INTRODUCTION
Protein folding is a process by which a linear sequence of amino acids gives rise to the native three-dimensional structure which is the functional state of a protein. The significance of protein folding is not limited to academic research alone, it also has tremendous importance in the biotechnology and pharmaceutical industries engaged in recombinant protein production. Understanding the mechanism of protein folding has far reaching implications in human and animal diseases, structure prediction from amino acid sequence, protein design and nanotechnology (Anfinsen, 1973; Jaenicke, 1987, 1999; Hartl, 1996; Radford, 2000).

The work on protein denaturation-renaturation began as early as 1920s when Anson and Mirsky showed that denaturation of hemoglobin is reversible (Anson and Mirsky, 1931; Anson, 1945). By these experiments it became clear that folding occurs in a spontaneous way. However, it was Christian Anfinsen and coworkers who in their classical work on denatured and reduced ribonucleaseA (RNase A) convincingly confirmed that the information required for the native structure formation resides in the amino acid sequence. After these studies on RNase A, work on protein folding began to gather momentum and the field of protein folding has emerged as a very important topic in basic as well as applied research in biology.

Advancements in recombinant DNA technology have made it possible to clone and express commercially important and rare proteins in large quantities in the prokaryotic host cell such as E.Coli. Although high levels of expression of proteins are achieved by using standard recombinant techniques, yet the expressed proteins often form insoluble and amorphous aggregates called inclusion bodies. Recovering the soluble and native proteins from inclusion bodies poses a major challenge and is a bottleneck in the biotechnology and pharmaceutical industries. Recovery of native proteins requires solubilization by chaotrope agents and subsequent in-vitro refolding. During in-vitro refolding, aggregation often competes with correct folding. Understanding the refolding process and devising strategies for preventing protein aggregation is of great...

In the recent years a growing number of human and animal diseases such as scrapie, Cruetzfeldt-Jakob, familial insomnia, Alzheimer's disease, cystic fibrosis, amyotrophic lateral sclerosis, \(\alpha_1\)-antitrypsin deficiency, etc. are known to be caused by either misfolding or by aggregate deposition or by mislocalization owing to misfolding of certain proteins. Prevention of these diseases requires an in depth understanding of the folding and the aggregation process, which in turn would be extremely helpful to design or search for molecules that could inhibit this process (Thomas et al., 1995; Prusiner, 1997; Kelly, 1998; Dobson, 1999; Dobson, 2001a,b; Soto, 2001).

As the complete human genome sequence has been unraveled, it is becoming extremely difficult to keep pace with the discovery of new proteins and solving their structure. It is, thus, imperative to develop more precise and accurate methods for structure prediction from amino acid sequence (Crawford, 1999; Heinemann et al., 2000). Understanding principles governing protein folding would be of great importance in designing proteins or peptides with novel sequences which could be used for therapeutic purposes. There are also several other phenomena such as ligand binding, protein-protein interactions and several aspects of structure-function relationship that require the understanding of the protein folding process (Nath and Udgaonkar, 1997). In order to understand the folding process it is essential to have the understanding of the native, the intermediate and the unfolded state and the interactions that stabilize or destabilize these states.

Native state

Native state is the functional and the fully folded state of a protein which is made up of single or multiple domains. Usually, smaller proteins possess a single domain while larger proteins can have several domains. In terms of physico-chemical properties, the native structure of a protein is highly cooperative in nature and from a physical point of
view represents a macroscopic system (Privalov, 1992). Protein molecules are also highly
dynamic and flexible. The static structure may not be entropically favorable, and being
flexible and dynamic helps them gain entropy without losing enthalpic contributions to

Non-covalent interactions play an important role in the stabilization of proteins.
Formation of the native state from an unfolded state leads to a decrease in the entropy of
the system. This situation is thermodynamically unfavorable. To overcome this effect,
there should be a rearrangement of interactions among different atoms (Privalov, 1992).
There is only -5 to -10 kcal/mole difference between the free energies of the native and
the denatured states (Privalov, 1979, 1989). The non-covalent forces that contribute to the
stability of native proteins are hydrophobic interactions, hydrogen bonding, van der
Waals interactions and electrostatic interactions (Dill, 1990; Jaenicke, 1991). Hydrophobic interactions refer to the tendency of non-polar compounds to associate in
aqueous media (Privalov and Gill, 1988) and are supposed to increase with increasing
temperature. When an unfolded random coil protein molecule is placed in water, the non­
polar residues tend to get buried inside the interior of the protein (Tanford, 1962).
Hydrogen bonds can contribute to the stability of proteins by forming intra as well as
inter molecular bonds (London, 1937; Kauzman, 1959; Klotz and Farnham, 1968; Jencks,
1987, Creighton, 1993; Murphy, 1995; Pace et al., 1996; Tanford, 1997; Takano et al.,
1998).

Denatured State

The theory of denaturation was proposed as early as in 1931 by Hsein Wu, in
which the denatured state is identified as alterations of the highly compact and ordered
structure into more or less open structure. This open structure can be achieved without
change in the amino acid sequence, i.e., without severance of any of the primary chemical
bonds which join one amino acid to the other (Wu, 1931; Tanford, 1961). Due to the lack
of structure, the unfolded state of the protein is less understood compared to the folded
state and there is no single denatured state of the polypeptide chain (Dill et al., 1995).
Instead, they form a poorly understood assembly of partially unfolded states, collectively known as the denatured state (Shortle, 1996). Unfolding leads to a complete loss of the secondary structural elements and an increase in intrinsic viscosity of the solution.

High concentrations of Guanidine hydrochloride (GdmCl) or urea often lead to complete protein denaturation. As suggested by thermodynamic evidence and statistical considerations, based on estimates of stabilization of energies and conformational entropies, several denaturing conditions lead to the same extent of unfolding (Creighton, 1978; Privalov, 1979). However, spectral properties show significant differences. The extent of denaturation of the order $U_{GdmCl} > U_{\text{urea}} > U_{\text{heat}} > U_{pH 2.0}$ is widely accepted (Jaenicke, 1999). NMR data have shown that even under extreme denaturing conditions some local residual structure is still preserved (Neri et al., 1992).

Models of Protein Folding

Framework model (Kim and Baldwin, 1982, 1990)

The framework model postulates that the native secondary structure is formed before the tertiary structure is locked in place. According to this model, folding begins with the formation of individual transient secondary structures and these structures are stabilized by packing. Further, folding is a hierarchical process in which simple structures are formed first from the local sequence and are marginally stable. The local structures interact to give rise to the intermediates with more complex structure and finally the native conformation (Baldwin and Rose, 1999a,b).

Hydrophobic Collapse model

According to this model a protein buries its hydrophobic side chains away from solvent water early during folding, forming a collapsed intermediate state from which the native structure develops by searching within the conformationally restricted states (Matheson and Scheraga, 1978; Dill, 1990). This idea originated from the fact that hydrophobic residues are generally buried inside the globular protein. According to this hypothesis hydrophobic collapse is the earliest event in protein folding. For the first time
using the protein barstar, it has been shown experimentally that a non-specific hydrophobic collapse occurs before the formation of the secondary or the tertiary structure (Agashe et al., 1995).

