutes at 37°C and make an induction with 2mM IPTG (0.47768 g in 2ml water and divide equally in to two flasks)

7) Leave overnight for at least 19 hours

8) Centrifuge immediately at 3000 rpm for 20 minutes at 4°C and proceed rapidly to purification.

Purification steps followed were;

1) Metal Affinity Chromatography Column: HisTrap HP 1 ml Binding buffer: 20 mM phosphate pH 7.4, 500 mM NaCl, 40 mM imidazole Elution buffer: 20 mM phosphate pH 7.4, 500 mM NaCl, 500 mM imidazole Gradient: 1 ml/min 0-100%

2) Dialysis Column: HisTrap HP 1 ml Binding buffer: 20 mM phosphate pH 7.4, 500 mM NaCl, 40 mM imidazole Elution buffer: 20 mM phosphate pH 7.4, 500 mM NaCl, 500 mM imidazole Gradient: 1 ml/min 0-100% B in 30 minutes

3) Cation exchange Column: HiTrap SP XL 1 ml Buffer A: 50 mM Na phosphate pH 7.0 Buffer B: 50 mM Na phosphate pH 7.0, 1 M NaCl Gradient: 1 ml/min 0-100% B in 10 minutes

Fig. 5.1 shows the final SDS-PAGE gel with the purified 48 residue N-terminal APE/Ref-1 peptide

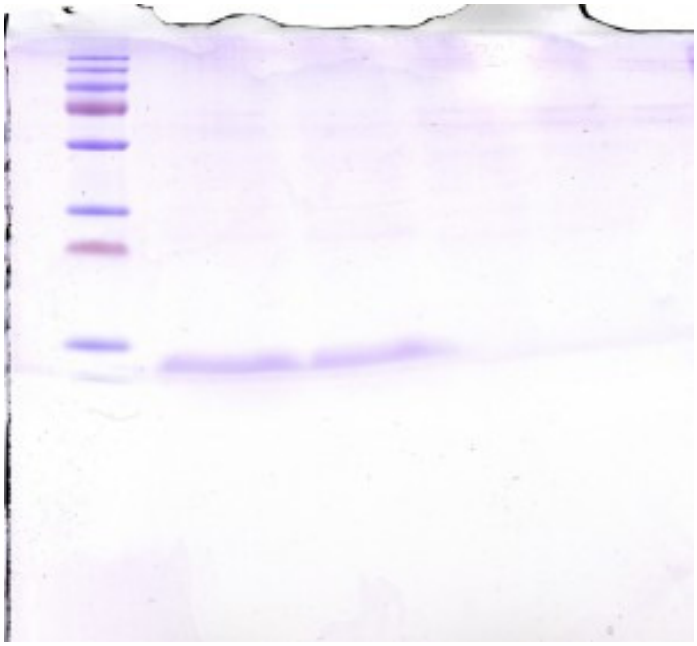


Fig. 5.1 SDS PAGE gel showing the single band of the purified final product obtained by the optimized protocol described above - The final yield of the N-15 labelled, 48 residue peptide was 10 mg/Litre of minimal media

References

[1] P H Yancey, M E Clark, S C Hand, R D Bowlus, and G N Somero. **Living with water stress: evolution of osmolyte systems.** *Science*, **217**(4566):1214–1222, 1982. 1, 7

[2] M B Burg. **Molecular basis of osmotic regulation.** *American Journal of Physiology*, **268**(6 Pt 2):F983–F996, 1995. 1, 3

[3] Paul H Yancey. **Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses.** *The Journal of experimental biology*, **208**(Pt 15):2819–30, August 2005. 1

[4] Maurice B Burg and Joan D Ferraris. **Intracellular organic osmolytes: function and regulation.** *The Journal of biological chemistry*, **283**(12):7309–13, March 2008. 1

[5] P H Yancey, A L Fyfe-Johnson, R H Kelly, V P Walker, and M T Auñón. **Trimethylamine oxide counteracts effects of hydrostatic pressure on proteins of deep-sea teleosts.** *The Journal of experimental zoology*, **289**(3):172–176, 2001. 1, 7

[6] A Garcia-Perez and M B Burg. **Renal medullary organic osmolytes.** *Physiological Reviews*, **71**(4):1081–1115, 1991. 1

[7] Peter W Hochachka and George N Somero. *Biochemical adaptation: mechanism and process in physiological evolution*, **30**. Oxford University Press, 2002. 1

[8] Daniel Harries and Jörg Rösgen. **A practical guide on how osmolytes modulate macromolecular properties.** *Methods in cell biology*, **84**(07):679–735, January 2008. 1

[9] D W Bolen and I V Baskakov. **The osmophobic effect: natural selection of a thermodynamic force in protein folding.** *Journal of molecular biology*, **310**(5):955–63, July 2001. 1, 5, 6

[10] Raj Kumar. **Role of naturally occurring osmolytes in protein folding and stability.** *Archives of biochemistry and biophysics*, **491**(1-2):1–6, November 2009. 1, 4, 6

[11] Matthew Auton and D Wayne Bolen. **Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale.** *Biochemistry*, **43**(5):1329–1342, 2004. 1, 5

[12] Matthew Auton and D Wayne Bolen. **Predicting the energetics of osmolyte-induced protein folding/unfolding**. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(42):15065–15068, 2005. 1, 5

[13] Y Liu and D W Bolen. **The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes.** *Biochemistry*, **34**(39):12884–12891, 1995. 1, 5

[14] C M Dobson. **Protein misfolding, evolution and disease.** *Trends in biochemical sciences*, **24**(9):329–32, September 1999. 2, 9, 12

[15] Christopher M Dobson. **Protein folding and misfolding.** *Nature*, **426**(6968):884–890, 2003. 2, 18

[16] I Baskakov and D W Bolen. **Forcing thermodynamically unfolded proteins to fold.** *The Journal of Biological Chemistry*, **273**(9):4831–4834, 1998. 3

[17] I V Baskakov, R Kumar, G Srinivasan, Y S Ji, D W Bolen, and E B Thompson. **Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor.** *The Journal of Biological Chemistry*, **274**(16):10693–10696, 1999. 3

