Introduction
**Introduction**

Proteins are major biomacromolecules necessary for carrying out various physiological functions. They are the polypeptides with characteristic sequence of amino acid residues which fold to acquire a unique three dimensional conformation to perform a given function. For proteins to be physiologically functional they must be properly folded to attain this unique conformation called native conformation which must be stable enough in the given condition to discriminate it thermodynamically from other possible conformations. The kinetically controlled former process is called protein folding and the latter is referred as protein stability. Both protein folding and stability is due to an interplay of weak non-covalent interactions to first bring the unfolded polypeptide chain to folded native conformation and then to maintain it's nativity in variable conditions in order for the protein to remain functional. A slight alteration in these weak interactions may lead to the unfolded polypeptide getting misfolded to a non native conformation. On the other hand, altered interactions may not interfere with critical interactions for protein folding and instead might only interfere with interactions necessary to keep protein kinetically stable. Thus the folded protein though is able to attain its almost native structure but is not stable enough and is prone to undesirable changes like truncation or aggregation.

**Forces determining protein stability**

Free energy of stabilization of globular proteins in solution is only 5 to 10 kcal/mol which is equivalent to only a few weak intermolecular interactions (Dill, 1990). Thus, the folded conformations are only marginally stable than unfolded, biologically inactive conformations under physiological conditions. This small net conformational stability is the result of much larger contributions from several important forces. The major destabilizing force is the conformational entropy (Pace, 1990). The entropy of an unfolded protein is large because rotation around the bonds in the polypeptide backbone and the side chains is less restricted than in a folded protein. Other major contributing forces are hydrogen bonding and the hydrophobic effect.
a) Hydrogen bonds

Hydrogen bonds form between an electronegative atom with a covalently bound hydrogen, DH (the donor), and another electronegative atom, A (the acceptor). When DH and A form a hydrogen bond, their electronic structure is perturbed very little compared to the changes that accompany covalent bond formation. In proteins, the electronegative atoms of most importance are O and N; the hydrogen bond of most importance is that between the amide hydrogen and the carbonyl oxygen in a peptide group. These hydrogen bonds make up 68% of the hydrogen bonds that occur in globular proteins (Stickle et al., 1992) and from it there also exist backbone-side-chain, side-chain-backbone, and side-chain-side-chain donor-acceptor contributions. Out of this side-chain-side-chain hydrogen bonding is believed to play a major role in stabilizing proteins (Ragone, 2001). The driving force in folding was initially thought to be this intramolecular hydrogen bonding (Pauling et al., 1951), then the hydrophobic effect (Kauzmann, 1959). In recent times, it has been argued that intramolecular hydrogen bonding is destabilizing (Dill, 1990; Honig & Yang, 1995), partially stabilizing and destabilizing (Honig, 1999) and once again, an important driving force (Takano et al., 2003; Pace et al., 2004). During the past two decades, the advent of protein engineering brought a hope that the newfound ability to introduce site directed mutants at will would provide ready answers to such unresolved questions (Oxender and Fox, 1987). The goal of studying hydrogen bonding mutants is to learn how much stability is gained when a polar group that is hydrogen-bonded to water in the unfolded state is dehydrated to form a specific intramolecular hydrogen bond in the folded state. The most common approach has been to replace a side chain involved in a hydrogen bond, such as Asn, with a side chain that cannot form a hydrogen bond, such as Ala (Serrano et al., 1992; Shirley et al., 1992; Green et al., 1992). These studies point out that a folded protein loses stability by burying polar groups but gains it back by forming hydrogen bonds. However, despite thousands of mutational experiments, disagreement about the energetic role of hydrogen bonding still remains.

Studies investigating influence of hydrogen bonds on packing density suggested that hydrogen bonds may contribute to protein stability, in part, by increasing packing density in the protein interior, and thereby increasing the contribution of van der Waals
interactions to protein stability (Schell et al., 2006). Molecular dynamics simulation studies dealing with thermodynamics of hydrogen bond breaking and formation in solutions of alcohol show that free energy of hydrogen bond formation is essentially independent of the environment (around 5 kJ/mol), suggesting that buried hydrogen bonds (e.g., in proteins) do not contribute significantly to protein stability (Spoel et al., 2006). On the other hand studies exploring the contribution of hydrogen bond reveal that H-bond strength varies from one protein to another stabilizing in one and destabilising in the other and presumably at different sites within the same protein (Shi et al., 2002). These studies suggest that contribution that polar groups make to protein stability depends strongly on their environment (Takano et al., 2003).

b) Electrostatic interactions

Specific interactions in proteins are largely electrostatic and are important in protein folding, stability, flexibility, and function. Different conformations adopted by protein in varying conditions can have electrostatic interactions contributing favorably or nonfavorably (Kumar & Nussinov, 2002; Matousek et al., 2007). Electrostatic interactions in the Denatured states can include specific non-native interactions that can even persist in the transition state for protein folding. These electrostatic interactions can be energetically significant and their modulation either by mutation or by varying solution conditions can have a major impact upon protein stability (Cho et al., 2008). There exist both close-range electrostatic interactions (salt bridges) and the long range electrostatic interactions in proteins. Salt bridges are formed by spatially proximal pairs of oppositely charged residues in native protein structures. A single salt bridge can contribute up to 3–5 kcal/mol to the free energy of protein folding (Anderson et al., 1990). Often salt-bridging residues are also close in the protein sequence and fall in the same secondary structural element. Salt bridges are rarely found across protein parts which are joined by flexible hinges, a fact suggesting that salt bridges constrain flexibility and motion. Recent computational and experimental evidence shows that salt bridges can be stabilizing (Permyakov et al., 2005; Ibarra-Molero et al., 2004) or destabilizing (Hendsch and Tidor 1994). Salt bridges are also reported to effect folding kinetics in a context dependent way (Ibarra-Molero et al., 2004). Structural and thermodynamic
comparisons of thermophilic and mesophilic proteins indicate that salt bridges contribute to reduced heat capacity change of unfolding (Lee et al., 2005). They can thus contribute significantly towards the thermophilic-mesophilic protein stability differential (Kumar & Nussinov, 2002; Nakamura, 1996) as salt bridge networks could accommodate stochastically the disorder of increased thermal motion to produce thermal stability (Missimer et al., 2007).

Long range electrostatic interactions can be both repulsive and attractive in nature also play important role in determining protein stability (Pace et al., 2000; Grimsley et al., 1999). Favorable long-range electrostatic interactions when present in denatured state lead to compaction of unfolded polypeptide chain thus reducing the net contribution of electrostatic interactions to protein stability (Funahashi et al., 2003) and may bring down the overall stability of the protein owing to the changes in the long-range electrostatic interactions. Long range interactions may also be sometimes insignificant in determining protein stability (Sun et al., 1991) while can be very critical in determining the protein dynamics (Fadrna et al., 2005) and meanwhile can also influence protein-protein interactions (Ramirez-Carrozzi & Kerppola, 2001).

c) Hydrophobic effect

The burial of hydrophobic residues is considered to be the major driving force for protein folding and stability (Dill, 1990; Rose et al., 1985; Nakai et al., 1988; Shortle & Meeker, 1986; Kellis et al., 1988, 1989; Alber and Matthews, 1987; Yutani et al., 1987; Matsumura et al., 1988; Shortle et al., 1990; Sandberg & Terwilliger, 1991). First one is to calculate the hydrophobicity scale of individual residues based on their transfer free energy from water to organic solvents like n-octanol (Herrmann et al., 1995), N-methylacetamide (Roseman, 1988) and cyclohexane (Lomize et al., 2002) assuming the solvent to mimic the folded protein’s interior. Studies with octanol provide reasonable transfer free energy $\Delta G_{tr}$ values for the peptide groups and amino acid side chains to estimate the contribution of the groups buried in folding to protein stability. The other method uses site-directed mutagenesis as a tool to investigate the forces that stabilize proteins focused on the hydrophobic effect. The approach has been to replace one hydrophobic side chain with another and measure $\Delta(\Delta G)$. To minimize the contribution
of steric strain, mutants are created to have a smaller side chain replacing the larger one. Many such studies on proteins like tryptophan synthase (Yutani et al., 1987), T4 lysozyme, barnase (Golovonov et al., 2000), staphylococcal nuclease (Chen et al., 2004), gene 5 protein from bacteriophage Φ (Zhang et al., 1996) and chymotrypsin inhibitor 2 (Ahmed et al., 2008) have been reported. Other studies deal with core packing and strain to assess the effect of adding larger hydrophobic residues to the interior of globular proteins (Lim et al., 1994). \( \Delta(\Delta G) \) values for hydrophobic mutants increase with the amount of nonpolar surface area buried, as expected. In order to compare the hydrophobic substitutions at the same accessibility the measured \( \Delta(\Delta G) \) values have been divided by the fraction buried for the side chain in the wild-type protein so that the \( \Delta(\Delta G) \) values are compared at the same accessibility, namely, 100% buried. In reality burial doesn’t happen to be 100% and cavities are formed due to non sufficient van der Walls interactions to fill the cavity formed completely which leads to an overestimated \( \Delta(\Delta G) \) values (Yamada et al., 1994) by \(-0.11 \text{ kcal/mol}\) (Ratnaparkhi & Varadarajan, 2000).