Nucleation-Condensation Growth model

According to this model a small unstable nucleus forms first during protein folding upon which the polypeptide chain can fold. The initial nucleation event is rate limiting and hence no intermediates are expected to be present further during folding and the folding is considered a two state process. Studies on the folding of chymotrypsin inhibitor 2 (CI2) has given strong evidence in support of this model (Wetlaufer, 1973; Itzhaki et al., 1995; Fersht, 1995, 1997; Neira et al., 1996).

Diffusion-Collision model (Karplus and Weaver, 1976, 1979, 1994)

The model was earlier proposed by Karplus and Weaver in 1976. According to this model, proteins are considered to be composed of several elementary microdomains like α-helices, β-strands etc. These microdomains move diffusively under the influence of internal and random external forces resulting in microdomain-microdomain collisions. A series of stepwise collisions lead to the formation of the native like conformation of the backbone. The final step of the folding process is the formation of the exact tertiary structure as a result of the close packing of the side chains.

Jig-saw puzzle model (Harrison and Durbin, 1985)

The main feature of this model is of multiple folding pathways. Each unfolded molecule follows a different path to reach the native conformation. This would make folding more robust to mutations that do not adversely affect the native structure. This model is analogous to a complex jig-saw puzzle where the pieces are joined together in a different order each time the puzzeld is assembled, but the end result is the same.
The new view of protein folding

One of the significant advances in our current understanding of protein folding is the emergence of a new view of protein folding (Baldwin 1994, 1995). The folding, according to this view, is considered as a parallel microscopic multipathway diffusion like process. The new view is based on broader notions like folding funnels and energy landscapes. The new view emphasizes on ensemble and multiple folding routes and lays less emphasis on specific structures and folding pathways (Dill and Chan, 1997).

Thermodynamic vs. Kinetic view of protein folding

The question whether the native state of a protein is at its global minimum of potential energy has been discussed for quite some time (Anfinsen, 1973). The thermodynamic view of protein folding was postulated by Anfinsen. According to this hypothesis, native conformations of proteins are at their global free energy minima, relative to all other states having identical bonded chemistry (Anfinsen, 1973). Anfinsen's hypothesis became widely accepted as the basis for protein folding and experimental proof came from small proteins whose folding and unfolding is reversible.

However, the native protein may not be at its global minimum. In the late 1960s Levinthal argued (Levinthal, 1968) that an unfolded polypeptide can adopt many conformations and the search for the native structure will require enormous time period (Later called as Levinthal's paradox). This can simply not be possible by a random searching of the native conformation. Levinthal concluded that proteins must fold by specific folding pathways and these pathways must necessarily lead to conformations which are lower in energy relative to other accessible states. Given the vast size of the conformational space, these low energy conformations may not be at global minima (Baker and Agard, 1994). The idea of folding pathways based on Levinthal's paradox necessitated the search for folding pathways and intermediates.
Folding intermediates

The involvement of folding intermediates during protein folding has been strongly argued by Kim and Baldwin, 1982 and their existence in small monomeric proteins has been experimentally confirmed (Kim and Baldwin, 1990). In the last two decades, a lot of effort has been made to identify and characterize folding intermediates in the folding pathway of proteins. Due to their transient nature it is very difficult to analyze the structure of folding intermediates. However, improvements in methodologies have made it possible to detect and characterize such intermediates.

Besides kinetic intermediates, there are several reports of equilibrium folding intermediates but it is not clear how these intermediates resemble the kinetic folding intermediates. The equilibrium intermediates of different proteins have common characteristics with native like secondary structure and compact molecular size but are devoid of specific tertiary structure. This state of the protein was termed “molten globule state” (Ohgushi and Wada, 1983; Kuwajima et al., 1985; Ikeguchi et al., 1986). They have shown by circular dichroism (CD) measurements that the equilibrium unfolding intermediate (the molten globule state) of α-lactalbumin is identical to its transient folding intermediates in kinetic refolding. Later, the molten globule state was proposed as a general intermediate of the protein folding process (Kuwajima 1989; Ptitsyn et al., 1990; Arai and Kuwajima, 2000).

The strongest evidence in favor of the folding intermediates has come from hydrogen exchange (H-X) pulse labeling measurements. Early structure in the folding of RNase A has been detected by this technique. The results indicate the presence of native like secondary structures in the early folding intermediate (Udgaonkar and Baldwin, 1988). Using the same technique, intermediates in the cytochrome-C (Cyt-C) folding have also been identified (Roder et al., 1988).
Early events in protein folding

The experimentally observed folding rates are very heterogeneous in nature. Some proteins are able to reach the native structure within milliseconds where as for others, it takes hours (Zettlmeissl et al., 1984). Early events during folding is traditionally monitored by spectroscopic changes in stopped-flow mixing of denatured proteins into the native buffer which allows to monitor the processes slower than one millisecond (Bieri and Kiefhaber, 1999). Using CD and fluorescence, Udgaonkar and coworkers showed that in the case of barstar, the polypeptide collapsed within 4 ms to a compact globule with a solvent accessible hydrophobic core but without secondary or tertiary structure. This was the first report demonstrating experimentally the hydrophobic collapse prior to secondary or tertiary structure formation (Agashe et al., 1995).

By using continuous flow devices the limit of detection of folding has been pushed to the 50-100 microseconds region (Chan et al., 1997; Takahashi et al., 1997; Shastry and Roder, 1998; Yeh and Rousseau, 2000; Kuwata et al., 2001). Photochemical triggering, pressure and temperature jump are other powerful methods to monitor the folding event in the nanosecond time scale (Jones et al., 1993; Callender et al., 1998; Jacob et al., 1999; Eaton et al., 2000). Fast folding kinetic studies have recently given new insights into the protein folding mechanism.

Late events in protein folding

The rate limiting step in the folding of a number of proteins involves the formation of the native conformation. Mutational analyses have shown that substantial tertiary structure exists in intermediates and transition states that lead directly to the native form. The hydrophobic core is significantly organized in the late intermediates which rearrange to achieve high packing densities seen in the native proteins (Matthews, 1991). One of the reasons for the slow process in protein folding is the isomerization of proline imide bonds (Brandts et al., 1975; Kiefhaber et al., 1990; Schultz et al., 1992; Nakano et al., 1993). Proline is unique compared to the other amino acids as it forms
imide bond rather than the amide bond. The X-Pro imide bond (X is any amino acid residue) is usually found to be in a trans conformation, whereas cis amide bonds have been observed rarely (Stewart et al., 1990). Out of the two isomers, trans conformation is thermodynamically favored over the cis in the denatured condition. The rates of cis-trans interconversion are slow in the folding reaction. The slow phase in protein folding is also due to the shuffling and formation of native disulfide bonds. Most of the secretory proteins are disulfide bonded (Fahey et al., 1977). These bonds are formed by covalent bonding between two cysteine residues. For smaller proteins with no disulfide bonds, the folding reaction is very fast, while for proteins which require disulfide bond formation, the refolding is slower (Saxena and Wetlaufer, 1970).

