Introduction
Cells are the structural and functional units of all living organisms. Rudolph Virchow, the nineteenth-century Prussian pathologist and researcher, concluded "Omnis cellula e cellula": every cell comes from another cell. This characteristic feature of cell is by the virtue of the genetic continuity vested in the single DNA (or RNA) molecule. The Structure of DNA allows for its replication and repair with near-perfect fidelity. The linear Sequence in DNA encodes proteins with three-dimensional structures. Proteins are long polymers of amino acids which constitute the largest fraction (besides water) of cells. Proteins are the molecular instruments through which genetic information is expressed. They are the most abundant biological macromolecules, occurring in all cells and all parts of cell. Proteins occur in great variety; some proteins have catalytic activity and function as enzymes, others serve as structural elements, signal receptors, or transporters that carry specific substances into or out of cells thus they are perhaps the most versatile of all bio-molecules.

All these properties of a protein are by virtue of its three dimensional structure which is determined by a process called "Protein folding". The general characteristics of protein folding are discussed in the present chapter.

1.1 **Protein folding: an Overview**

Pioneering experiments by Anfinsen over four decades ago showed that proteins can fold spontaneously *in vitro*, providing the fundamental insight that the linear sequence of the polypeptide chain contains all necessary information to specify a protein's three-dimensional structure (Anfinsen 1973). These findings suggested that protein folding inside a cell could be a spontaneous process. However, since the late 1980s it has become clear that in the cell, a large fraction of newly synthesized proteins require assistance from molecular chaperones to reach their folded states efficiently and at a biologically relevant time-scale (Hartl 1996). Several ideas to explain the mechanisms of folding have been proposed from experiment and simulation. The famous Levinthal paradox indicated that folding cannot occur by a stochastic exploration of all possible conformations, for even a small protein would need a biologically unrealistic timeframe to do so (Levinthal 1968). Indeed, the attractive and repulsive forces between neighbouring amino acid residues favour certain conformations of individual amino acids in the polypeptide chain, thereby dramatically reducing the conformational space and the number of possible folding pathways available.

Folding typically does not occur as an all-at-once process but involves intermediates. Their formation is considered the rule for larger proteins. Ninety percent of proteins in a cell have a significant tendency to rapidly collapse in aqueous solution into compact non-native conformations (Bartlett and Radford 2009; Brockwell and Radford 2007). Such intermediates either represent on-pathway "stepping stones" toward the native state or kinetically stable, misfolded conformations that may...
require substantial reorganization before the native state can be reached. The formation of metastable, non-native interactions during folding is interpreted as a consequence of the ruggedness of the funnel-shaped folding energy landscape (Fig. 1.1) (Jahn and Radford 2005; Lindberg and Oliveberg 2007; Onuchic and Wolynes 2004).

Energy landscape theory and the funnel concept provide the theoretical framework towards a quantitative understanding of the folding question (Bryngelson et al. 1995; Onuchic et al. 1997). This alternative view for the folding mechanism replaced the earlier idea that there must exist a single pathway for the folding event with clearly defined chemical intermediates (Englander and Mayne 1992; Kim and Baldwin 1990). Fig.1.1 represents the key features of a typical folding funnel. In such a funnel, there is not a unique folding pathway but a multiplicity of folding routes, all converging towards the native state.

At the top of the funnel, the protein exists in a large number of random states that have relatively high enthalpy and high entropy. There is a fight between the maximization of entropy keeping the protein as random as possible at the top of the funnel and the minimization of enthalpy trying to drag the protein down the funnel as it folds. Progress down the funnel is accompanied by an increase in native-like structure as folding proceeds.
Progress down the funnel is measured by a parameter Q, which is ratio of native contacts. Thus, the depth of the funnel represents energy, while the width represents entropy of the system. Folding is driven by an increase in Q as the collection of states rapidly inter-converts i.e. the molten globule state is formed with a Q around 0.27. The transition state occurs around 0.6 and intermediates normally get populated around a Q value of 0.7 (Fersht 1998). The downhill bias of the energy landscape is sufficient to drive random folding in a reasonable time period. Folding precedes one amino acid at a time in no particular order through an unlimited continuum of undefined intermediates and paths (Rumbley et al. 2001). There is an explicit assumption in the funnel mechanism that there is kinetic flow down through a series of states that are progressively lower in energy. A general description of different type of conformations stabilized during folding/unfolding of proteins is discussed below.