Moreover recent studies on protein stability in H2O and D2O (Efimova et al., 2007) establishes the importance of hydrophobic effect as D2O is a poorer solvent for non polar amino acids than H2O implying that larger and efficient packing of the hydrophobic core in D2O leads to enhanced stability.

**Thermodynamics of protein stability**

The conformational stability of a protein is defined as the free energy change, \( \Delta G \), for the reaction folded to unfolded under physiological conditions. The equilibrium between the two states can be represented as

\[
N \leftrightarrow D
\]

and conformational stability of a protein can be described as

\[
\Delta G = \Delta G_D - \Delta G_N = -RT \ln K = -RT \ln [D]/[N]
\]
where \([D]\) and \([N]\) represent the concentrations of denatured (D) and native (N) states, \(\Delta G_D\) and \(\Delta G_N\) represent the free energies of D and N, and K and \(\Delta G\) are the equilibrium constant and standard free energy change, respectively. The folding–denaturing transition in proteins is a highly cooperative process. In certain cases, as a rule for smaller proteins, it suffices to describe this transition within a 2-state approach involving the native state N and the denatured state D, only.

Since there are multiple perturbation modes to induce the transition, \(\Delta G\) is a multidimensional function of all the parameters which can independently be varied in an experiment, e.g. temperature \(T\), pressure \(P\), cosolvent \(x\), concentration of cosolvent \(C_x\), pH, etc.:

\[
\Delta G = \Delta G^0 + f(T, P, C_x, \text{pH}, \ldots)
\]

\(\Delta G^0\) depends on the reference state only, hence, is, for a transformation under fixed conditions, a constant. The function \(f\) contains all the system parameters which characterize the denaturing transition of a protein in a given solvent, e.g. the changes in entropy, volume compressibility, specific heat, thermal expansion, etc. Keeping all solvent associated parameters constant and considering variation in \(P\) and \(T\) only, since the transition is characterized by a latent heat, it is of 1st order, and, consequently, is governed by the Clausius–Clapeyron equation

\[
\frac{dP}{dT} = \frac{\Delta S}{\Delta V}
\]

\(\Delta S\) and \(\Delta V\) are the entropy and volume changes associated with the transition. Both quantities depend on the actual pressure \(P\) and temperature \(T\) where the transition takes place. The boundaries of the stability phase diagram, i.e. the area in a pressure–temperature plane where the protein is stable in its native state, can then be determined from a solution of this equation. The first derivatives of the volume with respect to pressure and temperature, namely

\[
\beta_T = (\delta V / \delta P)_T \quad \text{and} \quad \alpha^* = (\delta V / \delta T)_P,
\]
the absolute changes of the volume with pressure and temperature, are closely related to
the compressibility and the thermal expansion of the protein, respectively. Likewise, the
derivatives of the entropy with respect to pressure and temperature are associated with
the thermal expansion and with the specific heat capacity $C_p = T(\delta S / \delta T)_p$, respectively.
These are system parameters which are assumed to be roughly independent on pressure
and temperature and therefore $\delta S$ and $\delta V$ depend only linearly on $T$ and $P$, and, hence,
the equation can easily be integrated. The result is a general 2nd order curve in $P$ and $T$
whose shape may be elliptic, parabolic or hyperbolic:

$$aP^2 + bT^2 + 2cPT + 2fP + 2gT + \text{const} = 0$$

thus derived coefficients $a$, $b$, $c$, etc. are related to the changes of the system parameters
along the transition (Scharnagl et al., 2005), namely to $\Delta \beta_T$, $\Delta \alpha$, $\Delta C_p$, and to $\Delta S_0$ and
$\Delta V_0$, the entropy and volume change at the reference pressure and temperature $(P_0, T_0)$.

**Factors affecting protein stability**

**a) Effect of temperature**

High temperature induces denaturing transition accompanying the conversion of
structured protein to random coil form with solvation of hydrophobic amino acids. As a
consequence, water solvating hydrophobic molecules forms locally ordered structures
which have low entropy and low enthalpy due to well aligned hydrogen bonds in them
(Kim et al., 2005;). As melting of these structures requires energy, temperature induced
protein denaturation is accompanied by an increase in specific heat. Accordingly, the
change of the specific heat, $\Delta C_p$ associated with the temperature induced transition

$$\Delta C_p = C_{PD} - C_{PN},$$

is generally positive (Griko and Privalov, 1992). As long as $\Delta C_p$ remains positive the
complete solvation of hydrophobic core of the protein interior does not happen due to still
remaining structure and therefore the difference in enthalpy, $\Delta H$, between the native and the denatured state keep on increasing as the temperature is raised, according to

$$\Delta H(T) = H(T_1) + \Delta C_P [T-T_1]$$

Meanwhile the respective difference in entropy, $\Delta S$, also increases, since the conformational ordering melts away with the increasing temperature. At some critical temperature $T = T_m$, the enthalpic term, $\Delta H$, and the entropic term, $-T\Delta S$, cancel, rendering a free energy change $\Delta G$ of zero. At this temperature $T_m$, energetically more favorable denatured state the transition to the denatured state takes place because it is energetically more favorable (Trefethen et al., 2005; Schoeffler, 2004)

Protein stability is optimum at a certain temperature and increasing or decreasing from this temperature leads to reduced stability and thus like high temperature induced denaturation proteins are also subjected towards cold denaturation i.e. at lower temperatures. Lowering the temperature decreases the enthalpy term so that it eventually becomes negative and may compensate the entropy term, $T\Delta S$, which is positive due to decreasing entropy. The actual transition temperatures into the denatured state depend of course on pressure: High pressure at low temperature may destabilize the locally ordered structures (Cai et al., 2005) because it counteracts an optimum alignment of the hydrogen bonds. However, in the low pressure, high temperature regime, pressure may stabilize the respective structures to some extent (Marques et al., 2003)

### b) Effect of pressure

It has been shown in numerous studies that hydrostatic pressure may lead to disruption of the intermolecular forces maintaining native protein structure, which is accompanied by a decrease in the volume of the protein–water system, and simultaneous unfolding. In part, this is explained by the fact that more water molecules can fill in the protein’s void volumes when they become accessible to solvent upon dissociation. This results in a total volume contraction and is favoured under high pressure (Gross & Jaenicke 1994; Mozhaev et al., 1996; Silva et al., 1992; Royer 2002). Moreover, solvent-exposure of charged groups that have been involved in stabilization of protein assemblies
(as salt bridges) also contributes to the overall volume reduction through the so-called electrostriction effect, which consists in a tightly ordered arrangement of the solvent dipolar molecules around charged solutes. Another important, yet still debated (Boonyaratanakornkit et al., 2002), factor contributing to the pressure effect on polypeptide assemblies stems from the hydration of hydrophobic residues under high pressure (e.g. hydration volumes of model hydrocarbons such as benzene or methane are negative (Gross & Jaenicke 1994). Pressure denaturation studies provide a fundamental thermodynamic parameter for protein unfolding, the $\Delta V^0$, in addition to being an alternative method for perturbing the folded state, and thus elucidating its stability. Denaturation of proteins is usually studied at atmospheric pressure using high temperature, guanidinium hydrochloride or urea as denaturants. Interpretation of the results obtained using such methods may be complicated by the following: (i) varying the temperature changes both the volume and the thermal energy of the system at the same time and (ii) the thermodynamic parameters of denaturation by guanidinium chloride or urea are influenced by the binding of these molecules to proteins. The use of pressure is also advantageous from the fact that it is a rather mild denaturing agent and from methodological points of view: the transition to native conditions (renaturation) is achieved simply by releasing the pressure. Furthermore, the effects of pressure on proteins are generally found to be reversible, and seldom are they accompanied by aggregation or changes in covalent structure.