[18] R Kumar, J C Lee, D W Bolen, and E B Thompson. **The conformation of the glucocorticoid receptor af1/tau1 domain induced by osmolyte binds co-regulatory proteins.** *The Journal of Biological Chemistry*, **276**(21):18146–18152, 2001. 3

[19] R Kumar, J M Serrette, S H Khan, A L Miller, and E B Thompson. **Effects of different osmolytes on the induced folding of the N-terminal activation domain (AF1) of the glucocorticoid receptor.** *Archives of Biochemistry and Biophysics*, **465**(2):452–460, 2007. 3

[20] James Reid, Sharon M Kelly, Kate Watt, Nicholas C Price, and Iain J McEwan. **Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions**. *J Biol Chem*, **277**(22):20079–86, May 2002. 3

[21] Raj Kumar, Justin M Serrette, and E Brad Thompson. **Osmolyte-induced folding enhances tryptic enzyme activity.** *Archives of Biochemistry and Biophysics*, **436**(1):78–82, 2005. 3

[22] Luis Marcelo F Holthauzen and D Wayne Bolen. **Mixed osmolytes : The degree to which one osmolyte affects the protein stabilizing ability of another**. *Society*, pages 293–298, 2007. 3, 7

[23] Laishram R Singh, Nitesh Kumar Poddar, Tanveer Ali Dar, Raj Kumar, and Faizan Ahmad. **Protein and DNA destabilization by osmolytes: the other side of the coin.** *Life sciences*, **88**(3-4):117–25, January 2011. 3

[24] Timothy O Street, D Wayne Bolen, and George D Rose. **A molecular mechanism for osmolyte-induced protein stability.** *Proceedings of the National Academy of Sciences of the United States of America*, **103**(38):13997–4002, September 2006. 5

REFERENCES

[25] D Wayne Bolen and George D Rose. **Structure and energetics of the hydrogen-bonded backbone in protein folding.** *Annual Review of Biochemistry*, **77**(1):339–362, 2008. 6, 7

[26] J J O'Dowd, D J Robins, and D J Miller. **Detection, characterisation, and quantification of carnosine and other histidyl derivatives in cardiac and skeletal muscle**. *Biochim Biophys Acta*, **967**(2):241–9, Nov 1988. 7

[27] R Kohen, Y Yamamoto, K C Cundy, and B N Ames. **Antioxidant activity of carnosine, homocarnosine, and anserine present in muscle and brain.** *Proceedings of the National Academy of Sciences of the United States of America*, **85**(9):3175–3179, 1988. 8

[28] G A McFarland and R Holliday. **Retardation of the senescence of cultured human diploid fibroblasts by carnosine.** *Experimental Cell Research*, **212**(2):167–175, 1994. 8

[29] S E Gariballa and a J Sinclair. **Carnosine: physiological properties and therapeutic potential.** *Age and ageing*, **29**(3):207–10, May 2000. 8

[30] A R Hipkiss, J Michaelis, and P Syrris. **NONENZYMATIC GLYCOSYLATION OF THE DIPEPTIDE L-CARNOSINE, A POTENTIAL ANTIPROTEIN-CROSS-LINKING AGENT**. *FEBS Letters*, **371**(1):81–85, 1995. 8

[31] C Brownson and A R Hipkiss. **Carnosine reacts with a glycated protein.** *Free Radical Biology & Medicine*, **28**(10):1564–1570, 2000. 8

[32] A Boldyrev, R Song, D Lawrence, and D O Carpenter. **Carnosine protects against excitotoxic cell death independently of effects on reactive oxygen species.** *Neuroscience*, **94**(2):571–577, 1999. 8

[33] Alexander Boldyrev, Elena Bulygina, Toomas Leinsoo, Irina Petrushanko, Shiori Tsubone, and Hiroki Abe. **Protection of neuronal cells against reactive oxygen species by carnosine and related compounds.** *Comparative biochemistry and physiology Part B Biochemistry molecular biology*, **137**(1):81–88, 2004. 8

[34] Rinat Tabakman, Hao Jiang, Robert A Levine, Ron Kohen, and Philip Lazarovici. **Apoptotic characteristics of cell death and the neuroprotective effect of homocarnosine on pheochromocytoma PC12 cells exposed to ischemia.** *Journal of Neuroscience Research*, **75**(4):499–507, 2004. 8

[35] J E Preston, A R Hipkiss, D T Himsworth, I A Romero, and J N Abbott. **Toxic effects of beta-amyloid(25-35) on immortalised rat brain endothelial cell: protection by carnosine, homocarnosine and beta-alanine.** *Neuroscience Letters*, **242**(2):105–108, 1998. 8, 19

[36] Qiuli Fu, Haibin Dai, Weiwei Hu, Yanying Fan, Yao Shen, Weiping Zhang, and Zhong Chen. **Carnosine Protects Against Aβ42-induced Neurotoxicity in Differentiated Rat PC12 Cells**. *Cellular and Molecular Neurobiology*, **28**(2):307–316, 2007. 8

[37] H Abe. **Role of histidine-related compounds as intracellular proton buffering constituents in vertebrate muscle.** *Biochemistry Biokhimiia*, **65**(7):757–765, 2000. 8

[38] A R Hipkiss, J E Preston, D T Himsworth, V C Worthington, M Keown, J Michaelis, J Lawrence, A Mateen, L Allende, P A Eagles, and N J Abbott. **Pluripotent protective effects of carnosine, a naturally occurring dipeptide.** *Annals Of The New York Academy Of Sciences*, **854**:37–53, 1998. 8, 19

[39] Hong Yan and John J Harding. **Carnosine inhibits modifications and decreased molecular chaperone activity of lens alpha-crystallin induced by ribose and fructose 6-phosphate.** *Molecular Vision*, **12**:205–214, 2006. 8

[40] Hong Yan and John J Harding. **Carnosine protects against the inactivation of esterase induced by glycation and a steroid.** *Biochimica et Biophysica Acta*, **1741**(1-2):120–126, 2005. 8

[41] Norbert W Seidler, George S Yeargans, and Timothy G Morgan. **Carnosine disaggregates glycated alpha-crystallin: an in vitro study.** *Archives of Biochemistry and Biophysics*, **427**(1):110–115, 2004. 8