**Native state**

A 'Native' protein is one where the inter-atomic and intramolecular interactions are maximum, the conformation is rigid with very low free energy (Imoto 1997). The native state is not a single state but an ensemble of macroscopic states that are structurally very similar but are separated by measurable energy barriers.

**Intermediate States**

In equilibrium studies, only two states of the protein, viz., fully folded and fully unfolded states are considered to be the main species at any time in the stoichiometry determined by their partition function (Freire 1995; Privalov 1979). In between, these two states the intermediate states exist. Intermediates are an intrigue and are not universal in case of folding proteins and have been reported to arise as a consequence of energy barrier (Xie et al. 1993). Infact, the high cooperativity (Tanford 1968) and rapidity (Matouschek et al. 1990) do not favor accumulation of intermediate. Out of all intermediates reported molten globule (MG) and burst phase (BP) are the most characterized.

**Molten Globule:** Molten globule is stabilized in (Kuwajima 1989; Ohgushi and Wada 1983) mild denaturing conditions such as low pH, high ionic strength, low concentration of GdnHCl or Urea (Goto 1990; Jeng et al. 1990; Uversky and Ptitsyn 1994) and temperature (Griko et al. 1994; Kuroda et al. 1992). It is condensed and compact, with a stoke's radius slightly greater than that of the native state and occupies 10-20 % more volume than that of the folded state (Uversky and Ptitsyn 1994). It exhibits substantial secondary structure but many side chains lose persistent tertiary interactions of the native state (Jeng et al. 1990). The molecules behave as sticky due to the exposure of hydrophobic surfaces and bind to the hydrophobic molecules like ANS (Semisotnov et al. 1991). Inside a cell, molten globule state is involved in several biological and pathological processes like
membrane insertion, trans-membrane trafficking, and chaperone assisted refolding and that requires the proteins to become unfolded (Hartl 1994).

**Burst Phase:** All proteins for which kinetics of folding have been followed by circular dichroism (CD) show a substantial amount of secondary structure formation during the first few millisecond of refolding under strongly native condition (Kuwajima et al. 1985; Sosnick et al. 1996). This structure has been named as 'Burst Phase' (BP). BP intermediate is classified as the same species as MG but differs in many aspects from MG. It is shown (Baum et al. 1989; Jeng et al. 1990) that BP intermediates are highly dynamic states without stable hydrogen bonds. The characterization of BP becomes difficult also because of the scarcity of the methods that can follow folding reactions over μs time scale.

**The Denatured State**
Folding of a denatured protein starts from an undefined initial 'state' which represents an ensemble of readily inter-convertable conformers with equal or closely similar energies. Only under ideal solvent conditions (which do not exist because there is no good solvent for the backbone and the chemically diverse side-chains of the 20 natural amino acids), the polypeptide chain would approaches a random coil.

High concentrations of guanidinium chloride (GdnHCl) are assumed to cause 'complete denaturation'. As suggested by thermodynamic evidence and statistical considerations, based on estimates of stabilization energies and conformational entropies, other denaturing conditions such as elevated temperature, lead to the same extent of unfolding (Creighton 1978; Privalov 1979). On the other hand, spectral data show significant differences comparing, e.g. far-UV circular dichroism spectra of proteins in GdnHCl, urea or at high temperature and acidic pH. Here, for the degree of denaturation $U_{GdnHCl} > U_{urea} > U_{acid} > U_{pH 2}$ is widely accepted. There is evidence from recent NMR data, proving that even under extreme denaturing conditions, some local residual structure is preserved (Baldwin 1986). Obviously, the ‘denatured state’ comprises a large number of significantly different conformations in rapid equilibrium with each other. In the framework of the description of folding and stabilization of proteins as the descent in an ‘energy landscape’, denatured polypeptide chains would populate multi-minima sub states in the ‘funnel’ part of the landscape that are characterized by rms differences in the range of many Angstroms.
1.2 Thermodynamics of folding