Also with respect to the kinetics of the folding reaction, pressure studies are of particular use, as they allow to evaluate the volume profile during the folding process and to characterize the nature of the barrier to folding or unfolding and the corresponding transition state. Moreover, pressure studies present an important advantage due to the positive activation volume for folding, the result of which is to slow-down folding substantially, in turn allowing for relatively straightforward measurements of structural order parameters characteristic for folding intermediate states, which are difficult or even impossible to quantify on much faster time-scales corresponding to ambient pressure conditions.

$\Delta G = \Delta G_D - \Delta G_N$, the free energy change associated with protein denaturation, becomes lower as pressure is increased, at least above some threshold pressure. $\Delta G (T) =$
\( \Delta G (T1) + \Delta V \) [P-P1]. \( \Delta V = V_D - V_N \) is the volume change in going from the native to the denatured state. As a rule, \( \Delta V \) is negative because the structure of the native state has voids, for instance in the protein pockets, which are squeezed away in the denatured state so that its volume is smaller and, hence, the transition into the denatured state becomes favored under high pressure. In the low pressure, high temperature regime \( \Delta V \) can also be positive. Increasing the pressure up to some critical level \( P = P_0 \), the protein may eventually cross the boundary \( \Delta G = 0 \), and the transition to the denatured state takes place. The respective transition at low pressure is less straightforward to understand. First of all, we note that, in a large temperature range, the low pressure denaturation regime would require negative pressure, a condition which has, so far, not been realized experimentally. Accordingly, low pressure denaturation can experimentally be investigated in a rather limited temperature range only, for instance at high temperature. There is indeed a temperature range in which high pressure leads to a stabilization of the native state, and, consequently, low pressure to a destabilization associated with denaturation. High pressure protein stabilization takes place in a range where the change \( \Delta \beta_T = -(\delta \Delta V / \delta P)T \) is negative meaning that the denatured state is less compressed than the native state. If so, \( \Delta V = V_D - V_N \) will be positive and, since higher pressure stabilizes the smaller volume, the native state is favored. The question, however, what determines the temperature-pressure range with a negative \( \Delta \beta_T \), remains to be answered. At rather high pressure (i.e. outside this range) the denatured state is far from being a random coil state. It is plausible that unfolding to a random coil against high pressure is severely hindered. Instead, the high pressure denatured state is still kind of a globular state where the voids in the protein are squeezed to a high degree so that \( V_D < V_N \). On the other hand, in the lower pressure range and at sufficiently high temperature, unfolding to a random coil like state is still possible. Accordingly, the protein acquires a larger surface and, concomitantly, a larger volume. In addition, compression is much harder than in the native state because the compressible voids have vanished and the hydration shell is harder to compress than bulk water due to the ordered structures induced by the hydrophobic amino acids.
c) Effect of pH

Variation in pH causes changes in ionization status of titratable groups altering electrostatic interaction. The pH dependence of the thermodynamic stability ($\Delta G$ (F-U)) of a protein arises as a consequence of differential pK(a) values between folded (F) and unfolded states(U). A significant component of the free energy difference between native and denatured states is due to a small number of amino acids whose pKas are shifted anomalously in the native protein (Pace et al., 1990; Yang and Honig, 1993; Yang et al., 1992, Hu et al., 1992). General shifts of side-chain pK\textsubscript{a} values in unfolded states to lower than standard pK\textsubscript{a} values (by 0.3–0.4 pH units on average) have also been predicted, pointing toward the presence of stabilizing electrostatic interactions within unfolded states of proteins under non-denaturing conditions (Elcock, 1999; Tan et al., 1995) Other than this changes in pH also changes the overall net charge on the protein. When proteins possess both positively and negatively charged groups (e.g., at pH values close to the pi), anisotropic charge distribution on the protein surface could give rise to dipoles. In such cases, protein–protein interactions could be highly attractive, making assembly processes such as aggregation energetically favorable. Thus pH plays an essential role in determining protein stability.

The unfolding free energy, $\Delta G_{FU}$, can be calculated at any pH (Tollinger et al., 2003) as

$$\Delta G_{FU} = -RT \ln \left( \frac{\sum_{m=0}^{i} [UH_m]}{\sum_{m=0}^{i} [FH_m]} \right)$$

where $[UH_m]$ and $[FH_m]$ represent sums of concentrations corresponding to the binding of $m$ protons to $i$ binding sites for protons (ionizable groups) in the unfolded and folded states, respectively, $R$ is the universal gas constant, and $T$ is the temperature. $\Delta G_{FU}$ can be separated into a pH-independent term, $\Delta G_{FU}^H$ (representing non-electrostatic contributions to $\Delta G_{FU}$, as well as electrostatic contributions at a pH where all ionizable sites $i$ are protonated in both states), and terms related to protonation/deprotonation equilibria involving individual ionizable groups, $\Delta G_{FU}^{pH}(i)$ (representing the pH dependence of the
contribution of proton binding at site \( i \) to the overall \( \Delta G_{FU} \) as

\[
\Delta G_{FU} = \Delta G_{FU}^{(i)} + \sum_i \Delta G_{FU}^{pH}(i).
\]

Assuming an equilibrium between four species (folded and unfolded state, protonated, and deprotonated) at each site \( i \), individual, pH-dependent values of \( \Delta G_{FU}^{pH}(i) \) can be calculated for each group as

\[
\Delta G_{FU}^{pH}(i) = -RT \ln \left( \frac{[H^+] + K_a^{U}(i)}{[H^+] + K_a^{F}(i)} \right),
\]

where \( K_a^{U}(i) \) and \( K_a^{F}(i) \) are the ionization constants for group \( i \) in the unfolded and folded state, respectively. The maximum of \( \Delta G_{FU}^{pH}(i) \) between very high pH where both states are fully deprotonated and very low pH where both states are fully protonated is given by

\[
\Delta G_{FU,\text{max}}^{pH}(i) = 2.303RT \{ pK_a^{U}(i) - pK_a^{F}(i) \}.
\]

Relative values of the pH dependence of the thermodynamic stability for a system containing multiple \( i \) ionizable groups can be calculated by summation as \( \sum_i \Delta G_{FU}^{pH}(i) \). The calculations above provide a simple means to dissect the pH dependence of \( \Delta G_{FU} \) into additive contributions due to individual ionizable groups.

Above calculations do not include conformational adaptations in response to changes in the ionization state which tend to complicate the calculations of changes in stability as a function of pH and therefore this kind of changes must be taken into account before proceeding with the calculations.

d) Effect of ionic strength

Ionic strength of the solutions determines protein stability by multiple mechanisms e.g., the Hofmeister effect (Toth et al., 2008; Zhou, 2005; Cacace et al., 1997), preferential hydration (Hong et al., 2004; Wright et al., 2002, Arakawa and Timasheff, 1984), electrostatic effects and weak ion binding (Ramos and Baldwin, 2002; Maldonado et al., 2002; Nishimura et al., 2001). Combination of all these effects complicates the interpretation of salt effects. Ions in some cases may stabilize proteins by
Introduction

high affinity binding to specific sites. While the ligand induced stabilization is ion-specific and usually observed in the low salt concentration range, bulk ionic strength effects play a role in screening surface charge-charge interactions. Hofmeister effects, which occur at still higher salt concentrations, may strengthen the hydrophobic force by increasing the surface tension of the solvent, or stabilize peptide dipoles through specific ionic interactions. Recent results suggest that the efficiency of different salts to screen charge-charge interactions correlates with their denaturing strength and with the position of the constituent ions in the Hofmeister rankings (Perez-Jimenez et al., 2004). In the absence of site-specific ion binding, differential salt effects in reflect primarily differences in Coulombic interactions, e.g. between natural or designed variants of a protein. The Hofmeister effects that occur in parallel are usually insensitive to local changes in amino acid sequence. All these various mechanisms can affect the stability of both the native state and the unfolded state (Pradeep and Udgaonkar, 2004).