Most of our current understanding about protein folding comes from in-vitro and theoretical studies and little is known about the in-vivo folding aspects. In recent years, however, considerable progress has been made in understanding the structure and function of molecular chaperones which participate in the in-vivo protein folding reaction and prevent polypeptides from aggregation.

Molecular chaperones and in-vivo protein folding

Protein biosynthesis occurs on the ribosome and the synthesis proceeds from the N-terminal to the C-terminal. The nascent protein appears to fold cotranslationally, while the in-vitro refolding involves the rearrangement of segments of the full length protein (Fedorov and Baldwin, 1997; Hardesty et al., 1999). As the newly synthesized polypeptide emerges into the cellular environment, its contacts with the surrounding become important in deciding its fate. The newly synthesized polypeptide may be the target for interaction with other proteins that could favor aggregation. To overcome this problem, nature has evolved an efficient strategy involving molecular chaperones to prevent aggregation of proteins and in turn help them to fold to their native structure (Feldman and Frydman, 2000).
In vivo folding and biosynthesis are intimately coupled. As soon as enough amino acids protrude from the ribosome, they fold to the lowest available energy minimum for that length of the protein. For small, single domain proteins it can occur cotranslationally. For multidomain and multimeric proteins, exposure of hydrophobic residues may pose a major problem. Another aspect which affects the folding of proteins is its cytoplasmic environment. Cytoplasm is a crowded environment and in this environment the thermodynamic activity of partially folded states is increased 10-100 fold (Minton, 2000). This condition leads to the increase in the rate constant for folding which causes aggregation (van den Berg et al., 1999, 2000).

The problem of aggregation inside the cell is taken care of by molecular chaperones. They were originally discovered as factors for the formation of multisubunit assemblies such as bacteriophage capsids, chloroplast rubisco, RNA polymerase, and immunoglobulin (Seckler, 1997). Molecular chaperones belong to conserved protein families and have the ability to recognize and selectively bind non native proteins under physiological stress conditions. They, hence prevent irreversible aggregation reactions and keep the proteins on the correct folding pathway. Cells seem to have developed several functionally distinct chaperone families to support protein folding. The importance of molecular chaperones under stress conditions is highlighted by the finding that all the major heat shock family of proteins (Hsp104, Hsp90, Hsp70, Hsp60/GroEL and small heat shock proteins) suppress irreversible unfolding reaction (Buchner, 1996).

Hsp70 and GroE are well characterized molecular chaperone systems (Gething and Sambrook, 1992; Hartl, 1996; Bukau and Horwich, 1998; Grantcharova et al., 2001). Hsp70 has a molecular weight of 70kD and is found in prokaryotes and in most of the compartments of Eukaryotes. The basic function of this class of molecular chaperones is to bind and release hydrophobic segments of an unfolded polypeptide chain in an ATP-hydrolytic reaction cycle, wherein binding results in the stabilization of the unfolded state. This is followed by the controlled release of the folding polypeptide which allows progression along the folding pathway (Hartl, 1996; Feldman and Frydman, 2000).
The crystal structure as well as the cryo-electron microscope reconstruction of GroEL is available (Braig et al., 1994; Roseman et al., 1996; Xu et al., 1997). GroEL is made up of two stacked 7 membered rings of ~60kD. They cooperate with a smaller protein cofactor GroES which is a single heptameric ring of 10kD. GroEL is thought to promote productive folding through two actions. First, binding of nonnative polypeptide to the walls of GroEL central channel, mainly through hydrophobic contacts which leads to prevention of aggregation (Fenton et al., 1994) and partial unfolding of kinetically trapped intermediates, allowing them another chance for productive folding (Xu and Sigler, 1998). The second action is to facilitate folding after the polypeptide is released into an expanded central channel. Depending on the polypeptide, successful folding can require one or more rounds of binding and release (Fenton and Horwich, 1997). It has been reported that GroEL accelerates the millisecond folding kinetics of barstar but does not affect the slow refolding kinetics (Bhutani and Udgaonkar, 2000).

In vivo substrates of GroEL have been identified by mass spectrometry (Houry et al., 1999). Unlike the Hsp70 and GroE systems, small Hsp’s are less well known with the exception of α- crystallin (Buchner, 1996). The function of small Hsp’s do not seem to be regulated by ATP. From the in-vitro studies it has become clear that these proteins specifically bind non-native structures and prevent aggregation (Buchner et al., 1998a,b).

Subunit assembly

Nature has evolved multidomain and multimeric proteins in order

- To create function by the cooperation of two or more domains in the formation of one common active center.
- To become multifunctional, thus allowing regulated networks of processes to develop.
- To enhance the rate of self-organization by synchronous nucleation at multiple sites.
- To improve the stability by protecting the folding polypeptide chain in its substructures from cotranslational degradation (Jaenicke, 1999).
The probability of a protein to occur as multiple subunits increases with increasing molecular mass. Quaternary structure of a protein may be made up of homooligomers (e.g., citrate synthase, chloramphenicol acyl transferase, lactate dehydrogenase, P22 Tailspike protein, and GroEL proteins) or heterooligomers (e.g., lactose synthase, luciferase, tryptophan synthase, etc.). Subunit association involves a number of noncovalent interactions, e.g., hydrogen bonds, van der Waals, electrostatic, and hydrophobic interactions (Dill, 1990; Jaenicke, 1991) which give rise to the quaternary structure of a protein. The basic steps in the subunit assembly involve the formation of super secondary structures which collapse into sub domains and domains to give rise to monomers which ultimately form the oligomeric structure. In most of the cases formation of the functional state requires subunit assembly which in the case of enzymes can be monitored in-vitro by activity measurements. Preceding steps can be monitored by a variety of techniques like spectral analysis, cross linking, HPLC, etc. (Jaenicke, 1995; Seckler, 2000).

Salmonella Phage P22 Tailspike protein (TSP) is one of the most widely studied protein from the point of view of folding and subunit assembly. The folding and assembly pathway of this protein has been elucidated both in vivo and in vitro. The monomer folds via an intermediate and then forms a loosely packed structure called “protrimer” which finally gives rise to the native trimer. Another protein which has also been studied in great detail is bacterial luciferase. Luciferase assembly is a multistep process comprising intermediates both preceding and following the association steps (Seckler 1997, 1998).

While there has been considerable interest in understanding protein folding pathways and the assembly of proteins, another side of the folding reaction, viz., aggregation has been often ignored. The involvement of protein aggregation in proteins produced by recombinant DNA technology at high concentrations on one hand and its involvements in a number of diseases on the other, has drawn a considerable amount of attention to study their formation and prevention. (Jaenicke, 1987; Thatcher and Hitchcock, 1994; Jaenicke and Seckler, 1997; Jaenicke and Lilie, 2000).
Protein aggregation

The phenomenon of protein aggregation has often been observed during the in-vitro folding of multimeric and multidomain proteins. In addition to the folding problem, which is yet to be completely understood, the aggregation process during folding has started catching a great deal of attention. Understanding aggregate formation and its prevention, both in-vivo and in-vitro is a serious challenge for the scientific community.