The covalent backbone of a typical protein contains hundreds of individual bonds. Because free rotation is possible around many of these bonds, the protein can assume an unlimited number of conformations. However, each protein has specific chemical and structural characteristics, strongly suggesting that each has a unique three-dimensional structure.

The spatial arrangement of atoms in a protein is called its conformation. The possible conformations of a protein include any structural state that can be achieved without breaking covalent bonds. A change in conformation could occur, for example, by rotation about single bonds. Of the numerous conformations that are theoretically possible in a protein containing hundreds of single bonds, one or (more commonly) a few generally predominate under biological conditions. The need for multiple stable conformations reflects the changes that must occur in most proteins as they bind to other molecules or catalyse reactions. The conformations existing under a given set of conditions are usually the ones that are thermodynamically the most stable, having the lowest Gibbs free energy (G).

Understanding protein structure, folding and stability involves unraveling mechanisms by which natural proteins attain their special physical properties as compared to random heteropolymer. To explain these properties, it has been theoretically proposed that an energy gap exist between native and any of the other possible conformations of the proteins (Bryngelson et al. 1995; Karplus and Sali 1995). This gap requires the native sets of non-covalent interactions between amino acids to be more stabilizing than those in all other conformations. In most naturally occurring proteins, the folded conformations are generally only 5 to 10 kcal/mol more stable than unfolded, biologically inactive conformations under physiological conditions (Pace 1990). This stability of proteins is by the virtue of a delicate balance between large driving and opposing forces. Main opposing force is configurational entropy which arises from a large number of chain configurations accessible to the denatured state (Dill 1990; Pace 1990). Positive contribution (driving force) arises from non-covalent interactions. There are two types of non-covalent interactions in the protein: local and nonlocal, so defined in terms of the distance in the sequence between the interacting residues (Dill 1990). Importantly, they are geometrically different; local interactions participate in defining secondary structure while nonlocal is involved in defining tertiary structure. Both types of interactions can be either sequence independent, those between atoms of the backbone (i.e. α-helix main chain hydrogen bond) or depend on the chemical nature of the residues involved (Munoz and Serrano 1996).

The small net conformational stability is the result of much larger contributions from several important forces (Fersht 1993; Makhatadze and Privalov 1995; Matthews 1995). It is generally believed that proteins are stabilized by hydrophobic interactions, hydrogen bonding and van der Waal
interactions (Fig. 1.2). Hydrogen bonds allow the burial of the protein backbone groups within core of proteins whereas van der Waal interactions are responsible for packing. Hydrophobic forces favor the formation of protein hydrophobic core exploiting the tendency of non-polar amino acid side chains to escape from contact of water (Voet and Voet 2004). Conformational entropy poses folding because the polypeptide chains have many degrees of freedom that become fixed on folding. The large favorable contribution of these forces is offset by a large unfavorable entropy change on protein folding. The resulting stability of the protein, as measured by Gibbs free energy ($\Delta G$), is only in the order of few kilojoules. Studies on the stability of protein mutants in which hydrogen bond have been removed and calorimetric analysis of the $\alpha$-helix to coil transitions of a almandine peptide in water appear to suggests that both hydrogen bonding and hydrophobic effect make comparable contribution to the stability of globular proteins (Takano et al. 1997). The fact that the stability of globular proteins is low implies that no interactions are unimportant and even small interactions can make significant contribution to the $\Delta N^\circ G$ value (Dill 1990).