[42] Robert J Linhardt. **2003 Claude S. Hudson Award address in carbohydrate chemistry. Heparin: structure and activity.**, 2003. 8

[43] S E Stringer and J T Gallagher. **Heparan sulphate.** *The international journal of biochemistry cell biology*, **29**(5):709–714, 1997. 8

[44] D R Coombe and W C Kett. **Heparan sulfate-protein interactions: therapeutic potential through structure-function insights.** *Cellular and molecular life sciences : CMLS*, **62**(4):410–24, February 2005. 8

[45] F Meng, Y Park, and H Zhou. **Role of proline, glycerol, and heparin as protein folding aids during refolding of rabbit muscle creatine kinase.** *The international journal of biochemistry & cell biology*, **33**(7):701–9, July 2001. 8, 19

[46] Y-D Park, B-N Wu, W-X Tian, and H-M Zhou. **Effects of osmolytes on unfolding of chicken liver fatty acid synthase.** *Biochemistry. Biokhimiia*, **67**(8):914–7, August 2002. 8

[47] Hironobu Naiki, Suguru Yamamoto, Kazuhiro Hasegawa, Itaru Yamaguchi, Yuji Goto, and Fumitake Gejyo. **Molecular interactions in the formation and deposition of beta2-microglobulin-related amyloid fibrils**. *Amyloid*, **12**(1):15–25, Mar 2005. 8, 84

[48] Suguru Yamamoto, Itaru Yamaguchi, Kazuhiro Hasegawa, Shinobu Tsutsumi, Yuji Goto, Fumitake Gejyo, and Hironobu Naiki. **Glycosaminoglycans enhance the trifluoroethanol-induced extension of beta 2-microglobulin-related amyloid fibrils at a neutral pH.** *Journal of The American Society Of Nephrology*, **15**(1):126–133, 2004. 8, 21, 84

[49] Annalisa Relini, Silvia De Stefano, Silvia Torrassa, Ornella Cavalleri, Ranieri Rolandi, Alessandra Gliozzi, Sofia Giorgetti, Sara Raimondi, Loredana Marchese,

Laura Verga, Antonio Rossi, Monica Stoppini, and Vittorio Bellotti. **Heparin strongly enhances the formation of beta2-microglobulin amyloid fibrils in the presence of type I collagen.** *The Journal of biological chemistry*, **283**(8):4912–20, February 2008. 8, 9, 21, 84

[50] Neda Motamedi-Shad, Elodie Monsellier, Silvia Torrassa, Annalisa Relini, and Fabrizio Chiti. **Kinetic analysis of amyloid formation in the presence of heparan sulfate: faster unfolding and change of pathway**. *J Biol Chem*, **284**(43):29921–34, Oct 2009. 8, 84

[51] Peter G Wolynes. **Energy landscapes and solved protein-folding problems.** *hilosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences*, **363**(1827):453–464; discussion 464–467, 2005. 9

[52] K A Dill and H S Chan. **From Levinthal to pathways to funnels**. *Nat Struct Biol*, **4**(1):10–9, Jan 1997. 10

[53] Valerie Daggett and Alan Fersht. **The present view of the mechanism of protein folding**. *Nature Rev Mol Cell Biol*, **4**(June):497–502, 2003. 10

[54] Alan R Fersht. **Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism**. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(4):1525–1529, 2000. 10

[55] C P Schultz. **Illuminating folding intermediates.**, 2000. 11

[56] Dennis J Selkoe. **Folding proteins in fatal ways.** *Nature*, **426**(6968):900–904, 2003. 11

[57] Monica Bucciantini, Elisa Giannoni, Fabrizio Chiti, Fabiana Baroni, Lucia Formigli, Jesús Zurdo, Niccolò Taddei, Giampietro Ramponi, Christopher M Dobson, and Massimo Stefani. **Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases.** *Nature*, **416**(6880):507–511, 2002. 11

[58] Giampaolo Merlini and Vittorio Bellotti. **Molecular mechanisms of amyloidosis.** *The New England journal of medicine*, **349**(6):583–96, August 2003. 12

[59] Jean D Sipe, Merrill D Benson, Joel N Buxbaum, Shu-Ichi Ikeda, Giampaolo Merlini, Maria J M Saraiva, and Per Westermark. **Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis.** *Amyloid the international journal of experimental and clinical investigation the official journal of the International Society of Amyloidosis*, **17**(3-4):101–104, 2010. 12

[60] Massimo Stefani and Christopher M Dobson. **Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution.** *Journal of molecular medicine (Berlin, Germany)*, **81**(11):678–99, November 2003. 12, 18

[61] Vladimir N Uversky and Anthony L Fink. **Conformational constraints for amyloid fibrillation: the importance of being unfolded.** *Biochimica et Biophysica Acta*, **1698**(2):131–153, 2004. 12

[62] S E Radford, W S Gosal, and G W Platt. **Towards an understanding of the structural molecular mechanism of ₋2-microglobulin amyloid formation in vitro**. *Biochimica et Biophysica Acta*, **1753**:51–63, 2005. 12

[63] G Wistow. **Identification of lens crystallins: a model system for gene recruitment.** *Methods in Enzymology*, **224**:563–575, 1993. 13

[64] J J Harding. **Post-translational modification of lens proteins in cataract.** *Lens and Eye Toxicity Research*, **8**(2-3):245–250, 1991. 13

[65] J Horwitz. **Alpha-crystallin can function as a molecular chaperone.** *Proceedings of the National Academy of Sciences of the United States of America*, **89**(21):10449–10453, 1992. 13

[66] John A Carver, Agata Rekas, David C Thorn, and Mark R Wilson. **Small heat-shock proteins and clusterin: intra- and extracellular molecular chaperones with a common mechanism of action and function?** *IUBMB Life*, **55**(12):661–668, 2003. 13

[67] H Ecroyd and John A Carver. **Crystallin proteins and amyloid fibrils.** *Cellular and molecular life sciences CMLS*, **66**(1):62–81, 2009. 13