Mechanism of aggregation

One of the earliest systematic studies on aggregation goes back to 1974 in which the folding of enzyme tryptophanase revealed an intermediate at moderate denaturant concentrations that was responsible for aggregate formation (London et al., 1974). To map the step at which aggregation is committed, Goldberg and coworkers showed that early folding intermediates are responsible for aggregation in the case of lysozyme (Goldberg et al., 1991). The aggregate formation in P22 tailspike protein has been studied in great detail (Haase-Pettingell and King, 1988; Mitraki et al. 1991; Betts et al., 1997; Jaenicke and Seckler, 1997). Contrary to Goldberg’s observation, using P22 Tailspike protein and coat protein, King and coworkers showed that aggregation is a specific phenomenon (Speed et al., 1996).

Fink has proposed a model for aggregate formation. The basic assumption in this model is that protein folding involves intermediates, each consisting of structural building blocks. In the case of normal folding, hydrophobic surfaces of the structural units interact in an intramolecular manner to form the native conformation, while for aggregation the hydrophobic structures of the building blocks interact in an intermolecular fashion (Fink, 1998).

Structure of aggregates

The structure of protein aggregates has been determined by using attenuated total reflectance (ATR) FTIR. It has been observed that inclusion bodies and folding aggregates exhibit substantial secondary structure. Structure of a given protein in
inclusion bodies formed after refolding from a chemically denatured state or high temperature denatured state are the same (Oberg et al., 1994). An important observation regarding the formation of aggregates is that they have a tendency to have more β sheet content compared to their native state. This is a result of intermolecular interactions leading to beta sheet like interactions (Fink, 1998). The secondary structure of β lactamase (an α and β protein) inclusion bodies in *E. Coli* investigated using Raman spectroscopy, indicated that the amide I bond spectra of inclusion bodies were different from the native protein. There was an increase in the beta sheet content and a decrease in the alpha helical content (Przybycien et al., 1994). It has been observed for P22 Tailspike protein (Sargent et al., 1988) and interleukin-1 β (Oberg et al., 1994) that secondary structure of the aggregates were found to be very similar to those of the native proteins.

**Kinetic competition between folding and aggregation**

Rate of aggregate formation increases upon increasing the initial concentration of the unfolded protein. Because of the kinetic competition, the yield of correctly folded protein decreases when refolding is carried out at increasing protein concentration (Zettelmissee et al., 1979; Rudolph and Lilie, 1996; DeBernardez Clark et al., 1998; DeBernardez Clark et al., 1999). A kinetic competition between correct intramolecular and incorrect intermolecular interactions during refolding is considered to be responsible for the decrease in the yield of the renatured protein. In the case of aggregate formation, second order (or higher) reactions compete with first order folding. With the increasing concentration of the denatured protein, the rate of refolding remains constant, whereas the rate of aggregation increases as a consequence of which aggregation predominates at high denatured protein concentrations (Kiefhaber et al., 1991).

**Inclusion bodies**

Expression of genetically engineered proteins in bacteria often results in the accumulation of proteins as inactive and insoluble aggregates inside the cell called as inclusion bodies (DeBernardez Clark and Georgiou, 1991, Thatcher and Hitchcock, 1994;
DeBernardez Clark, 1998, 2001; DeBernardez Clark et al., 1999). Though inclusion body formation has mostly been observed in *E. Coli* and *S. typhimurium*, its occurrence has also been reported in algal chloroplast (Ketchner et al., 1995). The formation of inclusion bodies is often frustrating for researchers and a variety of methods have been used to isolate and solubilize them to generate the native protein. Expression of proteins as inclusion bodies is often advantageous as it reduces the number of steps for protein purification. The highly concentrated protein is also protected from proteolytic degradation. The inclusion body formation is also the best method to express the protein if it is toxic to the host cell (DeBernardez Clark and Georgiou 1991; Rudolph and Lilie, 1996).

Inclusion bodies are the most dense structures within the microbial cells. They are 0.2-0.5 μ in diameter, not bound by a membrane, and are morphologically dissimilar to any other intracellular structure found in prokaryotes (Thatcher and Hitchcock, 1994). There are many factors that influence the inclusion body formation including very high local concentration of the protein, reducing environment in the cytoplasm, lack of posttranslational modifications, improper interactions with chaperones and intermolecular crosslinking via disulfide bonds. Temperature also plays an important role in inclusion body formation.

Isolation and solubilization of inclusion bodies

Inclusion bodies are purified from cells by disrupting the cell wall using high pressure homogenization or sonication and the resulting suspension is centrifuged to remove the soluble fraction. The inclusion body containing pellet is usually washed with low concentration of urea, guanidine chloride or detergents such as Triton X-100 and sodium deoxycholate (Rudolph et al., 1997a). Solubilization of inclusion bodies is done by using high concentration of strong denaturants, most notably 8 M Urea or 6 M GdmCl, thiocyanate salts or SDS and a reducing agent like β-mercaptoethanol, DTT, DTE or cysteine. Normally a temperature above 30°C is used to facilitate the solubilization process (DeBernardez Clark, 1998).
Refolding of solubilized inclusion body is performed either by dilution or dialysis. During the removal of denaturant by dialysis, the protein is exposed to intermediate denaturant concentrations for an exposed period of time. This could be beneficial or detrimental for correct folding depending on the nature of the protein. Folding intermediates populated at intermediate denaturant concentration are sometimes highly susceptible to aggregation (Schumann et al., 1993a). In these cases the denaturant should be removed by rapid dilution into the refolding buffer. The preference of a given protein for refolding by dilution or dialysis cannot be deduced from the molecular properties of the protein (Rudolph and Lilie, 1996). To avoid the problem of aggregation and to improve refolding yield, various strategies have been applied.

Improving the refolding yield

*Preventing aggregation by pulse renaturation*

At high protein concentrations, refolding yield of multidomain and multimeric proteins decreases. Renaturation of these proteins would require very large renaturation buffer volumes. High concentration of correctly folded proteins can be obtained by the successive addition of denatured proteins to the renaturation solution (Rudolph and Fischer, 1990). By stepwise addition of the denatured protein to the refolding vessel, the concentration of the unfolded polypeptide is kept low throughout the refolding process. Care must be taken to give enough time in early stages of the folding pathway where folding molecules are quite susceptible to aggregation (Fischer et al., 1992). While optimizing the refolding conditions it is important to monitor the folding kinetics and to determine the critical concentration above which the refolding yield decreases. Since the residual denaturant concentration affects folding and activity, a concentration which does not interfere should be taken. (DeBernardez Clark et al., 1999).
Matrix-assisted prevention of protein aggregation

Creighton first demonstrated that refolding can be improved by the noncovalent binding of the unfolded polypeptide to an ion exchange matrix (Creighton, 1990). Transient binding of folding molecules to a matrix reduces misfolding and aggregation. Histidine tag (His-tag) or polyionic fusion constructs have been successfully used to improve the refolding yield of proteins. Aggregation can be completely prevented by this specific noncovalent immobilization of the folding polypeptides via the affinity tags at N or C terminus (Holzinger et al., 1996; Stempfer et al., 1996; Ishibashi et al., 2001; Rumlova et al., 2001).