Charges can be shielded by counter ions which are present in solvent surrounding and thus ionic cosolvents can regulate the electrostatic interactions on the protein surface. But the effect by these ions are highly context dependent as surface charges which are involved in thermodynamically stabilizing the protein shielding of which can lead to destabilization. On the contrary shielding of surface charges which are repulsive to other surface charges in the closer vicinity can lead to further stabilization of the protein. Also at high temperatures dielectric constant of water decreases which gives higher impact to electrostatic interactions. Thus at higher temperatures, high solvent ionic strength acts against repulsive surface interactions helping in protein stability while further destabilize proteins if the surface charges are attractive and protein stabilizing like in salt bridges. Because pH determines the type, total, and distribution of charges in a protein, salt-binding effects may be strongly pH dependent. High ionic strength can also facilitate charge shielding of repulsive interactions at pH<<pI and at pH>>pI but would otherwise counter attractive salt bridging at pH=pI.

At low concentrations, the predominant effect of ions in solution results from charge shielding, which reduces electrostatic interactions. However, at high concentrations of certain salts, in addition to charge-shielding effects, preferential binding of ions to the protein surface can result in a decrease in thermodynamic stability of the
native conformation and an increase in equilibrium solubility (Mertz and Leikin, 2004; Arakawa and Timasheff, 1984). Other salts that are preferentially excluded from protein surface show stabilizing or salting-out effects (Zhou, 2005).

With the addition of salt, the electrostatic interactions will be shielded and the pKa values can show shift from their normal expected values (Lindman et al., 2007). These shifts together with other salt mediated effects can even account for concentration dependent differential and anomalous salt effects as the net effect on protein stability is determined by alteration in intensity of different protein-salt interactions (Spencer et al., 2005). Thus salts affect protein stability by modifying the ionic strength of the solution, which overall can be slightly stabilizing or destabilizing depending on the nature of the specific charge distribution within the protein.

e) Effect of solvent environment

Since the stability of proteins results from a large number of counteracting enthalpic and entropic contributions which favor the folded state as compared to the unfolded state only marginally, it is clear that changes in the interactions of the protein with the solvent, for instance by adding a cosolvent, may have a severe influence on protein stability. The nature and the magnitude of the individual contributions of the protein solvent interactions to the free energies \( G_D \) and \( G_N \) of the respective states of the protein (denatured, D or native, N) are highly dependent on the solvent environment. Since water is the environment in which proteins exist and operate, the structure and dynamics of the hydration water is directly linked to protein flexibility and stability (Cioni et al., 2005; Soares et al., 2003). Consequently, chemical denaturants and cosolvents which change the properties of the hydration water can readily alter the equilibrium between different conformational ensembles. Free energy of transfer of amino acids from vacuum to water varies from that from vacuum to a different solvent. Thus free energy of individual amino acids differs in varying solvent conditions and along changes the thermodynamic stability of proteins comprising these amino acid residues (Pace et al., 2004). Accordingly, changing the chemical potential of the solvent by changing the concentration of cosolvents and/or denaturants provides a valuable tool for probing protein stability and protein-water interaction. For example, the addition of
glycerol to a solution of a native state protein leads to changes in structure and dynamics as reflected, for instance, in changes of volumes and compressibilities (Almagor et al., 1998; Priev et al., 1996).

The mechanisms of the cosolvent-protein interaction include following possibilities: (1) direct contact interaction of cosolvent molecules with the protein; (2) indirect effects via the perturbation of the hydration layer; (3) combination of (1) and (2), a disruption of the water structure in the hydration shell, so that water molecules are released and enable a direct interaction of cosolvent molecules with protein groups. Thus compatible cosolvents are those which do not disturb protein functionality, while noncompatible cosolvents, e.g. denaturants like urea, guanidinium hydrochloride, alcohols, etc. tend to induce a disruption of protein structure. There can be another category of molecules which tend to compensate and counter action of non-compatible cosolutes and stabilize the folded form against denaturation under external stress (examples are sugars, polyols, monomeric amino acids and methylamines). These are the osmolytes which are able to stabilize cells in vivo against dehydrating stress (e.g. salinity) and/or volume changes by maintaining an osmotic equilibrium. Some of the cosolvents called chaotrops are also known for disordering the water structure while some others known as kosmoprotots are known to induce disorder in the water structure whereas the so-called kosmotrops induce order. Due to the close relation between the structure of water and protein stability the coordination of water molecules by cosolvent molecules does have a significant influence on protein stability.

Combined effect of all these factors of solvent interaction however is also dependent strongly on the surface topology of the protein and the structure of the hydration shell as well as on the concentration of the cosolvent.

**Protein folding**

The protein folding problem is widely recognized as a challenge to investigators engaged in biological research at molecular level for previous many decades (Basharov, 2003). It is formulated simply as either how does a polypeptide chain fold to the native protein, or how does the amino-acid sequence specify unique three dimensional structure of the protein. A number of phenomenological models is proposed for the mechanism of
protein folding, such as the framework (Lin and Chang, 2007; White et al., 2005, Santra et al., 2004), diffusion-collision (Zhang et al., 2005; Myers and Oas, 2001), nucleated collapse (Djikaev, 2007; Kuo et al., 2005; Guo and Thirumalai et al., 1997; Dagget et al., 1996; Moult and Unger, 1991), hierarchic (Compiani et al., 2004; Tiana et al., 2003; Chakraborty and Peng, 2000), and so on models (Zhou and Karplus, 1999;). Different theories and several coarse grained (Lu and Liu, 2008; Cecconi at al., 2008; Cho et al., 2008) and lattice models (Wang and Klimov, 2008; Cellmer et al., 2005, Skolnick and Kolinski, 1991) have been proposed based on the considerations of statistical physics and mathematical statistics to describe protein folding process.

The protein folding problem is three different problems: the folding code—the thermodynamic question of how a native structure results from the interatomic forces acting on an amino acid sequence; protein structure prediction — the computational problem of how to predict the native structure of a protein from its amino acid sequence and folding speed (Levinthal's paradox) - the kinetic question of how a protein can fold so fast.

The hydrophobic interaction was believed to be a dominant component to the folding code that the folding code is distributed both locally and non-locally in the sequence, and that native secondary structures are more a consequence than a cause of folding forces. Much of the current trends in deciphering the folding code include use of purely physics-based methods, without knowledge derived from databases (such as statistical energy functions or secondary structure predictors), to explore native structures and folding processes. Advantages of these physics based approaches would be the ability to predict conformational changes, such as induced fit, a common and important unsolved problem in computational drug discovery; the ability to understand protein mechanisms, motions, folding processes, conformational transitions etc. However physics-based methods are currently limited by some inaccuracies in the force-fields and by huge computational requirements. Continuous experimental supplementation therefore is needed to upgrade these theoretical inaccuracies in order to improve predictions. The question of folding mechanism has driven major advances in folding experiments. Key advances include ability to measure folding events on timescales faster than a few milliseconds and to monitor individual chain monomers during folding. All these
developments were complemented by fast laser temperature-jump methods (Nguyen et al., 2003); mutational methods that give quantities called $\Phi$ values, which can identify those amino acids that control the folding speed; FRET methods that can watch the formation of particular contacts (Haas, 2005, Deniz et al., 2000); hydrogen exchange methods that see structural folding events (Zhou et al., 2006; Maity et al., 2004, Gorski et al., 2004); and extensive studies on model proteins, including cytochrome c, chymotrypsin inhibitor 2, barnase, apomyoglobin, src, fyn SH3 domains, proteins L and G, WW domains etc. These studies have led to the revelation that protein folding speeds, which vary over more than eight orders of magnitude, correlate with the topology of the native protein. Fast folders have mostly local structures like helices and tight turns, whereas slow folders, though not always, usually have more non-local structure, such as $\beta$ sheets (Silva et al., 2005; Shea et al., 2002).

