Chapter 1

Review of Literature:
Proteins are synthesized as linear polypeptide chains inside the living cell; but their biologically active conformation (usually referred as their native form) requires distinct three dimensional organizations. The cellular organelle that synthesizes protein is a nucleo-protein complex, the Ribosome.

1.1 The Ribosome:

Ribosomes from bacteria, archaea and eukaryotes consist of two subunits, a large and a small subunit. Ribosomes and their subunits are characterized by their differential sedimentation behavior upon centrifugation expressed in terms of Svedberg units (S). For example, ribosomes from bacteria and archaea are 70S, while their subunits are 50S (large subunit) and 30S (small subunit); similarly eukaryotic ribosomes and their subunits are 80S, 60S (large) and 40S (small) respectively (Wittmann, 1986). The 50S subunit contains 23S ribosomal(r) RNA (2904 nt), 5S rRNA (120 nt) and 33 different ribosomal(r) proteins, while the 30S subunit consists of 16S rRNA (1542 nt) and 21 different r-proteins. The 60S subunit comprises of 28S rRNA (4700nt), 5.8S rRNA (160 nt), 5S rRNA (120 nt) and approximately 49 r-proteins, whereas 40S subunit consists of 18S rRNA (1900 nt) and 33 r-proteins. The ribosomal subunits are connected by several well identified inter-subunit bridges (Frank et al., 1995; Yusupov et al., 2001), creating a solvent-accessible intersubunit space where the protein biosynthesis occurs (Moazed and Noller, 1991; Nissen et al., 2000). Although ribosome is a complex of RNA and protein, the catalytic function of ribosome is attributed to a particular segment of the large ribosomal RNA, like in bacteria the domain V of 23S rRNA. This is commonly recognized as the Peptidyl Transferase Center (PTC) of ribosome (fig.1.1), which is free from any ribosomal protein. The small ribosomal subunit on the other hand harbors the decoding center, where molecular recognition between the mRNA codon and the corresponding tRNA (transfer RNA) anticodon takes place. These tRNAs are extra-ribosomal adapter molecules that bring proper amino acids (specific for the mRNA codons) to the decoding center of the small ribosomal subunit e.g., 30S in bacteria. Distinct tRNA binding sites, termed as the A-site (accepting amino acylated tRNA), P-site (holding tRNA with the nascent peptide chain) and E-site (holds the deacylated...
tRNA before it leaves the ribosome), are located in the interface between the large and small subunits.

**Fig.1.1:** Ribosome & the peptidyl transferase center (PTC): 1A showing positions of PTC and the ribosomal proteins L4, L22 - within the 50S ribosome. 1B showing secondary structure of the PTC RNA as marked within the domain V of the 23S rRNA.

The solvent accessible space joining the large and the small ribosomal subunits is capable of accommodating approximately three tRNAs (Frank et al., 1995). Therefore, all of the catalytic activity of the ribosome is carried out by the RNA part of it, the r-proteins on the other hand reside on the surface and seem to stabilize the ribosome structure as well as do some accessory functions during protein synthesis, like recruiting proper translation factors etc.
The peptide exit tunnel in the Ribosome:

In the crystal structure of the 50S ribosome a passage (tunnel) has been identified as the nascent peptide exit path; exit site being at the opposite distal end of the peptidyl transfer center (fig.1.2).

Fig.1.2: Polypeptide exit tunnel is marked white in the 50S ribosome and the interior end of it showing PTC is shown in red color. Proposed exit site at the distal end of it is also marked.

The length of the tunnel has been reported to be about 100 Å, while the diameter is between 10 – 20 Å. The tunnel wall is built mainly by the nucleotides of domain I through domain V of the 23S rRNA and nonglobular parts of the ribosomal proteins L4 and L22 (Ban et al., 2000; Nissen et al., 2000). L4 and L22 proteins form the narrowest constriction along the tunnel (fig.1.1A), while the exit of the tunnel is flanked by ribosomal proteins L23 and L29. It has been shown that the length of the peptides protected by the ribosome is 30 to 40 residues. As the distance from the PTC to the distal end of the tunnel is about 80 Å, the distance of the tunnel traversed by each residue of the
nascent peptides is, on average, 2.0 to 2.7 Å. This distance is considerably less than the 3.5 Å per residue of nascent proteins in an extended conformation. An increasing body of experimental data supports the view that nascent peptides are indeed having α-helical conformation (Voss et al., 2006).