Protein engineering approach for prevention of aggregation

Hydrophobic patches on protein surfaces are one of the main source of aggregate formation. Aggregate formation can be prevented by altering the amino acids which form such hydrophobic patches. From in-vivo and in-vitro experiments it has been shown that mutations in turns and surface hydrophobic patches of a protein result in decreased aggregation (Knappik and Plückthun, 1995; Nieba and Plückthun; 1997). Replacement of Asn to Lys led to a decrease in the aggregation of erythropoietin derived from E.coli (Narhi et al., 2001).

In vitro folding of artificial fusion proteins

Hydrophobic patches on protein surface are known to be involved in the aggregate formation in proteins (Fink, 1998). One approach to reduce aggregation and increase the refolding yield is to make fusion constructs with hydrophilic proteins or peptides. The two proteins or domains serve as fusion partners and do not interfere with the refolding of the target protein. Refolding of recombinant insulin like growth factor1 (IGF-1) was carried out by using IgG binding domain from staphylococcal protein A as a fusion partner (Samuelsson et al., 1991). A similar strategy using small hydrophilic peptides as fusion partners has been successful in the oxidative folding of human granulocyte colony-stimulating factor (Ambrosius et al., 1996).
**Preventing aggregate formation by antibodies**

Monoclonal antibodies specific to hydrophobic patches can be used to prevent hydrophobic interactions among unfolded polypeptides and in turn in the prevention of aggregation. This strategy has been used in carboxypeptidase A (Katzav-Gozansky et al., 1996). Antibodies may also affect folding by stabilizing the native state of inherently unstable proteins. Reduced S-protein (a proteolytic fragment of ribonuclease A) which is difficult to refold because of low stability, could be renatured in the presence of monoclonal antibodies (Carlson and Yarmush, 1992)

**Cosolvent assisted protein refolding**

Cosolvents like glycerol and sucrose have long been used to stabilize the activity of enzymes and maintain the functional state of isolated cellular organelles under various forms of stress (Timasheff, 1998). Timasheff and coworkers in a series of elegant observations proposed the mechanism of stabilization of proteins by cosolvent additives (Gekko and Timasheff, 1981a,b; Arakawa and Timasheff, 1982, 1983, 1984, 1985; Arakawa et al., 1990; Bhat and Timasheff, 1992; Timasheff, 1993; Timasheff, 1998). Depending on their nature and the nature of proteins they either preferentially hydrate or preferentially bind to the proteins. Polyols and sugars are known to preferentially hydrate the proteins and in turn stabilize them (Kaushik and Bhat, 1998; Tiwari, 1999), while compounds like PEGs can preferentially bind to proteins at higher temperatures (Bhat and Timasheff, 1992). Various other groups of compounds like carboxylic salts (von Hippel and Schleich, 1969; Kaushik and Bhat. 1999; Tiwari, 1999), amino acids (Arakawa and Timasheff, 1983; Taneja and Ahmad; 1994; Rishi et al., 1998; Anjum et al., 2000) also preferentially hydrate the proteins and in turn stabilize them. Since stabilizing cosolvents are known to push the native state (N)\(\leftrightarrow\) denatured state (D) equilibrium in favor of the native state and folding and stability are interrelated phenomena, the use of cosolvents is catching great amount of attention for the in-vitro as well as the in-vivo folding of proteins.

The low in-vitro refolding yield of non disulfide bonded or disulfide intact proteins is mainly due to aggregate formation. A very efficient and cost effective strategy
to suppress aggregation is the inhibition of intermolecular hydrophobic interactions by using small additives (DeBernardez Clark, 2001). It is easy to remove cosolvent additives once the refolding is complete. However, the mechanism of action of additives on the refolding process is not very clear. They may influence the solubility and the stability of the native, denatured, and intermediate states, or they may alter the kinetics of refolding. Kinetics of refolding can also be altered by changing the physical properties like viscosity of the refolding buffer or slow or fast removal of the denaturant either by dialysis or dilution. Since exact mechanism of their action is not known, it is very difficult to predict the behavior of additives in the refolding of different proteins. Cosolvent additives that work well for a certain protein have proven to be disadvantageous for another protein.

Several success stories of the use of this approach have been reported in the literature. Labilizing agents like L-Arginine lead to substantial increase in the renaturation yield of recombinant human tissue-type plasminogen activator (Rudolph, 1990), antibody fragments (Buchner and Rudolph 1991), casein kinase (Lin and Traugh, 1993) and recombinant human interferon γ (Arora and Khanna, 1996). Although Arg.HCl has a guanidino group, L-arginine exerts minor effect on protein stability compared to GdmCl. A slight destabilizing effect of Arg.HCl has been reported in the case of cytochrome C (Taneja and Ahmad, 1994). Despite the successful use of L-Arginine, the mechanism of its action as a very efficient low molecular weight folding enhancer remains enigmatic. Polyethylene glycols (PEGs) have been successfully used to increase the refolding yield of carbonic anhydrase B (Cleland and Wang, 1990a,b; Wetlaufer and Xie, 1995), lysozyme (Moorthi, 1997), fusion protein of human growth hormone and glutathione S transferase (Kim and Lee, 2000), recombinant human tissue transglutaminase (Ambrus and Fesus, 2001). In the case of carbonic anhydrase B it has been shown that PEGs inhibit aggregation by complex formation with a molten globule like folding intermediate of the protein which is prone to aggregation. Low concentration of denaturants (GdmCl/Urea) have also been found to improve the refolding yield of α-chymotrypsinoogen (Orsini and Goldberg, 1978; Vani, 1996). Improving the refolding yield by the addition of nondenaturing concentrations of denaturant is, however, only possible if the native state is not significantly destabilized under such solvent conditions.
Lauryl maltoside, a detergent has been shown to increase the refolding yield of membrane bound protein rhodanese (Tandon and Horowitz, 1987). Another detergent CHAPS has helped in increasing the refolding yield of carbonic anhydrase B (Wetlaufer and Xie, 1995). Non-detergent sulphobetaines are known to enhance the refolding yields of unfolded hen egg white lysozyme and β- galactosidase manifold (Goldberg et al., 1996). Detergents and surfactants improve the refolding yield by binding to the folding intermediates which in turn prevents aggregation. Release of the detergent molecules from the folding intermediates can be facilitated by carefully selecting weakly binding detergents. The concentration of detergents needed varies for achieving the best refolding yield under particular conditions. Some are effective above their critical micelle concentration (CMC) and some below. The drawback in using detergents is that they are difficult to remove and may affect downstream chromatographic steps. In the micelle assisted refolding, Triton X-100 or lauryl maltoside used forms mixed micelles. These mixed micelles contain both polar and nonpolar moieties that can interact with various sites on the folding intermediates thereby preventing their aggregation (Zardeneta and Horowitz, 1994). Addition of haem and calcium improved the refolding of horseradish peroxidase (Smith et al., 1990). Proline has been shown to prevent aggregation of lysozyme by binding to the folding intermediates and trapping them into enzymatically inactive aggregation insensitive state (Samuel et al., 2000). Further, denatured and reduced lysozyme has been refolded at 1.6 mg/ml concentration in ethylammonium nitrate (EAN). EAN has been shown to prevent aggregation by interacting with hydrophobic groups and protecting the protein from intermolecular association (Summers and Flowers, 2000). The refolding yield of carbonic anhydrase II has been increased by using alcohols (Wetlaufer and Xie, 1995). Methylurea, ethylurea, formamide, methylformamide and acetamide in the 1.5-2.5 M concentration range have been used to enhance the protein refolding yield (Rudolph et al.,1997b). Refolding can be also increased by using cyclodextrin in the refolding buffer (Karupiah and Sharma, 1995). Glycerol has been used to improve the refolding yield of rhodanese (Gorovits et al., 1998) while addition of sucrose in the refolding buffer increased the refolding yield of β-lactamase (Valax and Georgiou, 1991).
A so called ‘artificial chaperone’ strategy has been employed to increase the refolding yield of citrate synthase, lysozyme and carbonic anhydrase B (Rozema and Gellman 1995, 1996a,b; Daugherty et al., 1998) in which a two step dilution strategy has been used. In the first step, the detergent is added which binds to the protein and in turn prevents intermediates from sticking to each other. In the second step, cyclodextrin is added to remove the detergent. Recently, a similar strategy has been applied for citrate synthase refolding using polyoxyethylenesorbitan followed by the addition of cycloamylose (Machida et al., 2000).

