7.12. PROTEIN FOLDING AND MISFOLDING[43]

"Protein folding" usually refers to the process that results in acquisition of the native structure from a completely or partially unfolded state. Protein folding studies started half a century ago and were substantially advanced in recent years by development of fast spectroscopic methods that allow characterizing intermediates on the folding pathway, single-molecule techniques and computational approaches. These modern approaches are solving the long-standing problems: (1) the short-lived intermediates cannot be characterized using traditional structural approaches, and (2) potential heterogeneity of folding trajectories of individual polypeptides within the ensemble. Protein folding-unfolding reactions give important insights into the factors stabilizing (and destabilizing) the native form of protein. They shed light on the question of the mechanism of the assembly of specific, organized protein conformations from newly synthesized polypeptide chain. The manner in which a newly synthesized chain of amino acids transforms itself into a perfectly folded protein depends both on the intrinsic properties of the amino acid sequence and on multiple contributing influences from the crowded cellular milieu. Folding and unfolding are crucial ways of regulating biological activity and targeting proteins to different cellular locations.

In 1969 Cyrus Levinthal proposed a puzzle in the theory of protein folding, the so-called Levinthal paradox. He noted that, since an unfolded polypeptide chain has a very large number of degrees of freedom (given a small 100 amino acids protein, and assuming 3 possible conformations per peptide bond (3^{100} conformations). Assuming that fastest motions happen at 10^{-15} sec time scale, and if the protein is to attain its minimum energy configuration by sequentially sampling all the possible conformations, folding of a single protein should take longer than the age of the universe (known as Levinthal's paradox). Proteins, however, fold spontaneously into single stable native conformations on a millisecond or even microsecond time scales. Thus, the Levinthal's paradox serves to demonstrate that an intensive, purely random search cannot succeed. The two views compete to explain how proteins fold into the minimum energy native conformation:

1. Folding proceeds through a predefined pathway[44]. The unit steps on this pathway correspond to cooperative folding/unfolding units, referred to as foldons. Although the foldons unfold and refold repeatedly under the native conditions, their addition is stepwise and is guided by the preexisting structure in a sequential stabilization process. The unfolding process is the reversal of the stepwise folding. This hypothesis is built upon Levinthal's original conjecture that because the protein cannot assume its minimum energy conformation by random sampling of ALL available conformations, it therefore should follow a distinct predetermined pathway and that the "native" state is not necessarily has to represent the global energy minimum.

[43] Additional discussion of this topic can be found in several recent reviews:
Dobson CM (2004) Seminars in Cell & Developmental Biology V.15 pp 3-16
Clark PL (2004) *TIBS* V.29 pp 527-534
Luheshi et al (2008) Current Opinion in Chemical Biology v.12 pp 25-31
Tartaglia et al (2007) *TIBS* v. 31 204-206

[44] Ref: Englander et al (2007) Quaterly Reviews of Biophysics 40 4 (2007) pp 287-326

Sequentially appearing stable intermediates were indeed identified over the years in some proteins.

2. Another (more generally accepted) view asserts that proteins fold along a funnel-like energy landscape rather than searching at random in conformational space. In this case, the search through conformational space can still be random, but energetically biased to proceed downhill. In contrast to specific "predetermined" pathway, the protein would stochastically follow one of many energetically endorsed pathways.

Due to the flexibility and dynamic nature of the incompletely folded polypeptide chains, residues located at a distance from one another could come in contact and form productive pairs. A very small number of such productive interactions may serve as a nucleation point for the folding reaction. Due to stochastic nature of chain fluctuations, folding of an individual polypeptide chain may take a number of trajectories along the complex energy landscape. The free folding landscape describes the free energy of a polypeptide chains a function of its conformational properties and is typically shaped like a funnel. Each point on the funnel surface represents a specific possible configuration of the chain and its energy value (**Figure 7.13**).