[68] Joseph Horwitz. **Alpha-crystallin**. *Exp Eye Res*, **76**(2):145–53, Feb 2003. 13

[69] Barry K Derham and John J Harding. **Effects of modifications of alpha-crystallin on its chaperone and other properties.** *The Biochemical journal*, **364**(Pt 3):711–717, 2002. 13

[70] K Kato, H Shinohara, N Kurobe, Y Inaguma, K Shimizu, and K Ohshima. **Tissue distribution and developmental profiles of immunoreactive alpha B crystallin in the rat determined with a sensitive immunoassay system.** *Biochimica et Biophysica Acta*, **1074**(1):186–190, 1991. 13

[71] Roger J W Truscott. **Age-related nuclear cataract-oxidation is the key.** *Experimental Eye Research*, **80**(5):709–725, 2005. 14

[72] Heath Ecroyd and John A Carver. **The effect of small molecules in modulating the chaperone activity of alphaB-crystallin against ordered and disordered protein aggregation.** *The FEBS journal*, **275**(5):935–947, 2008. 14

[73] Y Riaz, J S Mehta, R Wormald, J R Evans, A Foster, T Ravilla, and T Snellingen. **Surgical interventions for age-related cataract.** *Cochrane database of systematic reviews Online*, **Volume**(2):CD001323, 2006. 14

[74] Y Hiratsuka, K Ono, and A Kanai. **The present state of blindness in the world**. *Nippon Ganka Gakkai Zasshi. Acta Societatis Ophthalmologicae Japonicae*, **105**(6):369–373, 2001. 14

REFERENCES

[75] Sarah Meehan, Yoke Berry, Ben Luisi, Christopher M Dobson, John A Carver, and Cait E MacPhee. **Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation.** *The Journal of Biological Chemistry*, **279**(5):3413–3419, 2004. 14, 33, 54, 57

[76] Aileen Sandilands, Aileen M Hutcheson, Heather A Long, Alan R Prescott, Gijs Vrensen, Jana Löster, Norman Klopp, Raimund B Lutz, Jochen Graw, Shigeo Masaki, Christopher M Dobson, Cait E MacPhee, and Roy A Quinlan. **Altered aggregation properties of mutant γ-crystallins cause inherited cataract**. *the The European Molecular Biology Organization Journal*, **21**(22):6005–6014, 2002. 14

[77] F S Esch, P S Keim, E C Beattie, R W Blacher, A R Culwell, T Oltersdorf, D McClure, and P J Ward. **Cleavage of amyloid beta peptide during constitutive processing of its precursor.** *Science*, **248**(4959):1122–1124, 1990. 15

[78] C Haass, M G Schlossmacher, A Y Hung, C Vigo-Pelfrey, A Mellon, B L Ostaszewski, I Lieberburg, E H Koo, D Schenk, D B Teplow, and Et Al. **Amyloid beta-peptide is produced by cultured cells during normal metabolism**. *Nature*, **359**(6393):322–325, 1992. 15

[79] N Suzuki, T T Cheung, X D Cai, A Odaka, L Otvos, Jr, C Eckman, T E Golde, and S G Younkin. **An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants**. *Science*, **264**(5163):1336–40, May 1994. 15

[80] X D Cai, T E Golde, and S G Younkin. **Release of excess amyloid beta protein from a mutant amyloid beta protein precursor**. *Science*, **259**(5094):514–6, Jan 1993. 15

[81] Massimo Tabaton, Xiongwei Zhu, George Perry, Mark A Smith, and Luca Giliberto. **Signaling effect of amyloid-beta(42) on the processing of AbetaPP.** *Experimental Neurology*, **221**(1):18–25, 2010. 15

[82] Kun Zou, Jian-Sheng Gong, Katsuhiko Yanagisawa, and Makoto Michikawa. **A novel function of monomeric amyloid beta-protein serving as an antioxidant molecule against metal-induced oxidative damage.** *Journal of Neuroscience*, **22**(12):4833–4841, 2002. 15

[83] Rozena Baruch-Suchodolsky and Bilha Fischer. **Abeta40, either soluble or aggregated, is a remarkably potent antioxidant in cell-free oxidative systems.** *Biochemistry*, **48**(20):4354–4370, 2009. 15

[84] Jiahong Yao, Hui Li, and Zhenxin Zhang. **A case-control study on family history of Alzheimer's disease**. *Zhonghua liu xing bing xue za zhi Zhonghua liuxingbingxue zazhi*, **23**(1):54–56, 2002. 15

[85] U Igbavboa, G Y Sun, G A Weisman, Yan He, and W G Wood. **Amyloid beta-protein stimulates trafficking of cholesterol and caveolin-1 from the plasma membrane to the Golgi complex in mouse primary astrocytes.** *Neuroscience*, **162**(2):328–338, 2009. 15

[86] Bryan Maloney and Debomoy K Lahiri. **The Alzheimer's amyloid β-peptide (Aβ) binds a specific DNA Aβ-interacting domain (AβID) in the APP, BACE1, and APOE promoters in a sequence-specific manner: Characterizing a new regulatory motif.** *Gene*, **488**(1-2):1–12, 2011. 15

[87] Jason A Bailey, Bryan Maloney, Yuan-Wen Ge, and Debomoy K Lahiri. **Functional activity of the novel Alzheimer's amyloid β-peptide interacting domain (AβID) in the APP and BACE1 promoter sequences and implications in activating apoptotic genes and in amyloidogenesis.** *Gene*, 2011. 15

[88] Stephanie J Soscia, James E Kirby, Kevin J Washicosky, Stephanie M Tucker, Martin Ingelsson, Bradley Hyman, Mark A Burton, Lee E Goldstein, Scott Duong, Rudolph E Tanzi, and Robert D Moir. **The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide.** *PLoS ONE*, **5**(3):e9505, 2010. 15

[89] B A Yankner and M M Mesulam. **Seminars in medicine of the Beth Israel Hospital, Boston: b- Amyloid and the pathogenesis of Alzheimer's disease**, 1991. 15

[90] J Hardy and D Allsop. **Amyloid deposition as the central event in the aetiology of Alzheimer's disease.** *Trends in Pharmacological Sciences*, **12**(10):383–388, 1991. 15