It has now been clearly demonstrated that a variety of cosolvent additives can efficiently help in the refolding of proteins by preventing their aggregation. These cosolvents are of varied nature and not only stabilizers but also denaturants at low concentrations have helped in the refolding of proteins. This suggests that their mechanism of action could be quite different from each other. It has also been observed that certain compounds which help a protein fold better, may not help another at all (De Bernardez Clark et al., 1999). Despite several successes of the use of cosolvent additives in the refolding of proteins, no systematic studies on understanding their effect has been carried out and their mechanism of action is far from clear.

Protein misfolding, aggregation and disease

Proteins synthesized inside the cell are taken care of for proper folding. Failing to do so leads to their degradation. Despite these quality control measures, sometimes they get misfolded or form aggregates in the cell. Deposition of aggregates or the misfolded structures causes disturbance in the cellular physiology which in turn leads to a number of human and animal diseases such as Alzheimer’s disease, Cruetzfeldt-Jakob disease (CJD), bovine spongiform encephalopathy (BSE), cystic fibrosis, etc. (Thomas et al., 1995; Dobson, 1999; Morello et al., 2000; Soto, 2001). Pathologically diseases are classified as inflammatory, degenerative, infectious, neoplastic and conformational diseases. Diseases, which arise due to a change in the conformation of proteins, are called as conformational diseases (Carrell and Lomas, 1997).
Proteinaceous aggregates that form deposits and lead to malfunctioning of the cell are collectively termed as amyloid. Formation of β aggregates is supposed to be the main cause of Alzheimer's disease (Kisilevsky, 1998). The mechanism of conversion of normal cellular proteins into amyloid fibrils is not well known. However in-vitro conditions have been established where normal proteins can form amyloid structures. Formation of amyloid is not only restricted to certain proteins which can cause disease but occurs in other proteins as well (Guijarro et al., 1998; Schuler et al., 1999; Zurdo et al., 2001) having no direct connection to diseases. Studies on the amyloid formation also indicate that besides the amino acid sequence, the environmental factors such as pH, temperature, ionic strength and cosolvents can affect the polypeptide chain conformation (Kelly, 1998).

Prions are the most extensively studied group of proteins which cause conformational diseases (Prusiner, 1997). This is a group of diseases which includes CJD in humans, scrapie in sheep and BSE or mad cow disease. Normal cellular prion protein (PrPc) differs from its scrapie prion protein from (PrPsc) in their secondary and tertiary structure. The biological role of prion proteins is not known, but their expression is highest in neuronal cells (Daggett, 1998). It is a sialolglycoprotein bound to the plasma membrane. PrPc forms amyloid plaques in the brains of humans and animals leading to prion diseases. PrPc is mainly helical in nature (42%) with very little β sheet (3%) structure, while PrPsc contains a large amount of β sheet (43%) and less helical structure (30%) (Caughey et al., 1991; Pan et al., 1992; Safar et al., 1993). NMR structure of murine prion protein fragment (121-231) has been elucidated (Riek et al., 1996). The function of PrPc is not known but it is believed that PrPc binds copper and is involved in copper metabolism (Homshaw et al., 1995). From in-vivo experiments it has been observed that wild type mouse brain has a much higher level of membrane associated copper compared to PrP deficient mouse, which has correspondingly higher concentration of serum copper. From these studies it is concluded that PrPc is involved in the regulation of copper levels as a sieve or shuttle to intracellular compartments (Brown et al., 1997).
Prevention of misfolding in relation to disease

To control the amyloid pathway and to correct the folding defects, organic cosolvents (Chemical chaperones) have been used. Cosolvents are known to stabilize proteins (Gekko and Timasheff, 1981a,b; Schein, 1990) and this property has been used in the prevention of PrP\textsuperscript{Sc} formation. Glycerol, TMAO, and Dimethyl sulfoxide (DMSO) have been used to prevent conformational changes in the prion protein (Tatzelt et al., 1996). The structural information of PrP\textsuperscript{Sc} has helped in developing therapeutic strategies for the prion disease. One of the successes came from Kelly’s group involved in preventing transthyretin amyloid fibril formation, applicable to the prion disease as well (Miroy et al., 1996). In this strategy they designed a high affinity ligand for the folded form, making it difficult for the folded form to convert into the amyloid form. In another approach Phenylalanine rich peptidomimetics have been designed for Alzheimer’s disease, as they have been shown to be effective in preventing the assembly of β peptide into amyloid fibrils. Cystic fibrosis is caused by the deletion of Phenylalanine (ΔF508) which interferes with the folding of newly synthesized cystic fibrosis transmembrane conductance regulator (CFTR) polypeptide. It has been demonstrated that ΔF mutation does not affect its activity, thereby raising the possibility of therapeutic intervention by increasing the ΔF508 folding. Treatment of ΔF508 expressing cells with glycerol dramatically stabilized newly synthesized ΔF508 polypeptides and led to the accumulation of stable functional CFTR chloride channels in the plasma membrane (Sato et al., 1996). Glycerol has also been used in correcting trafficking and folding defects in aquaporin-2 water channels in nephrogenic diabetes insipidus (Tamarappoo et al., 1999). Certain forms of liver disease and emphysema have been linked to mutations that lead to the retention of misfolded proteins in the endoplasmic reticulum. One such example is D342K, which is the most common mutation found in the gene encoding α1-antitrypsin (α1-AT), referred to as the Z deficiency variant (α1-ATZ). Normally α1-AT is secreted into blood stream and body fluids. D342K mutation leads to lung emphysema in patients and its marked accumulation in the liver is responsible for the liver injury. In cells synthesizing α1-ATZ, treatment with glycerol produces a fivefold increase in the secreted
levels of the mutant enzyme (Carell and Lomas, 1997; Perlmutter, 1999; Burrows et al., 2000). Ligands with pharmacological selectivity can rescue the proper targeting and function of misfolded proteins including receptors and are called as pharmacological chaperones. These pharmacological chaperones might help to develop new treatments for conformational diseases (Morello et al., 2000).