The suggestion that the tunnel is the only route of nascent polypeptide exit seems attractive, if not for the fact that in almost all the structural studies done to establish the ribosomal polypeptide exit tunnel (Nakatogawa et al 2000, Halic et al 2006, Mitra et al 2005, Woolhead et al 2006), a ribosomal tunnel driven secM sequence has been used to stall the nascent protein. However, the view is supported by the evidence that proteins, which are translocated through the endoplasmic reticulum (ER) membrane, indeed emerge from the ribosomal exit tunnel (Beckmann et al., 2001, Menetret et al., 2000). Again, there are also experimental evidences suggesting that some proteins would leave the peptidyl transferase center along other routes. For example - globin folds and binds its heme group immediately after its active site residues have been synthesized in the PTC (Komar et al., 1997). Since competence to bind the heme group requires formation of the heme binding site, and since folding in the exit tunnel is precluded by its small size, this suggests that globin cannot leave the ribosome via the exit tunnel. Globin may, instead, leave the ribosome via the interface between the two ribosomal subunits. It has been observed in bacteria that some nascent peptide chains crosslink to the 16S ribosomal RNA (Choi et al., 1998). This is the major component of the small ribosomal subunit and cannot be accessed by a chemical crosslink from inner side of the tunnel. The tail spike protein of phage P22 (Clark and King, 2001) was found to remain associated with the 30S subunit after the ribosomal subunits become separated. In a cryo-electron microscopic study, three nascent proteins, Ig1, Ig2 and GFP were found to be largely concentrated in the entry cavity of the tunnel (PTC region) in the E. coli ribosomes stalled following translation (Gilbert et al., 2004) without using secM sequence in the nascent chain. In these and also in many earlier studies, the full-length protein is found to exit from a point, which is at a distance from the PTC. Whether the exit is from within or outside the tunnel thus remains unsettled. If different peptide chains can leave the ribosome along different routes, then how the choice between these exit alternatives is
made becomes an interesting question (Tenson et al, 2002). In the same line of argument, it has also been proposed that the peptide exit tunnel in the ribosome is branched.

1.2 Protein Synthesis in Bacteria:

The process of translation involves not only the ribosome, but additional protein factors (known as translation factors), many of which are GTPases activated by the ribosome. In general one can say that, during translation single tRNAs enter specific sites in the ribosome, read the code and deliver their amino acids one at a time to a growing polypeptide chain. The ribosome advances along the messenger RNA as protein synthesis goes on, releasing the “deacylated” tRNA through an exit site. The bacterial protein synthesis takes place in four steps, namely initiation, elongation, termination and ribosome recycling.

Initiation:

Translation initiation in bacteria takes place when the 30S subunit interacts with the Shine-Dalgarno sequence on mRNA that is complementary to the 3’ end of 16S rRNA (fig. 1.3).

Fig. 1.3: Translation initiation in bacteria.
Three initiation factors, IF1, IF2, and IF3 (Gualerzi and Pon, 1990) are also involved in this process. IF3 is known as 'anti-association factor', as it binds strongly to the 30S subunit and prevent its association with the 50S subunit. It also helps in selecting the initiator tRNA (fMet-tRNA\textsuperscript{Met}) by destabilizing the binding of other tRNAs in the P site of the ribosome (Hartz et al., 1990). On the other hand, IF1 binds to the A site of the 30S subunit (Carter et al., 2001), thereby preventing tRNA binding to the A site. IF2 is a GTPase, that binds preferentially to fmet-tRNA\textsuperscript{fmet}, and its affinity for the ribosome is increased by IF1 (Zucker and Hershey, 1986). After associating with the 30S subunit, fMet-tRNA\textsuperscript{fmet} binds to the IF2, and then IF2 transfers the tRNA into the partial P site. When the 50S subunit joins, it hydrolyzes GTP to GDP and Pi, causing a conformational change in the IF2 that causes IF2 to release and allow the 70S subunit to form. Surprisingly, recent kinetic data, however, indicate that the GTPase activity of IF2 is required neither for the proper placement of initiator tRNA in the P site nor for IF2 release (Tomsic et al., 2000).