[91] J Hardy and G Higgins. **Alzheimer's disease: the amyloid cascade hypothesis**. *Science*, **256**(5054):184–185, 1992. 15

[92] J Hardy and D J Selkoe. **The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics**. *Science*, **297**(5580):353–356, 2002. 15

[93] Eric Karran, Marc Mercken, and Bart De Strooper. **The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics.** *Nature reviews. Drug discovery*, **10**(September), August 2011. 15

[94] Christiane Reitz, Carol Brayne, and Richard Mayeux. **Epidemiology of Alzheimer disease**. *Nat Rev Neurol*, **7**(3):137–52, Mar 2011. 15

[95] Ron Brookmeyer, E Johnson, K Ziegler-Graham, and H M Arrighi. **FORECASTING THE GLOBAL BURDEN OF ALZHEIMER ' S DISEASE FORECASTING THE GLOBAL BURDEN OF ALZHEIMER ' S DISEASE**. *Public Health*, **3**(3):186–191, 2007. 16

[96] R Brookmeyer, S Gray, and C Kawas. **Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset**. *Am J Public Health*, **88**(9):1337–42, Sep 1998. 16

[97] P K Mölsä, R J Marttila, and U K Rinne. **Survival and cause of death in Alzheimer's disease and multi-infarct dementia.** *Acta Neurologica Scandinavica*, **74**(2):103–107, 1986. 16

[98] D Güssow, R Rein, I Ginjaar, F Hochstenbach, G Seemann, A Kottman, and H L Ploegh. **The human beta 2-microglobulin gene. Primary structure and definition of the transcriptional unit.** *The Journal of Immunology*, **139**(9):3132–3138, 1987. 16

[99] J Floege and G Ehlerding. **Beta-2-microglobulin-associated amyloidosis**. *Nephron*, **72**(1):9–26, 1996. 16, 17

[100] A Fuchs, J Jagirdar, and I S Schwartz. **Beta 2-microglobulin amyloidosis (AB2M) in patients undergoing long-term hemodialysis. A new type of amyloid.**, 1987. 17

[101] J M Campistol, A Torras, and J Lopez-Pedret. **Systemic character and visceral involvement of dialysis amyloidosis.**, 1987. 17

[102] M Jadoul. **Dialysis-related amyloidosis: importance of biocompatibility and age.** *Nephrology Dialysis Transplantation*, **13 Suppl 7**:61–64, 1998. 17

[103] M Sunde and C Blake. **The structure of amyloid fibrils by electron microscopy and X-ray diffraction**. *Advances in Protein Chemistry*, **50**:123–159, 1997. 18

[104] M Sunde and C C Blake. **From the globular to the fibrous state: protein structure and structural conversion in amyloid formation.** *Quarterly Reviews of Biophysics*, **31**(1):1–39, 1998. 18

[105] A K Chamberlain, C E MacPhee, J Zurdo, L A Morozova-Roche, H A Hill, C M Dobson, and J J Davis. **Ultrastructural organization of amyloid fibrils by atomic force microscopy.** *Biophysical Journal*, **79**(6):3282–3293, 2000. 18

[106] O S Makin and L C Serpell. **Examining the structure of the mature amyloid fibril.** *Biochemical Society Transactions*, **30**(4):521–525, 2002. 18

[107] J Iñaki Guijarro, Margaret Sunde, Jonathan A Jones, Iain D Campbell, and Christopher M Dobson. **Amyloid fibril formation by an SH3 domain**. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(8):4224–4228, 1998. 18

[108] Fabrizio Chiti, Paul Webster, Niccolò Taddei, Anne Clark, Massimo Stefani, Giampietro Ramponi, and Christopher M Dobson. **Designing conditions for in vitro formation of amyloid protofilaments and fibrils**. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(7):3590–3594, 1999. 18

[109] Marcus Fändrich and Christopher M Dobson. **The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation**. *the The European Molecular Biology Organization Journal*, **21**(21):5682–5690, 2002. 18

[110] I Z Nagy, K Nagy, and G Lustyik. **Protein and water contents of aging brain.** *Experimental Brain Research*, **Suppl 5**:118–122, 1982. 18

[111] I J Conlon, G A Dunn, A W Mudge, and M C Raff. **Extracellular control of cell size**. *Nature Cell Biology*, **3**(10):918–921, 2001. 18

[112] M Al-Habori. **Macromolecular crowding and its role as intracellular signalling of cell volume regulation.** *The international journal of biochemistry cell biology*, **33**(9):844–864, 2001. 18

[113] C H Schein. **Solubility as a function of protein structure and solvent components.** *Biotechnology Nature Publishing Company*, **8**(4):308–317, 1990. 19

[114] Y D Park, W B Ou, T W Yu, and H M Zhou. **Folding pathway for partially folded rabbit muscle creatine kinase.** *Biochemistry and cell biology Biochimie et biologie cellulaire*, **79**(4):479–487, 2001. 19

[115] Sung-Hye Kim, Yong-Bin Yan, and Hai-Meng Zhou. **Role of osmolytes as chemical chaperones during the refolding of aminoacylase.** *Biochemistry and cell biology Biochimie et biologie cellulaire*, **84**(1):30–38, 2006. 19

[116] Y Kusakari, S Nishikawa, S Ishiguro, and M Tamai. **Histidine-like immunoreactivity in the rat retina.** *Current Eye Research*, **16**(6):600–604, 1997. 19

[117] A A Boldyrev. **Carnosine and free-radical defence mechanisms.**, 1994. 19

[118] a M Wang, C Ma, Z H Xie, and F Shen. **Use of carnosine as a natural anti-senescence drug for human beings.** *Biochemistry. Biokhimiia*, **65**(7):869–71, July 2000. 19, 20

[119] J L Cummings, H V Vinters, G M Cole, and Z S Khachaturian. **Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities.** *Neurology*, **51**(1 Suppl 1):S2–17; discussion S65–7, 1998. 19