Cosolvents play an important role in preventing aggregation in-vitro as well as in-vivo. However, very little is known about their mechanism of action. In the present study Citrate Synthase, which is highly susceptible to aggregation during in-vitro refolding, and P22 Tailspike protein (TSP), one of the few proteins whose folding pathway is well characterized both in-vivo and in-vitro, have been used as model proteins and the effect of various cosolvent conditions on the refolding process has been studied in order to understand their mechanism of action.

Citrate Synthase (CS): A highly aggregation prone protein

Citrate Synthase (EC 4.1.3.7, citrate oxalacetate-lyase also known as condensing enzyme, CS) catalyzes the first reaction of tricarboxylic acid (TCA) cycle and thus plays a prominent role in cellular metabolism. In the presence of CS, acetyl-CoA and oxalacetate react to form citric acid and Co-A. This reaction is unique among those of the TCA cycle as it involves a new C-C bond formation. It is primarily a mitochondrial enzyme and located in the mitochondrial matrix. In eukaryotes, citrate synthase is coded by nuclear DNA and is synthesized in the cytosol as a precursor which is subsequently translocated into the mitochondria. It is a regulatory enzyme and its enzymatic activity is inhibited by ATP and long chain fatty acyl CoAs. The inhibitory studies result in a sigmoidal curve suggesting a cooperative interaction of these two effectors with the enzyme. Inspired by these observations, Wu and Yang made an attempt for physicochemical characterization of citrate synthase and its subunits (Wu and Yang, 1970). Molecular weight of CS is 1x10^5 Da as determined by sedimentation equilibria and a combination of hydrodynamic properties. Native citrate synthase is made up of two subunits which are similar (Srere, 1966; Wu and Yang, 1970; Singh et al., 1970; McEvily
and Harrison, 1986). CS is dissociable in 7.0 M GdmCl (Singh et al., 1970) or by extensive modification of lysine residues with succinic anhydride (Wu and Yang, 1970). The enzyme undergoes dimer-monomer equilibrium at low protein concentration, low pH and low ionic strength. Dimerization is favored in the presence of oxalacetate and citrate (McEvilly and Harrison, 1986). The amino acid sequence of porcine heart citrate synthase is known (Bloxham et al., 1982) and the three dimensional structure of the enzyme with and without substrate ligands has also been determined (Remington et al., 1982; Wiegand et al., 1984; Karpusas et al., 1990). Porcine heart citrate synthase is a dimer of identical subunits with 437 amino acid residues in each subunit. It is primarily a helical protein with 40 helices per dimer, packing tightly to form a globular molecule. Each subunit is clearly divided into a large and a small domain. The subunits are extensively interdigitated with one subunit making significant contribution to both the citrate and the CoA binding sites of the other subunit. The two subunits associate tightly to form the dimer and about 28% of the surface of the monomer becomes inaccessible to solvent upon dimerization. The forces which mainly contribute to the stability of subunits are hydrogen bonds and salt bridges. Upon substrate binding, the enzyme undergoes conformational changes (Remington and Huber, 1982). His 247 and Asp 375, two of the many residues which are conserved in citrate synthase, interact directly with the substrate molecule and have been implicated in the catalytic mechanism of the enzyme (Karpusas et al., 1990).

Citrate synthase is highly prone to aggregation and some early studies revealed very little reactivation of the enzyme upon GdmCl denaturation (West et al., 1990). This property of CS makes it one of the most popular model system for refolding studies and for studying the role of molecular chaperones in their folding (Buchner et al., 1991; Zhi et al., 1992). It has also been used as a model system for artificial chaperone assisted refolding of the protein (Rozema and Gellman, 1995; Daugherty et al., 1998). GroEL and GroES proteins belonging to the heat shock family of proteins have been used to increase the refolding yield of proteins as well as to understand the mechanism of action of molecular chaperones. In the case of CS, the refolding polypeptide quite often aggregates rapidly as shown by light scattering which is attributed to low refolding yield. Its
refolding is known to be concentration and temperature dependent. GroEL and GroES are known to increase CS refolding in a stoichiometric amount (Buchner et al., 1991). Srere and coworkers studied the influence of denaturants and folding assistants on CS refolding. Bovine serum albumin, oxalacetate, and glycerol have been used as folding assistants. Up to 82% refolding has been achieved with GroEL, GroES and ATP (Zhi et al., 1992). Role of GroES in the permissive and non permissive conditions has been studied by Schmidt et al. (1994). They have shown that under non permissive conditions, where spontaneous refolding could not occur, reactivation of the native state required the complete chaperonin system. Under permissive conditions, GroES was no longer mandatory.

Another heat shock protein Hsp90, abundant in the cytosol of eukaryotic cells, has been tested as a molecular chaperone using CS as a model protein. 15% refolding was observed in spontaneous refolding while up to 40% refolding yield was observed with the help of Hsp90. The action of Hsp90 does not depend on the presence of nucleotide triphosphate (Wiech et al., 1992). Role of N-terminal and C-terminal domains of Hsp90 was also probed by using CS as a model system. During refolding, both N and C terminal fragments have shown aggregation inhibition property. The chaperone function of Hsp90-associated proteins namely FKBP52 and p23 has also been tested by using CS a model protein. These components specifically suppress the aggregation of CS in a concentration dependent manner similar to that of small heat shock proteins and Hsp90 (Bose et al., 1996).

The molecular chaperone property of small heat shock proteins murine Hsp25, human Hsp27 and bovine αβ crystallin has been studied for the unfolding and refolding of citrate synthase. All these small Hsp's act as molecular chaperones in a stoichiometric amount and their interaction with the unfolded and the refolded protein seems to be ATP independent (Jakob et al., 1993). Small heat shock proteins from pea Hsp 18.1 and Hsp 17.7 were also shown to enhance the refolding of chemically denatured citrate synthase. They have also been shown to prevent the aggregation at 45°C and irreversible
inactivation at 38°C (Lee et al., 1995). Stable binding of several non-native CS molecules to one Hsp 25 oligomer led to the accumulation of a CS unfolding intermediate which was protected from irreversible aggregation (Ehrnsperger et al., 1997). \( \alpha \beta \) crystallin, a member of small heat shock protein family, has been shown to actively participate in the refolding of citrate synthase. Its molecular chaperonin property works in an ATP dependent manner (Muchowski and Clark, 1998).

Chaperone activity of the proteasome was tested by using CS as a model system. It works in an ATP dependent manner and the regulatory particle of the proteasome is involved in suppressing the aggregation of CS (Braun et al., 1999). The 26S proteasome is a large protease complex that catalyzes degradation of both the native and the misfolded protein. PA700, the regulatory subcomplex of the 26S proteasome, has also been shown to prevent CS aggregation and promote the refolding of citrate synthase (Strickland et al., 2000).