**Elongation:-**

At the end of the initiation process, an aminoacylated initiator tRNA in the P site of the ribosome and an empty A site, marks the beginning of the elongation cycle. An aminoacylated tRNA is first brought into the A site in a complex with the translation factor EF-Tu and GTP. The tRNA binding is stabilized by correct codon-anticodon interaction, which in turn triggers GTP hydrolysis by EF-Tu. This leads to the release of the aminoacyl end of A-site tRNA by EF-Tu; the tRNA then swings into the peptidyl transferase site of the 50S where peptide bond formation, involving deacylation of P-site tRNA and transfer of the peptide chain into the A site tRNA, takes place.
Ada Yonath and co-workers have shown a symmetry between the A and P sites of the ribosome (Bashan A & Yonath A, 2008) (fig.1.5). This symmetrical arrangement of the PTC allows the reactants involved in peptide bond formation to be positioned in favorable stereochemistry. After the peptidyl transfer, the ribosome has a deacylated tRNA at the P site and the peptidyl tRNA at the A site. Translocation of the tRNA and mRNA is facilitated by another GTPase translation factor EFG, thus the ribosome becomes ready for a next round of elongation with deacylated tRNA at the E site, peptidyl tRNA at the P site and an empty A site to receive the next cognate aminoacylated tRNA-EFTu complex.
Fig.1.5: Symmetry regions of the A and P sites of PTC are indicated by colors in the secondary structure, symmetry axis has been shown by arrow therein. Three dimensional positioning of the A and P site tRNAs are also shown (Bashan A & Yonath A, 2008).

Termination:-

Elongation cycle continues with polymerization of amino acids into the nascent polypeptide, until the mRNA reaches its stop codon (UAA or UGA or UAG) at the A site. In bacteria, two translation factors, RF1 and RF2, known as the “Class I release factors,” recognize the stop codons. Both of them recognize UAA, while UAG is recognized by RF1 and UGA is recognized by RF2. Binding of RF1/2 with the appropriate stop codon at the A site triggers the hydrolysis of the peptide chain from the P-site tRNA. Then another GTPase translation factor RF3, called “Class II release factor” comes into play to remove RF1/2. Hence after the peptide hydrolysis is over, the
A ribosome is left with an mRNA and a deacylated tRNA at the P site – usually referred as post-termination complex.

**Ribosome recycling:**

This is the last step of the translation process, involving dissociation of the post-termination complex (fig.1.6).

![Fig.1.6: Translation termination and ribosome recycling.](image)

This process makes the ribosomal subunits ready to initiate a fresh round of translation. It has been suggested in a cryo-EM study (Gao et al, 2007) that the RF3 brings about a major conformational alteration in between the two ribosomal subunits and that removes the last P-site tRNA. Although, a number of in vitro and in vivo studies (Janosi et al, 1998; Fujiwara et al, 2004; Hirokawa et al, 2005; Peske et al, 2005; Zavialov et al, 2005) have established the role of RRF (ribosome recycling factor), EFG (elongation factor G) and IF3 (initiation factor 3) in recycling the post-termination ribosome to re-engage them for a fresh round of translation. RRF alone has been shown to dissociate the 70S ribosome partially in a cryo EM study (Barat et al, 2007). As reported by others, GTPase activity of EFG plays crucial role along with RRF in this process and IF3 has been shown to sequester the free 30S subunit (Dallas et al, 2001) to stabilize the dissociated state (Hirokawa et al, 2005). The gene encoding RRF (*frr*) has been identified as an essential gene in bacteria (Janosi et al, 1994), deletion of which causes unscheduled translation.
termination (Janosi et al, 1998). No proper eukaryotic analogue of RRF has been identified yet, nor do we know how ribosome recycling takes place in eukaryotes.