[120] Francesco Attanasio, Sebastiano Cataldo, Salvatore Fisichella, Silvia Nicoletti, Vincenzo Giuseppe Nicoletti, Bruno Pignataro, Anna Savarino, and Enrico Rizzarelli. **Protective effects of L- and D-carnosine on alpha-crystallin amyloid fibril formation: implications for cataract disease**. *Biochemistry*, **48**(27):6522–31, Jul 2009. 20

[121] Jung Hoon Kang and Kyung Sik Kim. **Enhanced oligomerization of the alpha-synuclein mutant by the Cu,Zn-superoxide dismutase and hydrogen peroxide system.** *Molecules and Cells*, **15**(1):87–93, 2003. 20

[122] Masahiro Kawahara, Hironari Koyama, Tetsuya Nagata, and Yutaka Sadakane. **Zinc, copper, and carnosine attenuate neurotoxicity of prion fragment PrP106-126**. *Metallomics*, **3**:–, 2011. 20

[123] Ulf Lindahl and Jin-ping Li. **Interactions between heparan sulfate and proteins-design and functional implications.** *International review of cell and molecular biology*, **276**(09):105–159, 2009. 20

[124] Joseph R Bishop, Manuela Schuksz, and Jeffrey D Esko. **Heparan sulphate proteoglycans fine-tune mammalian physiology.** *Nature*, **446**(7139):1030–1037, 2007. 20

[125] Jack Van Horssen, Pieter Wesseling, Lambert P W J Van Den Heuvel, Robert M W De Waal, and Marcel M Verbeek. **Heparan sulphate proteoglycans in Alzheimer's disease and amyloid-related disorders.** *The Lancet*, **2**(8):482–492, 2003. 21

REFERENCES

[126] A D Snow and T N Wight. **Proteoglycans in the pathogenesis of Alzheimer's disease and other amyloidoses.** *Neurobiology of Aging*, **10**(5):481–497, 1989. 21

[127] J McLaurin, T Franklin, X Zhang, J Deng, and P E Fraser. **Interactions of Alzheimer amyloid-beta peptides with glycosaminoglycans effects on fibril nucleation and growth.** *The Federation of European Biochemical Societies Journal*, **266**(3):1101–1110, 1999. 21

[128] G M Castillo, W Lukito, T N Wight, and A D Snow. **The sulfate moieties of glycosaminoglycans are critical for the enhancement of beta-amyloid protein fibril formation.** *Journal of Neurochemistry*, **72**(4):1681–1687, 1999. 21

[129] M Goedert, R Jakes, M G Spillantini, M Hasegawa, M J Smith, and R A Crowther. **Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans.**, 1996. 21

[130] H K Paudel and W Li. **Heparin-induced conformational change in microtubule-associated protein Tau as detected by chemical cross-linking and phosphopeptide mapping.** *The Journal of Biological Chemistry*, **274**(12):8029–8038, 1999. 21

[131] Shuiliang Yu, Shaoman Yin, Nancy Pham, Poki Wong, Shin-Chung Kang, Robert B Petersen, Chaoyang Li, and Man-Sun Sy. **Ligand binding promotes prion protein aggregation–role of the octapeptide repeats.** *The FEBS journal*, **275**(22):5564–5575, 2008. 21

[132] Elina Sandwall, Paul O'Callaghan, Xiao Zhang, Ulf Lindahl, Lars Lannfelt, and Jin-Ping Li. **Heparan sulfate mediates amyloid-beta internalization and cytotoxicity.** *Glycobiology*, **20**(5):533–541, 2010. 21

[133] Jin-Ping Li, Martha L Escobar Galvis, Feng Gong, Xiao Zhang, Eyal Zcharia, Shula Metzger, Israel Vlodavsky, Robert Kisilevsky, and Ulf Lindahl. **In vivo fragmentation of heparan sulfate by heparanase overexpression renders mice resistant to amyloid protein A amyloidosis.** *Proceedings of the National Academy of Sciences of the United States of America*, **102**(18):6473–6477, 2005. 21

[134] Xiao Zhang and Jin-Ping Li. *Heparan sulfate proteoglycans in amyloidosis.*, **93**. Elsevier Inc., 1 edition, January 2010. 21

[135] S B Shuker, P J Hajduk, R P Meadows, and S W Fesik. **Discovering high-affinity ligands for proteins: SAR by NMR**. *Science*, **274**(5292):1531–4, Nov 1996. 23

[136] Daniel A Erlanson. **Fragment-based lead discovery: a chemical update**. *Curr Opin Biotechnol*, **17**(6):643–52, Dec 2006. 23

[137] Bernd Meyer and Thomas Peters. **NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors**. *Angew Chem Int Ed Engl*, **42**(8):864–90, Feb 2003. 23

[138] Murray Coles, Markus Heller, and Horst Kessler. **NMR-based screening technologies**. *Drug Discov Today*, **8**(17):803–10, Sep 2003. 23

[139] Jeffrey W. Peng, Jonathan Moore, and Norzehan Abdul-Manan. **NMR experiments for lead generation in drug discovery**. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **44**(3-4):225–256, July 2004. 23

[140] C Dalvit, P Pevarello, M Tatò, M Veronesi, A Vulpetti, and M Sundström. **Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water**. *J Biomol NMR*, **18**(1):65–8, Sep 2000. 23

[141] J M Moore. **NMR techniques for characterization of ligand binding: utility for lead generation and optimization in drug discovery.** *Biopolymers*, **51**(3):221–43, January 1999. 23

[142] B Meyer, T Weimar, and T Peters. **Screening mixtures for biological activity by NMR**. *Eur J Biochem*, **246**(3):705–9, Jun 1997. 23

[143] Gonnella, Lin, Shapiro, Wareing, and Zhang. **Isotope-Filtered Affinity NMR**. *J Magn Reson*, **131**(2):336–8, Apr 1998. 23

[144] P J Hajduk, J Dinges, G F Miknis, M Merlock, T Middleton, D J Kempf, D A Egan, K A Walter, T S Robins, S B Shuker, T F Holzman, and S W Fesik. **NMR-based discovery of lead inhibitors that block DNA binding of the human papillomavirus E2 protein**. *J Med Chem*, **40**(20):3144–50, Sep 1997. 23