Using artificial chaperone strategy, the refolding of CS was attempted wherein sequential use of detergents and cyclodextrin was carried out. Up to 65% refolding yield was obtained by this method. In the first step, unfolded CS was treated with detergents wherein hydrophobic patches of the unfolded polypeptides get shielded by the detergent. In the next step, cyclodextrin was added which initiates refolding by stripping the detergent from the protein-detergent complex. (Rozema and Gellman, 1995; Daugherty et al., 1998). Recently a similar strategy has also been applied in the refolding of CS by using cycloamylose (Machida et al., 2000).

P22 Tailspike Protein (TSP): A model system for in-vivo and in-vitro folding

The Phage P22 attaches to its host *Salmonella typhimurium* through elongated structures called as Tailspike protein (TSP). It is the product of gene 9 of phage and after synthesis it forms the infectious appendage to the phage (Israel, 1978; Sauer et al., 1982). TSP recognizes, binds and cleaves the O-antigen lipopolysaccharide projecting from the
outer membrane of the bacteria (Eriksson and Lindberg, 1977; Baxa et al., 1996; Baxa, 1998). The X-ray structure of P22 TSP has been solved by fragmentation approach. Two fragments, C-terminal (109-666) and N-terminal (1-124) were expressed, purified and crystallized (Miller, 1995; Miller et al., 1998). The structures were determined at 1.5 Å and 2.3 Å resolutions, respectively (Steinbacher et al., 1994, 1997). A Characteristic feature of P22 TSP is the presence of a parallel β-helix of 13 complete turns. First time such a structure was reported in bacterial pectate lyase C (Yoder et al., 1993). The cross section of TSP β-helix is kidney shaped but unlike other parallel β-helix structures the interior of this protein is hydrophobic. The P22 TSP structure has also been solved at 1.8 Å in complex with O-antigen oligosaccharides (Steinbacher et al., 1996).

After synthesis on ribosome, tailspike protein folds to a monomeric species. These species associate into protomer but are not fully folded. In the last step, protomer folds further to yield the thermally resistant native TSP (Goldenberg et al., 1982; Goldenberg and King 1982; Danner and Seckler, 1993; Seckler, 1997). The native TSP trimer is highly thermostable and resistant to denaturation by SDS at room temperature, while its folding intermediates or aggregated polypeptides are denatured by the ionic detergent. Hence, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) can be used to follow the native structure formation.

A large number of sites are known to be the target of point mutation in the tailspike gene. Due to clear differences between the physical properties of the native trimer and the intermediates, a class of mutations called temperature sensitive to folding (tsf) has been identified. tsf mutations alter the folding pathway without influencing the native protein (Smith et al., 1980). A characteristic feature of tsf folding mutations is that they acquire native structure at permissive temperatures but fail to do so at non permissive temperatures (Smith and King, 1981). A second kind of mutations which alleviate the folding defects caused by tsf mutation (Fane and King 1991) have been identified and called the suppressor (su) mutation. They have wild type characteristics and increase the refolding yield of the native TSP at higher temperatures. At low
temperatures, folding and assembly are identical for the wild type and all the mutant proteins.

In a series of exciting experiments, Seckler and coworkers dissected the unfolding and refolding pathway of P22 TSP in-vitro (Seckler, 1997, 1998, 2000). P22 completely unfolds at acidic pH (pH 3) and in 5 M urea (Seckler et al., 1989). After refolding, native trimer was formed with a high refolding yield (80%) at 10°C. The in vitro folding pathway of P22 TSP has been delineated by spectroscopic and hydrodynamic methods (Fuchs et al., 1991). The folding of P22 TSP proceeds through a monomeric folding intermediate containing substantial secondary structure shielding its aromatic side chains from the solvent. This monomeric intermediate which develops further structure and forms a folded monomer as observed by fluorescence spectra and circular dichroism studies indicates that it contains native like secondary and tertiary structure (Danner and Seckler, 1993; Miller et al., 1998). The assembly of individual subunits leads to a loosely folded protrimer in which the polypeptides are loosely associated and do not change subunits. Spectroscopic data and mutagenesis studies showed that the trimer maturation involves significant structural rearrangements in the carboxy terminal part of the protein as suggested by the crystal structure. The detailed mechanism of P22 TSP folding mutants has been characterized by Danner and Seckler (1993). The folding kinetics and yields of tsf and su mutants were compared with respect to the wild type. At 10°C, subunit folding rates were similar for wild type, tsf and su mutants. At 25°C, tsf mutant showed a decrease in the rate of subunit folding and su mutant an increase in the rate when present along with tsf mutant background. These data indicate that tsf mutations are destabilizing while the su mutations are stabilizing the folding intermediates. Stability of folding intermediates were assessed by temperature upshift and downshift experiments (Danner and Seckler, 1993; Beissinger et al., 1995).

N terminal deleted protein (ΔN P22 TSP) has been produced by proteolytic cleavage as well as by recombinant expression of the gene (Danner et al, 1993; Miller et al., 1998). The spectral and oligosaccharide binding properties of ΔN P22 TSP were
found to be similar to the full length protein. It is also SDS resistant and highly thermostable. It represents simpler version of the full length TSP and the availability of structure makes it a suitable model for folding and assembly studies. Folding and assembly pathway of the shortened protein is a mirror image of the full length protein (Miller, 1995; Miller et al., 1998).

Outline of the Present Research Problem

Citrate synthase and bacteriophage P22 Tailspike Protein are both highly prone to aggregation during refolding depending on their concentration and temperature. In the present study, cosolvents like polyols, sugars and polyethylene glycols have been used in assisting the refolding of citrate synthase. Polyols and sugars are known to increase the thermal stability of proteins by favoring the native state over the denatured state. It is, therefore expected that they should also favor the native state formation in the process of refolding, since molecular interactions involved in the stabilization of the native state are the same for the folding process as well. On the other hand, polyethylene glycols are known to prevent aggregation of proteins during refolding by virtue of binding to the exposed hydrophobic surfaces and their exclusion from the native state further facilitates the folding of proteins.

Although, there are several reports in the literature where the cosolvents used in the present study have been employed to enhance the refolding yield of proteins during in-vitro refolding, yet the mechanism of their action is far from clear. In the present study the focus has been to first check the ability of these compounds in improving the refolding yield of CS followed by understanding their mechanism of action. Among the polyols used, glycerol has been used successfully in the refolding of proteins and owing to its high solubility in water, can be explored for a wide range of its concentrations. Since among the polyols glycerol led to the highest increase in the refolding of CS, extensive studies on understanding the mechanism of action of glycerol were undertaken in particular. Glycerol effect was further studied extensively for the folding of the trimeric protein P22 TSP. The advantage of using this protein is that it is prone to
aggregation at higher temperatures and at the same time its folding and assembly pathways have been characterized in detail. In order to understand the mechanism of the effect of glycerol on the folding pathway of P22 TSP, the compensation effect of glycerol towards urea – induced destabilization of folding intermediates was investigated by SDS-PAGE analysis and fluorescence kinetics. Further, in order to ascertain at what stage glycerol affects the folding intermediates, temperature up shift experiments were carried out wherein the protein was shifted to a higher temperature at different times of folding in the presence and absence of glycerol.