1.3 How does protein fold within the living cell?

After they are synthesized in the ribosome, the linear polypeptide chains fold into compact three dimensional (native) forms to play their characteristic structural and functional roles in the cell. A proper mechanism explaining how the new born proteins achieve the functional forms is still unavailable.

Conformation of the nascent protein on the ribosome:

The rate of protein synthesis is not the same throughout the process of decoding one messenger RNA. This in principle depends on the abundance of the cognate tRNAs in the cytosol. The concentrations of the full set of tRNAs are known for only few organisms, among them are *Escherichia coli* and *Bacillus subtilis*. The copy number of tRNA has been shown to positively correlate with the frequency of synonymous codon usage. Some codons from each synonymous set of codons are more frequently used (usually called 'codon bias') and tend to dominate in highly expressed genes. Although it is not clear why some codons are preferred over others, but it has been demonstrated that the codon distributions might be related to secondary-structural elements in proteins. For example: α-helices are encoded by more frequent codons, whereas β-structures, loops and disordered structures are enriched in infrequent codons (Thanraj et al, 1996). It has also been demonstrated that the translation attenuation temporally divides a long protein chain into segments by producing local secondary structures in the attenuated regions (Zhang et al, 2009). These observations go in full agreement with the studies of co-translational protein folding (Fedorov et al, 1995), however in disagreement with the report that the ribosomal peptide exit tunnel can only accommodate α-helices (Voss et al, 2006). Some recent observations on the other hand indicated a significant proportion of native and non-native tertiary structure formation of nascent protein within the ribosomal exit tunnel (O'Brien et al, 2010). Within the tunnel of ribosome, a crevice has also been identified
aside the typical macrolide binding pocket, which has been suggested to be involved in the co-translational folding of protein (Amit et al, 2005).

Molecular chaperones are protein folding modulators in cell:

Molecular chaperones have been identified as a class of proteins inside the living cell that assist new born proteins in forming their biologically active forms. However, the first protein to be called a chaperone was the one that assisted the assembly of nucleosomes from folded histones and DNA in the nucleus (laskey et al, 1978). Later on the idea was extended and applied generally by John Ellis to describe a functionally related group of diverse proteins that assist the folding or assembly of other proteins (Ellis, 1987). The definition of Molecular chaperone as given by R.J.Ellis (book named “Molecular Chaperones” Publication Date: September 30, 1993 | ISBN-10: 0412550601) is: “Molecular chaperones are defined as functional classes of unrelated families of proteins that assist the correct non-covalent assembly of other polypeptide containing structures in vivo, but are not components of these assembled structures when they are performing their normal biological functions”. The general characteristics of chaperones have been extensively studied by others as well (Bukau & Horwitch,1998; Gutsche et al., 1999; Feltham & Gierasch, 2000). Chaperone DnaK with its cohorts DnaJ or GrpE (in E.coli), their eukaryotic homologues Hsp70 proteins and the DnaJ equivalents – constitute one major family of chaperones. The chaperonins like GroEL, GroES in E.coli and their eukaryotic counterparts Hsp60 family – constitute the second major family of molecular chaperones. It was estimated (Bukau, 2000) that in prokaryotic cytosol, DnaK-J can fold 5-18% of proteins (>30KDa) and GroEL-ES can fold 10-15 % of proteins (~20-60 KDa); while in eukaryotic cytosol Hsp70/Hsp40 can fold around 15-22% of proteins. Hence the above molecular chaperones can assist in folding a relatively small fraction of proteins in the cell.