Electrostatic interactions arises due to interaction between two charged or polar groups, which are described in terms of pair wise columbic interactions between individual atoms: Hydrogen bond is a type of dipole-dipole interaction and van der Waals interaction, an induced dipole interaction, are electrostatic in nature (Burley and Petsko 1988; Honig and Yang 1995; Jencks 1987; Klotz and Farnham 1968). The association of two ionic groups of opposite charge is known as an ion pair or salt bridge. The energy of association, $U$ required to bring two electric charges $q_1$ and $q_2$ from infinite distance to a separation distance $r$ is:

$$U = \frac{kq_1q_2}{Dr}$$

Here $k=9.0 \times 10^9$ J.m.C$^{-2}$ and $D$ is dielectric constant of the medium in which the charges are immersed ($D=1$ for vacuum). Since ion sable group are constrained in a folded protein, the distance $r$ between the charged groups will be less in the folded state than in unfolded state. In addition, because the solvent water has high dielectric constant ($D=80$) compared to the interior of the protein ($D=4$), the effective dielectric constant for each pair of the charges is lower in folded protein. For these reason the effect of the charged groups on a protein should be much greater in folded state than in unfolded state (Imoto 1983; Tanford 1961). Electrostatic interactions occur between charges on the proteins group (Fairman et al. 1996). Such charges are present at the amino- and carboxyl- termini and on many ion sable side chains (Murphy 1995). Contribution of electrostatic interactions is highly dependent on the spatial location of charged group. In a typical protein about 75% charged residues occur in ion pair and very few ion pairs are buried (insolated) and ion pairs that are exposed to the aqueous solvent tend to be, but poorly conserved among homologous proteins.
The oxidation of the sulfhydryl groups of two cystein residues (intramolecule: ribonuclease; intersubunit: dimeric protein insulin)

Dipole molecules attract each other by van der Waals force (transient and weak: 0.1-0.2 kcal/mol)
Hydrophobic interaction, a tendency of hydrophobic groups or molecules being excluded from interact with hydrophilic environment

Fig. 1.2: A schematic depiction of important interactions and their relative contributions to the protein stability.

Current advancement in protein engineering via site directed mutagenesis (Carra et al. 1994a, b; Kiefhaber et al. 1990; Monera et al. 1994; Nicholson et al. 1988; Ogasahara et al. 1998; Spassov et al. 1995; Waldburger et al. 1995) and chemical modification of amino acids (Goto and Nishikiori 1991; Hagihara et al. 1994) has made it possible to explore the role of electrostatic interactions in protein stability. The contribution of a single salt bridge to the free energy of stabilization to the native state is about 3-5 kcal/mol. Mutational studies with cold shock proteins shows that the mesophilic protein could be converted in to a highly thermostable form by changing the glutamic acid residues at position 3 and 66 to arginine and leucine respectively (Perl et al. 2000). Overall, it can be said that ionic interactions, though very important in providing specificity and stability to the protein structure, are highly dependent on the spatial location of charged group.

1.3 Folding of monomeric and multimeric proteins

The field of protein folding has traditionally focused almost exclusively on the study of individual domains. This attention to single-domain protein fragments or small proteins has been for two reasons. First, large proteins can be difficult to study. Second, it is assumed that a protein domain is an autonomous folding unit and that folding principles identified from these studies of single domains will be generally applicable. A domain is a structural, functional and evolutionary component of
proteins, which can often be expressed as a single unit. There are a number of additional advantages: (i) domain folding is an efficient way of excluding wrong intramolecular interactions in the case of large protein molecules, (ii) it protects the nascent polypeptide chain from proteolysis and (iii) it may be considered as a simple mechanism to proceed from monomeric to multimeric proteins by ‘domain swapping’ (Bennett et al. 1995).