**Protein destabilization, misfolding and aggregation**

Interest in the problem of protein destabilization or misfolding and aggregation has increased in recent years due to sharp rise in the number and volume of therapeutic proteins produced commercially and the recognition of the central role of protein aggregates in degenerative diseases.

Nonnative protein aggregation describes the assembly from initially native, folded proteins of aggregates containing nonnative protein structures (Chi et al., 2003). Aggregation is often irreversible. Protein aggregation behaviors such as onset, aggregation rate, and the final morphology of the aggregated state (i.e., amorphous precipitates or fibrils) have been found to depend strongly on the properties of a protein's solution environment, such as temperature, pH, salt type, salt concentration, cosolutes, preservatives, and surfactants (Campioni et al., 2008; Nielsen et al., 2007; Sasahara et al., 2007; Chi et al., 2003; Manno et al., 2004) as well as the relative intrinsic thermodynamic stability of the native state (Galka et al., 2008; Scharnagl et al., 2005; Hill et al., 2005, Minton, 2000).

The reaction order for the rate-limiting step determines the apparent order of the aggregation reaction (Chi et al., 2003). A number of proteins have been found to follow first order aggregation kinetics (Golub et al., 2008; Hoiberg-Nielsen et al., 2006; Weijers,
suggesting that the rate-limiting step is unimolecular (e.g., a conformational change) but some of the molecules like rhG-CSF show bimolecular reaction following a second-order reaction limited by collision frequency which shows that protein conformation alone cannot explain the aggregation behaviors. The aggregation transition states of some proteins have been identified as a structurally expanded species within the protein native state ensemble (Clark, 2005; Chi et al., 2003).

Protein molecules also assemble to form higher order aggregates. Molecular assembly processes occur as a result of attractive intermolecular interactions (Chi et al., 2003). Thus, an understanding of protein aggregation also requires information about the nature and magnitude of these interactions. The osmotic second virial coefficient \( B_{22} \) is a thermodynamic solution parameter that directly quantifies overall protein–protein interactions on the molecular level. Positive \( B_{22} \) values indicate the overall dominance of repulsive forces between protein molecules, where protein–solvent interactions are favored over protein–protein interactions. Negative \( B_{22} \) values reflect overall attractive forces between proteins, with protein–protein interactions being favored over protein–solvent interactions (Sinibaldi et al., 2008; Winzor et al., 2007). Morphology of the solid phases formed are predominantly determined by the mechanisms of molecular approach, reorientation, and incorporation of native proteins, which are governed by the strength and range of protein colloidal interactions (Hoyer et al., 2004). Protein stability and aggregation is therefore controlled by both conformational stability and colloidal stability, and, depending on the solution conditions, either could be rate limiting. To successfully stabilize protein against aggregation, solution conditions need to be chosen not only to stabilize the protein native conformation but also to stabilize protein against attractive intermolecular forces.

Proteins are only marginally stable and are highly susceptible to both chemical and physical degradation. Chemical degradation refers to modifications involving covalent bonds, such as deamidation oxidation, and disulfide bond shuffling (Chi et al., 2003; Hau et al., 2002 Reubsaet et al., 1998). Physical degradation includes protein unfolding, undesirable adsorption to surfaces, and aggregation (Bolanos-Garcia, 2008; Schmitt et al., 2007). Nonnative aggregation is particularly problematic because it is encountered routinely during refolding, purification, sterilization, shipping, and storage
processes. Aggregation can occur even under solution conditions where the protein native state is highly thermodynamically favored and in the absence of stresses. Partially unfolded states however adopt a collapsed conformation that is more compact than the unfolded state and has substantial secondary structure and little tertiary structure, have large patches of contiguous surface hydrophobicity and are much more prone to aggregation therefore than both native and completely unfolded conformations (Fink, 1995). There exists an ensemble of native substates with a distribution of structural expansion and compaction. Kendrick et al. showed that the aggregation of rhIFN-γ proceeds through a transiently expanded conformational species within the native state ensemble (Kendrick et al., 1998).

Outstanding progress has been made in the development of therapeutic strategies targeting these diseases. Three promising approaches (Rochet, 2007) include: (1) inhibiting protein aggregation with peptides or small molecules identified via structure-based drug design or high-throughput screening; (2) interfering with post-translational modifications that stimulate protein misfolding and aggregation; and (3) up regulating molecular chaperones or aggregate-clearance mechanisms. Ultimately, drug combinations that capitalise on more than one therapeutic strategy will constitute the most effective treatment for patients with these devastating illnesses.

How do cosolvents affect protein stability, dynamics, folding and aggregation?

Effect of cosolvents on protein stability

Thermodynamic description for a ternary solution can be described with components water (w), cosolvent (x) and protein (p). The corresponding concentrations of the solutes are \( n_w \) and \( n_p \), respectively (Scharnagl et al., 2005). Each component is characterized by the respective chemical potential \( \mu_i = \mu_i^0 + RT \ln(c_i) \) (i =w, x, p) with activity coefficient \( c_i \) and standard chemical potential \( \mu_i^0 \). The protein is present at infinite dilution, hence, \( c_p = 1 \). Nonideal behavior arises due to the dependence of the standard states of cosolvent and protein on the concentration of the cosolvent.
Cosolvent-induced effects on protein stability manifest themselves through changes in the free energy of the solvated protein. The change in protein stability can be related to the changes in the concentration of the components of the solvent around the protein. A change in the number of water and cosolvent molecules in the protein phase results from the interaction with the protein (either direct or indirect). This may lead to an excess population of one sort of molecules in the vicinity of the protein compared to the reference phase. Possible interaction mechanisms are (i) association of cosolvent molecules with the protein eventually in competition with water; (ii) inaccessibility of the protein to a specific cosolvent molecule due to steric reasons; (iii) solvent reorganization. Thereby, the range of influence of the protein on the local solvent composition does not only include directly bound molecules. Due to long-range interactions, the influence extends over several layers of waters. Cosolvent can penetrate into the solvation shell and exchange with water and hence, $C_{px}$ has contributions from both water and cosolvent affinities. The addition of cosolvent therefore creates the possibility to probe the role of hydration water for the stability of proteins. The influence of the cosolvent on protein stability is directly related to the change in preferential binding coefficient of the cosolvent (or, equivalently, the change of hydration) upon denaturation. A change in one
parameter implies as a direct consequence, changes in all other parameters. A requirement for the stabilization of the native structure by a cosolvent is a negative value of \( \Delta C_{px} \), suggesting that the proteins are preferentially hydrated or the cosolvent is preferentially excluded from the local domain of the protein. Therefore, in order to stabilize the native state, the value of \( C_{px}(D) \) must be more negative than \( C_{px}(N) \), a condition tantamount with the finding that the cosolvent is more strongly excluded from the denatured protein than from the native one.

On a microscopic scale, the stabilization or denaturation of a protein due to the interaction with a cosolvent can be understood in terms of the interactions of the cosolvent molecules with those residues which come into contact with water upon denaturation. The formation of excess or deficit population in the vicinity of the protein with respect to a certain component of the solvent has contributions from (i) nonspecific steric exclusion of cosolvent or solvent molecules on the basis of excluded volume as an entropic consequence of a higher population of one component, and (ii) specific binding in competition with water molecules. The relative contributions from the two processes depend on the nature of the cosolvent, i.e. on its size, flexibility, charge, polarity, etc. and on the surface topology of the protein. The interpretation of cosolvent-induced stabilization of a protein in its native state on the basis of excluded volume effects relates stabilization to the increase of the steric–repulsive interactions in the water–cosolvent mixture relative to the pure water solvent. The stabilization or destabilization due to this effect depends on the geometrical structure of the denatured state in relation to the native state as well as on the average size of cosolvent molecules compared to water. If cosolvent molecules are significantly larger than water molecules, cosolvent is excluded from a certain volume shell around the protein, whereas water is not. Since excluded volume is proportional to the solvent-accessible surface area (SAS) of the protein. The exclusion of cosolvent due to steric reasons therefore tends to stabilize protein conformations with lower SAS, a condition found mostly in compact proteins.