Trigger factor (TF) has been identified as another chaperone which remains associated with the ribosome at the exit site of the peptide exit tunnel and it has been suggested that it protects the newly synthesized unfolded protein (Lakshmipathy et al 2007, Merz et al 2008) while it emerges out of the ribosome. Distinct attachment sites of TF to the
ribosome at the proteins L23/L29 have also been identified. But the E.coli trigger factor mutants have no obvious phenotype and can very well survive at the physiological temperature.

Is there other folding modulator than the molecular chaperones?

The fact that only a considerably small fraction of new born proteins can get assistance in folding from the chaperones prompted scientists to look for the presence of other protein folding modulators inside the cell. Different experimental observations suggested that thorough search should be conducted to check if ribosome itself has the ability to form the newly synthesized protein chain. [Zubay (1973) in vitro translation experiment giving rise to active proteins may be an example of folding in S30 mix.]

1.4. Ribosome as protein folding modulator:

In such a quest of finding additional protein folding modulators – specially within the ribosome itself – it was first discovered from our lab that the ribosomes from different sources like E. coli, B. subtilis (prokaryotes), wheat germ (plant), rat liver (animal) and *Methanosarcina barkery* (archaeabacteria) can refold a number of proteins with diverse physicochemical properties like – Bacterial Alkaline Phosphatase, Beta Galactosidase, Glucose 6-Phosphate Dehydrogenase, Glucose Oxidase, Lactate Dehydrogenase, Malate Dehydrogenase, Bovine Carbonic Anhydrase, Human Carbonic Anhydrase, Beta Lactamase, Restriction Endonuclease EcoR1, Restriction Endonuclease BamH1, DnaK HspH from *Bradishizobium japonicum*, GFP, Horse radish peroxidase, Luciferase etc. (Das et al., 1992; Bera et al., 1994; Das et al., 1996; Chattopadhyay et al., 1998; Ghosh et al., 2003). It was also shown that Ricin-A chain (Argent et al., 2000) and Rhodanase (Hardesty et al., 1995), could be folded by ribosomes (Argent et al., 2000).

The only criterion for the choice of test protein was that the protein (enzyme) should be amenable to quick and quantitative assay. Assuming that only correctly folded protein possess enzymatic activity, the measurement of enzymatic activity gain accounted for the...
extent to which the protein could be folded by ribosomes. In the in vitro experiments for each protein, the level of regain in activity with ribosome was around 80-100%.

Upon separating subunits, the protein folding activity was detected in the large ribosomal subunit from various sources (Das et al., 1996). Then the r-proteins were increasingly depleted from the large ribosomal subunit with increasing concentration of lithium chloride, and the activity was traced in its large rRNA (e.g.; 23S rRNA in case of bacteria) (Das et al., 1996; Chattopadhyay et al., 1996). Finally, the region in the domain V of *E. coli* 23S rRNA having the peptidyl transferase activity (the peptidyl transferase center, PTC) was found to have the protein folding activity (Chattopadhyay et al., 1996; Pal et al., 1997; Pal et al., 1999; Chowdhury et al., 2002; Sanyal et al., 2002; Samanta et al., 2008). Even when this segment of ribosomal RNA was cloned and transcribed in vitro, the transcribed PTC-RNA sequence could fold denatured protein equally well. Oligonucleotides complementary to the single-stranded segments in the central loop of domain V could inhibit protein folding when annealed to 23S rRNA (Chattopadhyay et al., 1996).