[145] Aldino Viegas, Anjos L Macedo, and Eurico J Cabrita. **Ligand-based nuclear magnetic resonance screening techniques**. *Methods Mol Biol*, **572**:81–100, 2009. 24

[146] Christian Ludwig and Ulrich L Guenther. **Ligand based NMR methods for drug discovery**. *Front Biosci*, **14**:4565–74, 2009. 24

[147] Till Maurer. **NMR studies of protein-ligand interactions.** *Methods in molecular biology (Clifton, N.J.)*, **305**:197–214, January 2005. 24

[148] Moriz Mayer and Bernd Meyer. **Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy**. *Angewandte Chemie International Edition*, **38**(12):1784–1788, June 1999. 25, 35

[149] M Mayer and B Meyer. **Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor.** *Journal of the American Chemical Society*, **123**(25):6108–17, June 2001. 25, 52

[150] Katalin E Kövér, Patrick Groves, Jesús Jiménez-Barbero, and Gyula Batta. **Molecular recognition and screening using a 15N group selective STD NMR method**. *J Am Chem Soc*, **129**(37):11579–82, Sep 2007. 25

[151] Thomas Haselhorst, Anne-Christin Lamerz, and Mark von Itzstein. **Saturation transfer difference NMR spectroscopy as a technique to investigate protein-carbohydrate interactions in solution**. *Methods Mol Biol*, **534**:375–86, 2009. 25, 26, 35

[152] Astrid Blume, Michael Fitzen, Andrew J Benie, and Thomas Peters. **Specificity of ligand binding to yeast hexokinase PII studied by STD-NMR.** *Carbohydrate research*, **344**(12):1567–74, August 2009. 25, 35

[153] Christopher a Lepre, Jonathan M Moore, and Jeffrey W Peng. **Theory and applications of NMR-based screening in pharmaceutical research.** *Chemical reviews*, **104**(8):3641–76, August 2004. 27

[154] R K Dewan and V a Bloomfield. **Molecular weight of bovine milk casein micelles from diffusion and viscosity measurements.** *Journal of dairy science*, **56**(1):66–8, January 1973. 27

[155] Kevin F. Morris and Charles S. Johnson. **Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy**. *Journal of the American Chemical Society*, **114**(8):3139–3141, 1992. 28, 35

[156] C.S. Johnson. **Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications**. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **34**(3-4):203–256, May 1999. 28

[157] Victor A Bloomfield. **Survey of Biomolecular Hydrodynamics**. In *Biophysics*, pages 1–16. 2000. 28

[158] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst. **Investigation of exchange processes by two-dimensional NMR spectroscopy**. *The Journal of Chemical Physics*, **71**(11):4546–4553, 1979. 30, 36

[159] L Braunschweiler and R.R Ernst. **Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy**. *Journal of Magnetic Resonance (1969)*, **53**(3):521 – 528, 1983. 32, 36

[160] Ad Bax and Donald G Davis. **MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy**. *Journal of Magnetic Resonance (1969)*, **65**(2):355 – 360, 1985. 32, 36

[161] Anat Frydman-Marom, Meirav Rechter, Irit Shefler, Yaron Bram, Deborah E Shalev, and Ehud Gazit. **Cognitive-performance recovery of Alzheimer's disease model mice by modulation of early soluble amyloidal assemblies**. *Angew Chem Int Ed Engl*, **48**(11):1981–6, 2009. 34, 63

[162] Brian Cutting, Sachin V Shelke, Zorica Dragic, Beatrice Wagner, Heiko Gathje, Soerge Kelm, and Beat Ernst. **Sensitivity enhancement in saturation transfer difference (STD) experiments through optimized excitation schemes**. *Magn Reson Chem*, **45**(9):720–4, Sep 2007. 35

[163] Alexej Jerschow and Norbert Müller. **Suppression of Convection Artifacts in Stimulated-Echo Diffusion Experiments. Double-Stimulated-Echo Experiments**. *Journal of Magnetic Resonance*, **125**(2):372 – 375, 1997. 36

[164] M Piotto, V Saudek, and V Sklenár. **Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions**. *J Biomol NMR*, **2**(6):661–5, Nov 1992. 36

[165] T.L. Hwang and A.J. Shaka. **Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients**. *Journal of Magnetic Resonance, Series A*, **112**(2):275 – 279, 1995. 36

[166] D Marion and K Wüthrich. **Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins**. *Biochem Biophys Res Commun*, **113**(3):967–74, Jun 1983. 36

[167] D S Wishart, B D Sykes, and F M Richards. **Relationship between nuclear magnetic resonance chemical shift and protein secondary structure**. *J Mol Biol*, **222**(2):311–33, Nov 1991. 36

[168] Leo C James and Dan S Tawfik. **Conformational diversity and protein evolution–a 60-year-old hypothesis revisited**. *Trends Biochem Sci*, **28**(7):361–8, Jul 2003. 37

[169] Gregory M Lee and Charles S Craik. **SUP_Trapping moving targets with small molecules.** *Science (New York, N.Y.)*, **324**(5924):213–5, April 2009. 38

[170] David D Boehr, H Jane Dyson, and Peter E Wright. **An NMR Perspective on Enzyme Dynamics**. *Engineering*, 2006. 38, 39

[171] J M Yon, D Perahia, and C Ghélis. **Conformational dynamics and enzyme activity**. *Biochimie*, **80**(1):33–42, Jan 1998. 38

[172] J. A. Mccammon and S. C. Harvey. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, April 1988. 38

[173] L E Kay, D A Torchia, and A Bax. **Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease**. *Biochemistry*, **28**(23):8972–9, Nov 1989. 38

[174] A G Palmer, 3rd, C D Kroenke, and J P Loria. **Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules**. *Methods Enzymol*, **339**:204–38, 2001. 38

[175] Y Bai, J J Englander, L Mayne, J S Milne, and S W Englander. **Thermodynamic parameters from hydrogen exchange measurements**. *Methods Enzymol*, **259**:344–56, 1995. 40, 43

[176] S W Englander and N R Kallenbach. **Hydrogen exchange and structural dynamics of proteins and nucleic acids**. *Q Rev Biophys*, **16**(4):521–655, Nov 1983. 40, 41