Since the denatured state usually is characterized by a more open structure with larger SAS, excluded volume will stabilize the native state. In the surface tension model (Jackson and Sternberg, 1994; Lee and Timasheff, 1981), the stabilization of the native state is attributed to the increasing surface tension of water upon the addition of cosolvent.
which increases the free energy requirement to accommodate the increased surface area of more open denatured states. On the other hand, preferential binding can enhance or attenuate the stability. Le Chatelier's principle implies that if a cosolvent binds preferentially to the native protein, the native state will be stabilized and denaturation becomes less favorable as the cosolvent concentration increases. Conversely, cosolvents that bind preferentially to the denatured protein will destabilize the native state. Another model which is employed to analyze cosolvent induced stability changes is the transfer free energy model (Auton et al., 2008; Alonso and Dill, 1991). Transfer energies of individual amino acids are used to predict the transfer free energy of the whole protein. In this model, steric effects and binding contributions are simultaneously included. The energy transfer data indicate that the osmophobic theory explains protein stabilization. The transfer Gibbs energy values estimated from the $\Delta G_{tr}$ values were used for the analysis of protein stability measurements in osmolytes. In such analysis, both denatured structure information and native structure information are needed. The transfer Gibbs energy values, from water to an osmolyte, of the denatured and native states were calculated using Eq.

$$\Delta G^0_D - \Delta G^0_N = \Sigma \alpha \Delta G_{tr}$$

Here, $\alpha$ represents the fractional exposure of an amino acid residue in the native or denatured state against each amino acid, and $\Delta G_{tr}$ represents the transfer Gibbs energy of the amino acid. The $\alpha$ value for the native state ($\alpha_N$) is obtained from the accessible surface areas (ASAs) of the amino acid residues in the native structure and the amino acids. In the case of the denatured state, we have to prepare a denatured structure for such analysis. $\alpha_N = \Sigma \text{ASA}_N / \text{ASA}_{\text{amino}}, \alpha_D = \Sigma \text{ASA}_D / \text{ASA}_{\text{amino}} $ (3) Here, $\text{ASA}_{\text{amino}}, \text{ASA}_N,$ and $\text{ASA}_D$ are the ASAs of an amino acid, the native state, and the denatured state, respectively.

Various thermodynamic models for the cosolvent effect on protein stability as well as the volumetric analysis can be combined in the rigorous statistical mechanics frame of the Kirkwood-Buff theory (Schellman, 2005), which allows for relating the excess number of cosolvent molecules around the protein to their radial distribution function, which is available from experimental as well as from computer simulation data.
The pair correlations of proteins with water and osmolytes are the determining structural factors for the proteins’ response to the presence of osmolytes (Rosgen et al. 2007). Only very recently, however, has effort been devoted to calculate thermodynamic solvation of proteins from experimental data (Shimizu et al., 2006; Rosgen et al., 2005; Schurr et al., 2005; Shulgin and Ruckenstein, 2005; Shimizu, 2004; Shimizu and Boon 2004). Solvation effects of water and osmolyte as well as their impact on protein stability is correlated. Hydration is found to be more sensitive to osmolyte size than osmolyte type (excluded volume effect of hydration). Hydration effects are of minor importance for the unfolding energetics of proteins in different osmolyte solutions and for the energetics of the bulk osmolyte. Water self-correlations are largely unaffected by osmolyte concentration and type. The energetic contribution of water self-hydration turns out to be insignificant to both the energetics of osmolytes and proteins in osmolyte solutions. This is consistent with a finding that the water–water correlation in the vicinity of the chaotrope (urea) and the cosmotrope (TMAO, trimethyl-amine-N-oxide) are very similar with regard to both angles and distances between water molecules (Gallagher and Sharp, 2003). Differences between osmolytes stem mostly from osmolyte self-solvation. Energetically, neither bulk water nor protein hydration is the main player in osmolyte concentration-dependent effects on protein stability and osmolyte energetics. Very predictable and monotonic solvation properties might be prerequisite for a molecule to be useful as a biological osmolyte. The very simple activity coefficient (Rosgen et al., 2004 a, b) and solvation behavior (Rosgen et al., 2005; 2007) seems to be a general property of osmolytes. Fine-regulation of the chemical activities of cellular components have to rely on the constant solvation behavior without sudden switches.

Thus, the combined effect of factors like solvophobicity, surface tension, excluded volume, water structure changes and electrostatic repulsion are all responsible for preferential exclusion and the effects exclusion has on protein properties (Auton et al., 2006). The effects lead to significant decrease in protein conformational entropy in native and denatured states in native conditions contributing to the mechanism of protein stabilization by stabilizing cosolvents like naturally occurring osmolytes (Chen et al., 2006).
Thus, the protein stability has thermodynamic as well as kinetic aspects. Thermodynamic stability, which is a positive value for the unfolding free energy at physiological temperature does not guarantee that the protein will remain in the native state during a given time scale, since irreversible protein alterations (even if they occur from lowly populated unfolded or partially unfolded states) may deplete the native state in a time-dependent manner. A strong scanning-rate dependence of the thermal denaturation transitions therefore is a signature of kinetic control for many protein systems (Remmele et al., 2005; Jayaraman et al., 2005; Sanchez-Ruiz et al., 1988). Many proteins therefore are likely to be naturally selected to have significant kinetic stability which include examples like α-lytic protease, Cu/Zn superoxide dismutase, the viral capsid protein SHP11 and human low-density lipoprotein. The interest of understanding protein kinetic stability is emphasized by the fact that some emerging molecular approaches to the inhibition of amyloidogenesis focus on the increase of the kinetic stability of the protein native state. Moreover, kinetic stability may be of considerable biotechnological importance as the proteins and the solvent conditions employed in technological applications often imply irreversible denaturation and kinetic control of the stability. Recent work exploring the kinetic stability of lipase (Rodriguez-Larrea et al., 2006) suggests that a solvation barrier arising from the asynchrony between breaking of internal contacts and water penetration may contribute to the kinetic stability. Thus general existence of water-solvation barriers that contribute to protein kinetic stability seems to be consistent with the fact that the dehydrated and lyophilized enzymes denature irreversibly at very high temperatures, indicating very high kinetic stability. Thus enhancing protein stability for biotechnological applications may in many cases mean enhancing protein kinetic stability and solvent mediated approach could be adopted to improve the kinetic aspects of the stability too.

Effect of cosolvents on protein dynamics

Stabilizing forces that protect proteins from denaturation may or may not be distinct from those forces that rigidify the protein. Crystallographic studies have identified thermophilic proteins as possessing smaller Debye-Waller factors than their mesophilic counterparts, suggesting a link and inverse correlation between thermal
stability and protein dynamics (Vihinen, 1987). In addition, there have been a number of studies, both experimental and computational, that provide a correlation for (Wagner and Wuthrich, 1979; Lazaridis et al., 1997; Zavodszyky et al., 1998; Svingor et al., 2001; Tsai et al., 2001) or show no evidence for (Colombo et al., 2008; Fitter and Heberle, 2000; Hernandez et al., 2000; Fitter et al., 2001; Grottesi et al., 2002) a link between dynamics and stability.

Proteins do not exist in a unique conformation, but can exist in a very large number of somewhat different structures called conformational substates. So an instantaneous structure is characterized by the positions of all \( N \) atoms in the protein, the hydration shell, and some part of the bulk solvent. An instantaneous structure is a point in the conformation space and the protein motions are transitions among these points. This establishes that the solvent is the master and the protein the slave, as far as large-scale motions (which involve the largest amplitude motions that proteins make so that the fluctuating solvent should constrain these motions) are concerned (Frauenfelder et al., 2006).