The results of in vitro studies were further supported by in vivo experiments. Since the protein folding activity was detected in the PTC (Chattopadhyay et al., 1996; Pal et al., 1997; Pal et al., 1999; Chowdhury et al., 2002; Sanyal et al., 2002; Samanta et al., 2008) of the large ribosomal subunit of bacteria, it could be stopped by antibiotics like chloramphenicol or lincomycin which bind to the PTC (Long et al., 2003; Douthwaite, 1992). So, the synthesis of protein was terminated with streptomycin or kasugamycin which bind to *E. coli* ribosomal 30S subunit (Carter et al., 2000; Schuwirth et al., 2006) leaving the 50S subunit unaffected. The nascent polypeptides, already made, took some time to fold which was recorded. But if the PTC binding chloramphenicol or lincomycin was added, the nascent proteins could not fold, perhaps because the PTC region responsible for synthesis as well as folding of nascent protein remained inaccessible due to antibiotic binding. The difference between protein activities in presence of streptomycin and in presence of chloramphenicol was significant and was a measure of protein folding in *E. coli*. When experiments were done with Δ trigger factor and Δ Dnak DnaJ trigger factor triple mutant *E. coli* cells, similar results were obtained, indicating that the ribosome mediated folding was independent of DnaK, DnaJ and Trigger factor, which
are known as ribosome associated chaperones. When the same protein whose folding was measured in vivo was chemical unfolded and was allowed to refold in presence of the 23S rRNA along with 50S and 30S specific antibiotics, its percent recovery of native activity decreased dramatically with the 50S specific antibiotics, whereas 30S specific antibiotics showed no effect even at very high concentrations. This indicated towards the presence of a common feature of in vivo and in vitro protein folding by the PTC RNA.

1.5. Association of Nascent and Chemically Unfolded proteins with the ribosome:

As a consequence of the above in vitro and in vivo observations, the next aim was to check whether it is possible to trap the full length nascent protein on the ribosome. For that purpose, a C-terminal His-tagged rhizobacterial protein HspH was overproduced in E.coli cell (Basu et al, 2008a). The C-terminally histidine tag helped to trap only the full length population of the protein from cell using Ni column. UV radiation was used to crosslink the nascent proteins to the ribosome population. This fixation allowed having a snapshot of the intracellular association pattern of nascent protein with the ribosome at a certain time point of cell growth. We isolated a large population of 50S bound full length nascent protein along with the relatively small 70S, indicating that the release of nascent protein from 50S, produced post translationally by splitting 70S, into the cytosol is a slow process compared to translation and its termination.

In parallel in vitro experiment (Basu et al, 2008b), FITC labeled native protein was denatured by guanidine and UV crosslinked in presence of 70S in ice. The ribosome sample was then ultra-centrifuged through 5%-20% sucrose density gradient and the FITC fluorescence was traced to be maximum at the 50S subunit peak in the gradient. So, the 70S was split by the folding protein. We observed earlier that the binding deacylated tRNA to the P site of 70S inhibits ribosome mediated protein folding (PhD thesis of Dr.Arunima Basu, CU). When ribosome having mutation at the P loop of 23S rRNA (responsible for tRNA binding at that position) were mutated, those mutant ribosomes fully retained refolding capability when subjected to P site tRNA binding conditions (PhD thesis of Dr. Arunima Basu thesis, CU).
1.6. The splitting of 70S ribosome by chemically unfolded protein suggested further experiments:

Slow release of nascent protein from the 50S, which we observed previously, prompted us to check whether unfolded protein has any role in recycling ribosome. Chemically unfolded protein was added to the 70S ribosome and the product was analyzed in sucrose density gradient (Basu et al, 2008b). The 70S ribosome was found to be completely dissociated and the kinetics of the process was followed using fluorescence light scattering technique. The chemically denatured protein dissociated 70S very rapidly (Basu et al, 2008b). It was also confirmed that neither the little amount of carried over denaturant in the final reaction mixture, nor the native form of a protein could dissociate the 70S (Basu et al, 2008b). When the 70S was pre-bound with deacylated tRNA specifically at the P-site or both the P and E – sites, unfolded protein was unable to split it (PhD thesis of Dr. Arunima Basu, CU). These observations encouraged us to further investigate the unfolded protein mediated splitting of 70S ribosome in association with the well known translation factors responsible for ribosome recycling. The exact individual /collective role of some of these factors, like EF-G, RRF etc. became the subject of discussion in the ribosome mediated nascent protein folding project.