**Folding of monomeric proteins**

**Single domain proteins:** Equilibrium denaturation and kinetic studies on folding/unfolding of proteins have largely been done for monomeric proteins. The high degree of reversibility, unimolecular folding kinetics, less complexity in handling and feasibility of NMR structure solution have made small monomeric globular proteins an attractive system for studies. In addition, the molecular biology tools, rapid mixing stopped flow techniques and hydrogen exchange methods have helped to study a wide variety of proteins (Van Nuland et al. 1998). The small globular monomeric proteins can be considered as single domain proteins. Usually, such proteins have been found to fold very rapidly on the microsecond time scale without accumulation of any intermediate (two state model) one such example is chymotrypsin inhibitor 2 (Li and Daggett 1996) however, during unfolding of many single domain protein intermediate states do exists e.g. Barnase (Li and Daggett 1998).

**Multidomain proteins:** Computational analyses of sequenced genomes have shown that most proteins contain more than one domain. Proteins have evolved through a process of extensive duplication and shuffling of domains. The modular nature of domains has facilitated the generation of novel and complex protein functions from a limited set of domain families. It is important to address the effects that the presence of neighbouring domains has on protein folding properties. The effects of neighbouring domains can be threefold. The first concern is thermodynamics, are domains stabilized by their neighbours? The second concern is kinetics, is the folding or unfolding rates altered and is the folding pathway of the domains the same in single-domain and multidomain constructs? The third issue is how domains in multidomain proteins avoid misfolding with their neighbours.

To investigate in detail the effects of neighbouring domains, one has to study both the stability (thermodynamics) and folding rates (kinetics) of the domain in isolation as well as in the multidomain protein. However, there can be difficulties with this approach. The kinetics and thermodynamics of multidomain proteins can be complex to analyse. Also, in multidomain proteins it can be difficult to define where one domain ends and where the next begins. Artificial shortening of domains can lead to lower stability measurements or non-cooperative folding. Furthermore, it can be difficult to express a soluble form of an individual domain when there are extensive inter-domain interfaces.
This might indicate that the interfaces have an important role in reaching and maintaining the correct fold, but it could simply be an effect of 'stickiness' if the exposed interface of the isolated domain is largely hydrophobic in nature, this can lead to aggregation in the cell during expression. There are hardly any evidences of how different characteristics of the domain interface affect folding. Intuitively, a large complementary interface will not only allow domains to adopt specific conformations relative to each other but will also function to stabilize a protein.

**Folding of multimeric proteins**

Multimeric proteins have been less extensively studied (Neet and Timm 1994), although they provide an opportunity to assess the relative stabilities of intra- and inter-subunit interactions. Reversible unfolding studies of large, oligomeric proteins are hampered by the difficulty of achieving quantitative refolding. Though the folding of individual polypeptide chains of multimeric proteins may be similar to that of single chain proteins, they would in addition be accompanied by simultaneous interactions with the other monomers. The nature of interactions in monomeric and multimeric proteins can be the same but denaturation/renaturation is much more complex in oligomeric proteins (Seckler and Jaenicke 1992). Protein association requires a specific site on each of the reactants, so it can occur only after these sites have appeared. Each association reaction should thus occur only after separate folding of each of the reactants.

Folding of an oligomeric protein from its denatured and separated chains probably begins like that of single-chain proteins and proceeds until the formation of a specific binding site that can recognize another monomer. At this critical step, the folding pathway shifts from being intramolecular to intermolecular, to yield a dimeric species. This homo- or hetero-dimer may need further folding steps to become either a native protein or an intermediate with an adequate specific site that allows a second association step to take place. The overall folding pathway of an oligomeric protein is thus a succession of mono-molecular folding steps and bi-molecular association steps (Jaenicke 1987).