The studies demonstrate that stabilizing osmolytes restrict the increase in conformational space in the presence of chemical denaturants, causing more restricted, native-like protein fluctuations (Doan-Nguyen and Loria, 2007). This could possibly limit access to higher energy conformational substates that would ultimately lead to protein denaturation. Meanwhile hydrogen exchange studies on proteins (Wang et al., 1995; Idiyatullin et al. 2003) have revealed that rate of exchange of slow exchanging amide protons are further slowed down in presence of stabilizing osmolytes. Because stabilizing osmolytes oppose an increase in protein surface area exposure, slowing of the hydrogen exchange rates strongly indicates that the cosolvents affect the dynamics of the compact unfolded state ensemble of the protein where the exchange was protected to a significant extent. The fact that other components (intermediate and fast) of the hydrogen exchange rates could not be affected in a significant way suggested that the compatibility of stabilization of the osmolytes is decided by the unperturbed dynamics of the functionally critical flexible and dynamic regions of the native state ensembles. Thus naturally occurring stabilizing cosolvents seem to be evolutionary selected on their ability to discriminate between the dynamics of various conformational ensembles.
The hydration water dynamics and their dynamical coupling with the protein are essential for protein dynamics and biological function. Inelastic neutron scattering experiments on Staphylococcal nuclease (Nakagawa H et al., 2008) looking into the effect of hydration on protein dynamics at differing hydration levels have revealed that the partial hydration is sufficient to affect the harmonic nature of protein dynamics, and that there is a threshold hydration level to activate the anharmonic motions. Thus, hydration water has been found to control both the harmonic and anharmonic protein dynamics and the hydration effects are strongly dependent on both temperature and hydration. These results indicate that to understand protein dynamics the hydration water dynamics should be revealed.

**Cosolvents and enzyme activity**

Stabilizing osmolytes are preferentially excluded from the protein domain and therefore no direct interaction between the osmolyte and the protein is expected to bring a change in $K_m$ and $k_{cat}$ values. At the same time compatible osmolytes might affect the association of substrate with enzyme through solvation effects on substrates or enzyme active sites, or by means of effects on the thermodynamic activity of substrates or enzyme (Abel et al., 2008; Agarwal, 2005; Chang et al., 2008). Polyols are observed to bring no significant effect on $K_m$ and $k_{cat}$ of enzymes like lysozyme and RNase-A (Haque et al., 2005). The lack of effect on activity parameters could be due to that polyols have little or no effect on the solvation properties of substrates and enzyme active sites. Similar effects are also observed in case of prolines (Diamant et al., 2001). Compatible osmolytes must therefore be able to discriminate between the solvation characteristics of structurally and functionally critical domains of the protein.

**Effect of cosolvents on protein folding**

Solvation has a dramatic effect on the energy landscape of the proteins as the concept of amino acids having helical and beta sheet forming propensity incorporates the environmental solvent effects (Levy et al., 2001). This means that kinetic studies looking into transition from unfolded state to folded state can provide insights on how solvent could possibly alter these accumulations of various conformational ensembles depending on their respective stabilization/destabilization under defined solvent conditions. This
Introduction

should however also depend on whether solvent mediated equilibrium effects reflect native state stabilization or unfolded state destabilization.

Studies of unfolding rates of some proteins like DHFR with varying solvent ionic strength have revealed that slower unfolding rates are followed at high ionic strength conditions without altering unfolding rates at native conditions suggesting kinetic stabilization of the protein predominantly from destabilization of the unfolded states as observed for some thermophilic enzymes (Gloss et al., 2008). This indicates that haloadaptation harnesses the effects of increased salt concentrations on the properties of the aqueous solvent to enhance protein stability. Also, simulation studies looking into the effects of incorporating water molecules while observing peptide folding in silico, have revealed that peptide conformers with high solvent-accessible hydrophobic surface area have low hydration density around hydrophobic residues, whereas a concomitant higher hydration density around hydrophilic residues was observed. The dewetting effect was found to stabilize the fully folded state. The results suggested that dehydration-driven solvent exposure of hydrophobic surfaces may be a significant factor determining peptide conformational equilibria and could be a potential driving force in peptide folding (Daidone et al., 2007). This enhances the possibility that hydrating cosolvent molecules could significantly alter the rates of folding depending on their respective solvation effects varying conformational entities accumulating during protein folding. This is further evidenced by studies on cutinase exploring the effect of trehalose on its folding which is found to favor an intermediate off-folding pathway and might even decelerate the distinct folding event (Melo et al., 2003). On the other hand, pressure-jump-induced kinetics on RNase A has also implicated a two-dimensional energy surface containing a pressure- and temperature-dependent barrier between two isomeric unfolded states (Font et al., 2006). Analysis of the activation volume of the kinetic phases revealed a temperature-dependent shift of the unfolding transition state to a larger volume. However, the observed effect could be compensated by glycerol. This indicated towards the hydration dependent folding/unfolding mechanism of ribonuclease.

The fact that the folding reaction is significantly influenced by the nature of the bulk solvent (Parker et al., 1997) is evidenced by solvent viscosity effects on the folding rates. Viscogenic agents however are known to affect folding rates not only by increasing
solvent viscosity but also by increasing protein stability (Jacob et al., 1999; Pradeep and Udgaonkar, 2007; Mishra et al., 2007). Studies with both protein and solvent confined within a space limit has shown that solvent-mediated affect could lead to destabilization of the native state and unfolding happened to a relatively compact form of the unfolded state (Lucent et al., 2007). Thus, the confinement of solvent has a significant impact on protein kinetics and thermodynamics.

**Effect of cosolvents on protein aggregation**

Presences of various kinds of additives have been found to prevent protein aggregation (Dong et al., 2004). Ligands and cosolutes that alter protein conformational stability also influence the rate of formation of non-native aggregates (Chi et al., 2003). For example, in the presence of polyanions, aggregation of acidic fibroblast growth factor (Fan et al., 2007) and native recombinant keratinocyte growth factor (Derrik et al., 2007) is greatly inhibited. It has also been shown that the addition of weakly interacting preferentially excluded solutes can reduce the rate of protein aggregation. For example, sucrose has been shown to inhibit aggregation of hemoglobin (Kerwin et al., 1998), rhIFN-γ (Webb et al., 2001), keratinocyte growth factor (Chen and Arakawa, 1996), immunoglobulin light chains (Auffray and Rougeon, 1980), and rhGCSF (Thirumangalathu et al., 2006).

Protein folding aids can be categorized into two groups. Molecules like acetamide, acetone, thiourea and L-arginine stabilize unfolded protein or folding intermediates. The presence of these additives decrease the folding rate with increase in their concentration and can minimize the aggregation at a particular concentration. The other group include molecules like glycerol which act as protein stabilizers (Baynes and Trout, 2004; Mishra et al., 2007). In the presence of these kinds of folding aids, both the refolding rate and yield were enhanced by increasing their concentration to a proper value. So the cooperative application of the two kinds of folding aids could result in favorable refolding rate and better protein yields. Some of the aggregation suppressors like ammonium sulphate salts have also been correlated to their surface tension property to be described as one among the factors to prevent thermal aggregation (Hirano et al., 2007). Cosolvents as folding aids or aggregation suppressers could also act indirectly by
facilitating molecular chaperones enhancing their ability to assist refolding and minimize aggregation (Lange et al., 2005).

**Outline of the current research problem**

Proteins are the major biomacromolecules involved in variety of physiological functions. They are the polypeptides which fold to acquire a particular three dimensional conformation so as to carry out the given function. As discussed earlier that any alteration in the specific conformation of proteins may result in variety of physiological disorders. To avoid any such alterations, polypeptide chains have to first get properly folded from the unfolded state and the folded state has to be stable enough under varying conditions to carry out the required function. Thus either misfolding or destabilizations of the proteins lead to serious disorders.