1.7. ‘Native’ and ‘Unfolded’ states of protein: Classical Models:-

A polypeptide chain that makes up protein has thousands of atoms and hence millions of possible inter atomic interactions that could lead to millions of possibilities for its conformational rearrangements. However, only one of those conformations is biologically functional and called as its ‘native’ state. This native state is also thermodynamically the most stable conformation of the protein. Thus ‘unfolding’ of a protein means disruption of its native stability and involves disruption of those inter-atomic interactions. The ideal unfolded protein is in the random coil, in which the rotation angle about each bond of the backbone and side-chains is independent of that of the bonds distant in the sequence, and where all conformations have comparable free energies, except when atoms of the polypeptide chain come into close proximity. It has
been a practice to use either chemical denaturants (like guanidine hydrochloride/urea) or heat to unfold the native form of a protein in order to study its refolding. These in vitro studies were designed with an aim to better understand the mechanism of protein folding, which was otherwise difficult to follow in vivo. Christian B. Anfinsen in his pioneering in vitro work (Anfinsen, 1973) demonstrated the re-formation of native structure of an enzyme by diluting out the denaturant and concluded that the information for correct folding of a protein lies in its amino acid sequence. It has been shown that upon dilution, an unfolded protein regains some of its secondary structures in a millisecond order of time (Chaffotte et al, 1992); although biological activity gain is a much slow process compared to that. Nevertheless, the information that a protein can fold spontaneously immediately raised a fundamental question — among the enormous possible conformations, how does protein choose its ‘correct’ conformation that is biologically active? Levinthal concluded (Levinthal et al, 1968) that the pathway to achieve the native form of a protein can not be random, especially in the biological time scale.

Protein folding pathways and folding intermediates:

In an attempt to solve the protein folding problem, Ptitsyn proposed a model illustrating stepwise folding of protein, latter called as “frame work model” (Ptitsyn, 1973). It was postulated that the interactions within the protein molecule takes place in steps; stressing upon the rapid formation of α-helices and β-hairpins at the initial folding step followed by subsequent gluing of these helices and hairpins into native like globular form. In the in vitro refolding experiments, all these folding intermediates have been observed to date. A stable, equilibrium state of a ‘weakly denatured’ protein — termed as the “molten globule” state was suggested in 1973, then discovered and studied first in 1980-s (Ptitsyn, 1995); latter on identified as a metastable kinetic folding intermediate. According to the “frame work model”, further crystallization of protein structure within this molten globule is the last step of the folding process.

Observing that proteins’ native states often contain a hydrophobic interior of nonpolar amino acid side chains (combined with charged side chains that are neutralized by salt bridges), leaving most of the polar or charged residues on the solvent-exposed protein
The "hydrophobic collapse" model of protein folding was hypothesized (Agashe et al, 1995). According to the hypothesis, the hydrophobic collapse is relatively early event in the folding pathway that occurs before the formation of many secondary structures and native contacts present in the fully folded tertiary structure. This collapsed intermediate is also referred to as a 'molten globule' and corresponds to a partially folded state whose energy is lower than that of the completely unfolded state but higher than that of the native state. “Jigsaw model” for protein folding, on the other hand, proposed that each protein folds by a distinct pathway (Gassner et al, 1996).

Energy landscape of protein folding (Folding funnel concept):-

In physics, an ‘energy landscape’ is mapping of all possible conformations of a molecular entity or the spatial positions of interacting molecules in a system, and their corresponding energy levels (Gibbs free energy) on a two- or three-dimensional Cartesian coordinate system. The phenomenon of protein folding was largely an experimental venture until the formulation of energy landscape theory of proteins was proposed by Joseph Bryngelson and Peter Wolynes in the late 1980s and early 1990s. This approach was based on the assumption that in the path of evolution the amino acid sequences being selected for the natural proteins in the way that interactions between side chains largely favor the molecule's acquisition of the folded state; usually referred as the ‘principle of minimal frustration’.