Figure 7.13

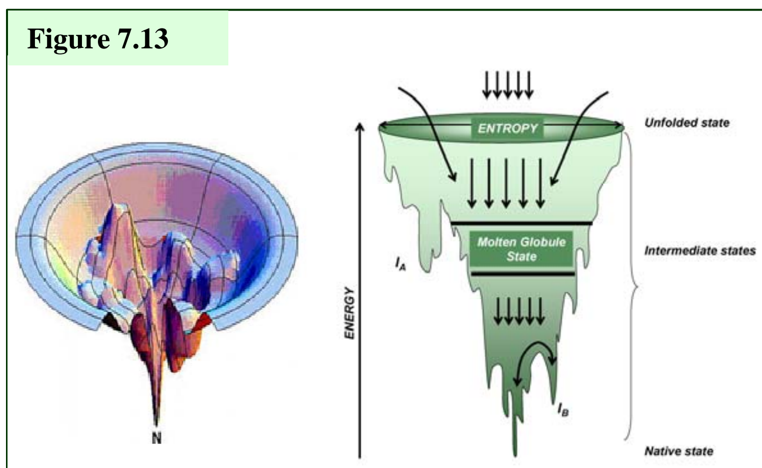


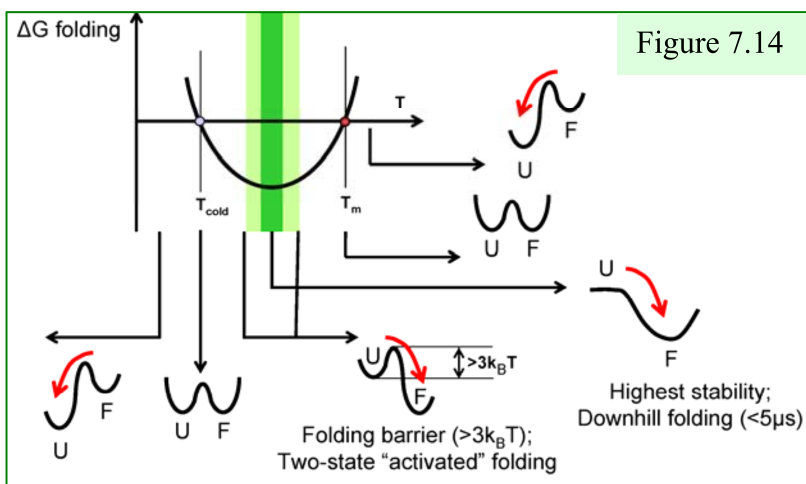
Figure 7.13 (right) shows a simplified version of a rugged protein folding funnel which describes the system in terms of configurational entropy (x-axis), i.e. the number of available states, and the free energy of conformation with the chain entropy term subtracted (y-axis). The y-axis may also represent the degree of nativeness (the fraction of native contacts formed). The exhaustive description of the protein at every point on every possible path towards the native state will yield a multidimensional funnel taking in account all parameters (i.e. all possible configurations of rotatable bonds and amino acid contacts) describing the protein structure.

The 3D funnel on the left is a pictorial representation of the funnel-like energy landscape where an ensemble of denatured (unstructured) chains populates the outer rim characterized by high conformational entropy.

Under physiological conditions, the region of configuration space associated with the native state has the lowest free energy and therefore is thermodynamically the most stable. The ensemble of denatured states can flow downhill through many alternative routes. The trajectory down the funnel restricts the number of conformations accessible to the chain as the native state is approached. To compensate for the loss in the high degree of disorder, the more favorable enthalpy is associated with stable native-like interactions observed in the bottom of the funnel. The separation between the top and the bottom of the funnel represents other energy contributions (chain enthalpy, solvent entropy and enthalpy) to each chain conformation.

The energy landscape theory and kinetics of protein folding:

The energy landscape theory states that folding of a protein does not follow a single predefined pathway, instead it suggest that protein folding is a collective self-organization process that may occur via any of the number of possible intermediates. Hence, the protein can take a multiplicity of routes down the funnel assuming that the correctly folded native state is down in the very bottom (**Figure 7.13**). The fraction of chains (or a probability of a chain) to follow a particular route is weighted by the relative heights of the barriers on this route. An important prediction from the energy landscape theory is that the two components of the free energy, enthalpic and entropic can compensate for one another. The balance between the entropic and enthalpic contributions is defined by the temperature. Thermal stability of proteins (in particular globular proteins) differ from that charachteristic to most of the ordered molecular systems. In addition to transition between disordered state at high temperatures and ordered (native) state at low temperatures, where the thermal fluctuations are suppressed, many proteins exhibit additional transition attributed to cold-induced denaturation (**Figure 7.14**). At higher temperatures, the imput of the heat into the system allows for more disorder. The increase in the entropy at temperatures above melting temperature of particular protein compencates for loss of the specific interactions and subsequent increase in the enthalpy and as result, allows the protein to unfold. At very low temperatures (below the cold denaturation temperature), the free energy of hydrating the non-polar residues in the protein core becomes increasingly more negative, which promotes protein unfolding. The presence of two transitions implies that each protein has a temperature region at which it will be the most stable. At the temperature of maximal stability, the folding landscape becomes very smooth allowing very fast transition between unfolded state and folded state (under 5 μs).