[177] V J Hilser and E Freire. **Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors**. *J Mol Biol*, **262**(5):756–72, Oct 1996. 40

[178] T Sivaraman and a D Robertson. **Kinetics of conformational fluctuations by EX1 hydrogen exchange in native proteins.** *Methods in molecular biology (Clifton, N.J.)*, **168**:193–214, January 2001. 40

REFERENCES

[179] A HVIDT. **A DISCUSSION OF THE PH DEPENDENCE OF THE HYDROGEN-DEUTERIUM EXCHANGE OF PROTEINS**. *C R Trav Lab Carlsberg*, **34**:299–317, 1964. 40, 41

[180] Y Bai, J S Milne, L Mayne, and S W Englander. **Primary structure effects on peptide group hydrogen exchange.** *Proteins*, **17**(1):75–86, September 1993. 40, 41, 49

[181] A Hvidt and S O Nielsen. **Hydrogen exchange in proteins**. *Adv Protein Chem*, **21**:287–386, 1966. 41

[182] J L Marmorino, D S Auld, S F Betz, D F Doyle, G B Young, and G J Pielak. **Amide proton exchange rates of oxidized and reduced Saccharomyces cerevisiae iso-1-cytochrome c**. *Protein Sci*, **2**(11):1966–74, Nov 1993. 43, 44

[183] E Kupce and R Freeman. **Wide-band excitation with polychromatic pulses**. *J.Magn.Reson.Ser.A*, **108**:268–273, 1994. 47

[184] H Geen and R Freeman. **Band-selective radiofrequency pulses**. *J.Magn.Reson*, **93**:93–141, 1991. 47

[185] A J Shaka, P B Barker, and R Freeman. **Computer-optimized decoupling scheme for wideband applications Computer-optimized decoupling scheme for wideband applications and low-level operation**. *J. Magn. Reson.*, **64**:547–552, 1985. 47

[186] F Delaglio, S Grzesiek, G W Vuister, G Zhu, J Pfeifer, and A Bax. **NMRPipe: a multidimensional spectral processing system based on UNIX pipes**. *J Biomol NMR*, **6**(3):277–93, Nov 1995. 48

[187] M Okon, P Bray, and D Vucelić. **1H NMR assignments and secondary structure of human beta 2-microglobulin in solution**. *Biochemistry*, **31**(37):8906–15, Sep 1992. 48

[188] Gideon Schwarz. **Estimating the Dimension of a Model**. *Annals of Statistics*, **6**(2):461–464, 1978. 49

[189] O D Monera, C M Kay, and R S Hodges. **Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions**. *Protein Sci*, **3**(11):1984–91, Nov 1994. 53

[190] S L Mansfield, D a Jayawickrama, J S Timmons, and C K Larive. **Measurement of peptide aggregation with pulsed-field gradient nuclear magnetic resonance spectroscopy.** *Biochimica et biophysica acta*, **1382**(2):257–65, February 1998. 59, 65

[191] U Soomets, R Mahlapuu, R Tehranian, J Jarvet, E Karelson, M Zilmer, K Iverfeldt, M Zorko, a Gräslund, and U Langel. **Regulation of GTPase and adenylate cyclase activity by amyloid beta-peptide and its fragments in rat brain tissue.** *Brain research*, **850**(1-2):179–88, December 1999. 59, 65

[192] Y Fezoui, D M Hartley, J D Harper, R Khurana, D M Walsh, M M Condron, D J Selkoe, P T Lansbury, Jr, A L Fink, and D B Teplow. **An improved method of preparing the amyloid beta-protein for fibrillogenesis and neurotoxicity experiments**. *Amyloid*, **7**(3):166–78, Sep 2000. 61

[193] D S Wishart, C G Bigam, A Holm, R S Hodges, and B D Sykes. **1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects**. *J Biomol NMR*, **5**(1):67–81, Jan 1995. 63

[194] K R Harris and L A Woolf. **Temperature and Volume Dependence of the Viscosity of Water and Heavy Water at Low Temperatures**. *J. Chem. Eng. Data*, **49**:1064–1069, 2004. 66

[195] Giuliana Verdone, Alessandra Corazza, Paolo Viglino, Fabio Pettirossi, Sofia Giorgetti, Palma Mangione, Alessia Andreola, Monica Stoppini, Vittorio Bellotti, and Gennaro Esposito. **The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition**. *Protein Sci*, **11**(3):487–99, Mar 2002. 79, 82, 95

[196] Hua Fung and Bruce Demple. **A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells**. *Mol Cell*, **17**(3):463–70, Feb 2005. 121

[197] Carlo Vascotto, Laura Cesaratto, Leo A H Zeef, Marta Deganuto, Chiara D'Ambrosio, Andrea Scaloni, Milena Romanello, Giuseppe Damante, Giulio Taglialatela, Daniela Delneri, Mark R Kelley, Sankar Mitra, Franco Quadrifoglio, and Gianluca Tell. **Genome-wide analysis and proteomic studies reveal APE1/Ref-1 multifunctional role in mammalian cells**. *Proteomics*, **9**(4):1058–74, Feb 2009. 121

[198] Gianluca Tell, Giuseppe Damante, David Caldwell, and Mark R Kelley. **The intracellular localization of APE1/Ref-1: more than a passive phenomenon?** *Antioxid Redox Signal*, **7**(3-4):367–84, 2005. 121

[199] S Xanthoudakis, G G Miao, and T Curran. **The redox and DNA-repair activities of Ref-1 are encoded by nonoverlapping domains**. *Proc Natl Acad Sci U S A*, **91**(1):23–7, Jan 1994. 121

[200] M A Gorman, S Morera, D G Rothwell, E de La Fortelle, C D Mol, J A Tainer, I D Hickson, and P S Freemont. **The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites**. *EMBO J*, **16**(21):6548–58, Nov 1997. 121

[201] Brittney A Manvilla, Kristen M Varney, and Alexander C Drohat. **Chemical shift assignments for human apurinic/apyrimidinic endonuclease 1**. *Biomol NMR Assign*, **4**(1):5–8, Apr 2010. 121