Interactions that do not favor folding are selected against, although some residual ‘frustration’ does exist as expected. The “folding funnel” hypothesis is a specific version of the energy landscape theory of protein folding, which assumes that a protein’s native state (having a well defined tertiary structure) corresponds to its free energy minimum (fig.1.7) in solution usually encountered in cells. In contrary to the believe that the protein folding is a simple optimization problem with one minimum, in reality, free energy contains both the enthalpic as well as the entropic contributions, and hence make a direct optimization impossible.
However, the ‘folding funnel’ hypothesis is closely related to the ‘hydrophobic collapse’ hypothesis in the sense that the latter includes both the entropic and the enthalpic contributions to explain the stabilization of the final native form of protein. The driving force for protein folding, according to the hydrophobic collapse model, is the stabilization associated with the sequestration of hydrophobic amino acid side chains in the interior of the folded protein. This allows the solvent water to maximize its entropy by the process called “desolvation”, accounted into lowering the total free energy.

The theory is supported by both computational simulations of model proteins and experimental studies (Sessions et al 2004, Liu et al 2005, Rodriguez-Larrea et al 2006) and it has been used to improve methods for protein structure prediction and design.
1.8. Protein engineering and directed evolution:

Each protein is unique by virtue of its amino acid sequence. By changing one or more amino acids of the sequence, which are crucial for three dimensional structure formation of that protein, one can in principle produce a novel protein. Designing and construction of new proteins with desired function is the goal of this emerging field, named “protein engineering” (Jäckel et al, 2008) – which has vast application in different fields of biology. With the robust use of recombinant DNA technology, changing amino acids in a protein chain has become relatively simple. Nevertheless, the main challenge is in obtaining a functional form of the protein thus produced. Reason for this is perhaps the absence of any quantitative working hypothesis to relate the amino acid sequence of a protein with its functional three dimensional form. Hence a designer of protein depends mainly on intuition based trial, even to find lowest free energy conformation through bioinformatics search.

In quest of finding the protein folding modulators inside the living cell, the chaperones were discovered (discussed earlier). However, it was found that chaperones can assist in folding of a small fraction of proteins in reality and the mechanism of chaperone action only provided with qualitative information to a protein designer without giving any precise knowledge about the structure formation of a protein from its linear amino acid sequence. “Directed evolution” on the other hand is a technique very often used in protein engineering (Jäckel et al, 2008) to exploit the power of natural selection to evolve new proteins. The probability of success in a directed evolution experiment is directly related to the total library size i.e. evaluating more mutants increases the chances of finding one with the desired properties. These approaches are recently being used to search for a quantitative model of protein folding, which would relate the amino acid sequence information with the topology of its three dimensional structure formation.
1.9. PTC-RNA mediated protein folding: specificity in nucleotide-amino acid interaction:

It was found from our lab that the PTC sequence interacts with the chemically unfolded form of a number of unrelated proteins through the same five RNA nucleotides (Samanta et al, 2008). The amino acids of different proteins that interact with those nucleotides are not random either; there is a strong consensus – one of the amino acid-nucleotide pairs being conserved in that small set of data (Das et al, 2010). Regarding the position of those amino acids, they mostly lie in random coils on the surfaces of the crystal structures of globular proteins (Samanta et al 2008, Das et al 2010). These observations prompted us to further search for a quantitative mechanism of protein folding based on the involvement of PTC-RNA.

1.10. Aim of the present work:

Based on our previous observations we asked a few questions that we considered relevant:

1) How far the in vivo behavior of nascent protein (which is unfolded) is comparable to the behavior of chemically unfolded protein(s) in the ribosomal environment? This is the subject of chapter 2 of the thesis.

2) Should the nascent protein be left out of the scheme of ribosome recycling by RRF-EFG? (No RRF analogue has been identified for eukaryotes yet) If so, then how two parallel ribosome recycling pathways could be relevant in bacteria? This is the subject of chapter 3 of the thesis.

3) Interaction of in vitro transcribed PTC sequence with the chemically unfolded protein has a temporal component. Comparing in vitro and in vivo interaction patterns, is it possible to figure out a molecular mechanism of PTC mediated protein folding? This is the subject of chapter 4 of the thesis.

The above questions have been addressed through the works described in the thesis.