Cataract is one such disease where the constituent proteins of the lens gets aggregated that interfere with its transparency subsequently leading to blindness. Mammalian lenses are made of long fiber cells, enclosing a major cytoplasmic component lens–specific proteins, the crystallins. α-crystallins (molecular weight ~7,50,000-1,200,000) and β-crystallins (β\textsubscript{h}~180,000;β\textsubscript{1}~60,000) are heterogenous oligomeric proteins, while γ-crystallins are monomers (molecular weight~ 20,000) (Bloemendal and de Jong, 1991). These constitute an estimated 35% of the net weight of the lens. Using a mixture of these different sized protein assemblies to fill the lens fibre cells ensures polydispersity and prevents crystallization. In order to fulfill their optical function, crystallins have to be first and foremost soluble. Since they have to last longer in relation to the whole life span of the organism they must also be stable. Crystallins contribute to the transparency and refractive power of the lens by short-range interactions among themselves and cytoskeletal elements in a highly concentrated matrix (Bloemendal and de Jong, 1991; Delaye and Tardieu, 1983). A disruption in this interactive order (either through alteration in solubility or in stability) may be caused by different types of protein condensates, such as aggregates, coexisting liquid phases or crystals. Any such disruption results in increased light scattering and lens opacity or cataract.

Cataractous crystallin proteins may be divided into two categories: α-crystallins
and β γ-crystallins. α-crystallin is a member of small heat shock proteins acting as molecular chaperone (Raman et al., 1997), is thought to bind to unfolded polypeptide chains during the times of stress and is thus crucial for preventing aggregation. It forms polydisperse multimers having molecular masses ranging from 300 to 1200 kDa, depending on the solvent conditions and other variables. It is a predominantly β-sheet with less than 20% helix content.

β γ-crystallins are however small (20-30 kDa) proteins primarily composed of antiparallel β-sheets. β-crystallins and γ-crystallins are structurally similar. They both comprise four Greek key motifs separated into two domains. β-crystallins form domain-swapped dimers in solution owing to their flexible sequence. Truncation and deamidation of βB1 leads to altered conformation. These alterations happen extensively in ageing human lenses and may be important for age onset cataract formation.

The γ-crystallins are monomeric in solution. In addition, γ-crystallins are the only known crystallins having attractive forces between molecules and are therefore more likely to form cataractogogenous aggregates (Tardieu et al., 1992). Also, as with age, chaperone action of α-crystallin gets reduced due to destabilization, therefore it becomes necessary to set aggregation prone γ-crystallins as an alternative target to design drugs for cataract. γ-crystallins differ from the α- and β-crystallins in one important aspect: the interactions between the γ-crystallins are attractive (Tardieu et al., 1992). This feature reduces the osmotic pressure in the lens but it also makes γ-crystallins more susceptible to aggregation and phase separation, a phenomenon that diminishes the homogeneity of the lens and causes cataract. Also α-crystallin, in order to show its chaperone behavior, forms complexes with β- and γ-crystallins. In vitro studies on these complexes have revealed that α-β-crystallin complexes remain soluble for a long time in contrast to α-γ-crystallin complexes that form precipitates as a function of time. Such differential behavior in vitro provides a strong indication that γ-crystallins pose a more severe cataract risk than β-crystallin. Human γD crystallin is the model of our choice as this is the most abundant among the expressed γ-crystallins in the lens.

However, Human γD crystallin as the model for solvent mediated protein stability
Figure 2: Conformational constraints in Human γD crystallin showing two β greek key domains joined by a hinge with A. buried cysteines and B. prolines. (PDB code:1HK0)
studies has been selected on the basis of some additional background studies:

A) Proteins tend to compromise between rigidity (for stability) and flexibility (for folding, function and degradation) as a result of which the free energy of stabilization of globular proteins in solution is equivalent to only a few weak intermolecular interactions. Stabilizing cosolvents or ligands therefore can bring about additional increments in this free energy of denaturation. These can play their role in both ways either by bringing in the enthalpic effects or through entropy by reducing the flexibility (cystine bridges, increased proline content etc.), or by water release from residues buried upon folding and association. Protein stability and function can also be maintained by increased ion binding and glutamic acid content (as in halophiles) allowing it to compete for water during high salt conditions. Similarly, proteins facing the outside extremes of pH adopt to the conditions by possessing anomalously high contents of ionizable amino acids in order to have buffering ability. Thus, the basic mechanisms of molecular adaptation include changes in packing density, charge distribution, hydrophobic surface area and in the ratio of polar:non-polar or acidic:basic residues (Jaenike 2000). Human γD crystallin seems to have all these adaptations as it possesses high packing density, distribution of acidic and basic residues, calcium binding property (to possibly provide the protein with ion buffering ability), distribution of surface charges (to facilitate protein-protein attractive order) and surface salt bridges (to overcome the solubility concerns under high protein concentration).

B) Proteins can exist in concentrated or crowded solutions (Minton, 2001; Hall and Minton, 2003; Rivas and Minton, 2004) where the concentration of which might not be high in a solution, is forced to exist in a considerably reduced volume fraction of the total solution volume due to the presence of an inert solute at high concentration. Human γD crystallin is different in a way that they exist in crowded conditions but crowding is not solute mediated, instead the protein itself existing at very high concentration create the protein-protein mediated crowding. Concentrated and crowded solutions are often encountered in pharmaceutical milieu as well as in physiological environment. The behavior of a protein molecule in such an environment is significantly affected by the presence of other molecules. The primary consequence is the alteration of the activity or effective concentration of protein in solution, which further results in change in protein
structure, function and its stability (Minton, 2005; Saluja and Kalonia, 2008). In physiological systems, the consequences of high protein concentration coupled with rather minor structural alteration and sequence mutations are expressed in the form of various diseases and disorders due to protein assembly processes (Hardy and Gwinn-Hardy, 1998; Koo et al., 1999 and Lansbury, 1999). Examples include cataract (Stradner et al., 2004), neurodegenerative diseases including Alzheimer's and Parkinson's disease (Meehan et al., 2004), systemic amyloidosis (Harper and Lansbury, 1997), polyglutamine disorders like Huntington's disease (Koo et al., 1999), etc. From a pharmaceutical perspective, high protein concentrations in solutions pose formulation challenges originating from protein solubility, manufacturing challenges due to high viscosity of some of these solutions and often result in compromised stability of the protein in solutions with regard to self-association and aggregation (Shire et al., 2004).

C) Using small molecules to bring in desired conformational changes or to inhibit protein misfolding in recent times has helped develop potential therapeutic strategies against misfolding diseases (Rochet, 2007). Small molecules like members of polyol series (glycerol, erythritol, xylitol and Sorbitol) and amino acids and their methyl derivatives (glycine, sarcosine, betaine, etc.) are well known to occur in nature as protein stabilizers (Arakawa and Timasheff, 1982a and b, 1983, 1985). These osmolytes stabilize proteins through preferential hydration and have been found to enhance thermal stability of several proteins (Ahmad, 1999; Radha et al., 1998; Kaushik and Bhat, 1999; Kaushik and Bhat, 2003). Studies on the effect of osmolytes on protein folding have also been reported (Mukaiyama, 2008). Refolding studies on Citrate synthase (Mishra and Bhat, 2005) and on Carbonic anhydrase (Yoshimoto et al., 2003) have revealed that by carefully manipulating the solvent environment one could enhance the refolding of the proteins considerably. Meanwhile osmolytes are also reported to influence aggregation and amyloid fibril formation in different ways (Yang et al., 1999; Ignatova and Gierasch, 2006) All this information opens tremendous scope for utilization of these molecules straight away or molecules derived from these as potential therapeutic molecules against many protein destabilization or misfolding related disorders (Tanaka et al., 2004, 2005).

In light of these background studies, we have carried out detailed conformational studies of the wild type Human γD crystallin to understand how its physico-chemical
In this context therefore studies on solvent mediated effects on stability, folding and aggregation were carried out with following objectives:

- To study the conformational properties of Human \( \gamma \)-D-crystallins extensively and alteration in these properties when subjected to varying solvent conditions in terms of pH, ionic strength, presence of naturally occurring osmolytes of polyols, Polyethylene Glycols, amino acids and their derivatives etc., so as to optimize and define proper conditions for maximum stability and least aggregation.

- To characterize the aggregation prone intermediate on the folding pathway and to specify alteration in its population/physico-chemical properties under varying solvent conditions so as to obtain maximum refolding yield, in turn defining the best stabilizing conditions.

- To further explore the role of non-specific attractive interactions among the \( \gamma \)-D-crystallins and the possible solvent mediated factors involved therein.