**1.4 Converging concepts of protein folding in vitro and in vivo**

Proteins in the test tube and in the cell are subject to the same laws of physics, so what is special about folding under cellular conditions, and why are chaperones necessary? The increasing availability of highly sensitive biophysical techniques to study folding *in vitro* and in cellular systems is now providing new insights into these issues (Bartlett and Radford 2009). These studies also shed light on the process of aggregation, a potentially dangerous off-pathway reaction that can cause disease and must be prevented by molecular chaperones.
foldons (Englander et al. 2007; Lindberg and Oliveberg 2007). Examples of such minimal nucleation motifs are the two-stranded-helix motifs found in $\alpha/\beta$ domain proteins. Multistate folding behaviour with populated intermediates would be observed when multiple foldons are separated and do not act cooperatively (Lindberg and Oliveberg 2007) or when foldons misassemble, resulting in a kinetic block of folding. The propensity to misfold increases with topologically complex fold types that are stabilized by long-range interactions (for example, $\alpha/\beta$ domain architectures) or when proteins contain multiple domains that are separate in the native state but may interact during folding (Netzer and Hartl 1997; Wright et al. 2005).

New, fluorescence-based techniques now allow protein folding and aggregation to be observed in vivo in real time (Ignatova and Gierasch 2004). These and other studies indicate that the tendency of partially folded proteins to aggregate is greatly enhanced in the highly crowded environment of the cell, largely explaining the requirement of molecular chaperones (Ellis and Minton 2006). Whereas folding experiments in vitro are typically performed in dilute solution to minimize aggregation, in the cell, folding occurs in the presence of 300–400 g l$^{-1}$ of protein and other macromolecules. The
folding experiments in vitro are typically performed in dilute solution to minimize aggregation, in the cell, folding occurs in the presence of 300–400 g l⁻¹ of protein and other macromolecules. The resulting excluded volume effects substantially enhance the affinities between interacting protein molecules, including folding intermediates.

The translation process can potentially further increase the risk of misfolding and aggregation, because incomplete polypeptide chains cannot fold into stable native conformations. Additionally, the exit channel of the large ribosomal subunit, which is 100 Å long but, at most, 20 Å wide, largely precludes folding beyond the formation of α-helical elements (Lu and Deutsch 2005; Woolhead et al. 2004) and thus prevents the C-terminal 40–60 residues of the chain from participating in long-range interactions (Kramer et al. 2009). As a consequence, productive folding can occur only after a complete protein or at least a domain (~50–300 amino acids) has emerged from the ribosome, consistent with the general rules of folding established in vitro and supported by recent simulations of nascent chain folding (Elcock 2006). Because translation is relatively slow nascent chains are exposed in partially folded, aggregation-sensitive states for prolonged periods of time. Moreover, non-native intrachain contacts formed during translation could block folding upon completion of synthesis. Molecular chaperones therefore interact co-translationally with nascent polypeptides and inhibit their premature (mis)folding (Fig. 1.3). For example, the chaperone Trigger factor binds to the small titin I27 chain (~120 amino acids) until its complete β-sandwich domain has emerged from the ribosome (Kaiser et al. 2006).

The cellular chaperone machinery ensures that folding is efficient for most proteins (Vabulas and Hartl 2005). We define a molecular chaperone as any protein which interacts, stabilizes or helps a non-native protein to acquire its native conformation but is not present in the final functional structure. Chaperones are involved in a multitude of cellular functions, including de novo folding, refolding of stress-denatured proteins, oligomeric assembly, intracellular protein transport and assistance in proteolytic degradation. Numerous classes of structurally unrelated chaperones have been described (Chang et al. 2007; Tang et al. 2007). Many of these are known as stress proteins or heat shock proteins, as they are upregulated by cells under conditions of conformational stress in which the concentration of aggregation-prone folding intermediates increases. Chaperones are usually classified according to their molecular weight (Hsp40, Hsp60, Hsp70, Hsp90, Hsp100 and the so-called small Hsp proteins). Although all these components have the capacity to prevent aggregation, only certain members of the Hsp100 family in bacteria and fungi can actively dissociate aggregates for subsequent protein refolding or degradation (Weibezahn et al. 2005). The cellular chaperone machinery forms complex networks that are indispensable for protein quality control and maintenance of protein homeostasis (Balch et al. 2008).