Folding kinetics of many analyzed proteins can be described as a "nucleation – condensation", where nucleation corresponds to the formation of a folding nucleus of a small number of key residues. The first step is the movement of hydrophobic R-groups out of contact with water. This drives the collapse of the polypeptide into a compact and dynamic "molten globule". Formation of the globule is driven by entropy since it allows for a great number of conformations. The dynamic nature of this state also allows acquisition of the native-like environment for a small number of key residues that constitute the core of the folding nucleus. The remainder of the structure then condenses around the folding nucleus. In the case of the small proteins (fewer than 100 amino acids) collapse of the polypeptide chain into the stable compact structure occurs only if majority of the critical contacts are formed. Once it happens, the structure rapidly converts into the folded state. Remarkably, there is almost perfect correlation between folding rates of small proteins and the contact order of their structures (average separation along the polypeptide chain of the residues found to contact each other in the native state).

For the fast folding proteins, nucleation step occurs rapidly and allows formation of only “productive” (native-like) contacts. The protein then undergoes rapid “downhill” folding indicative of the absence of any significant energy barriers on the path between unfolded and native states. At suboptimal temperatures or in structurally more complex proteins, the “correct” nucleation may present an energy barrier due to the presence of multiple similar energy states displaying non-native contacts. On its way to the folded state the protein has to surmount this barrier – the non-native contacts have to be broken and the native-like contacts have to be formed. If the barrier height is significant (>3k_BT), the protein fill follow the “activated” or two-state folding mechanism characterized by single-exponential kinetics. The barrier height for such proteins can be obtained from kinetic measurements carried out at different temperatures (the nature and treatment of activated processes will be discussed in details in the next lecture). At the temperatures near the most stable zone, the free energy barrier diminishes below 3k_BT resulting in the incipient downhill folding characterized by non-exponential kinetics. For fast folding proteins, the shape of the folding landscape (downhill vs. low barrier folding) can be greatly affected by a protein sequence – a single strategically positioned amino acid substitution may result in switching of the folding mechanism without disrupting the topology of the folded state. The solvent conditions (temperature and the presence of non-polar solvent) represent other factors that influence the mechanism of folding.

Rougthness of the funnel reflects local minima, which can trap suboptimal folding intermediates. As the chain folds to lower energy conformations, it might populate intermediate states along the sides of the funnel (represented as I_A and I_B). In statistical mechanics terms, these traps represent the degree of frustration of the polypeptide sequence. Dependending on their depth, the barrier between the trap and native conformations, these kinetic traps may hinder or promote formation of the native structure.

It is worth noting here the importance of making distinction between “folding pathway” and “folding mechanism”[45]. While the former provides a comprehensive description of a temporal order in which the contacts are made and consolidated, the latter (see **Figure 7.15**) describes the relative abundance of secondary structure elements in the tertiary structure of folding chains irrespective of exact location of these elements.

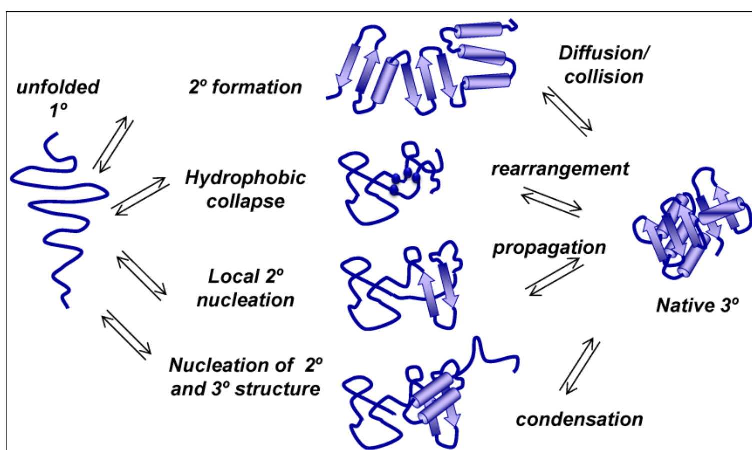


Figure 7.15 Four classic protein folding mechanisms

Some structurally related proteins fold with the same mechanism and through similar pathways. Proteins containing the immunoglobulin-like (Ig-like) β-sandwich fold represent an extensively studied example of this situation. These proteins fold by a nucleation-condensation mechanism with the residues that form the obligate folding nucleus being highly conserved within the Ig superfamily.

[45] Beautifully reviewed by Nickson and Clarke (2010) Methods

Related proteins may follow the same folding pathway (i.e. proceed through the same folding intermediates) but with different kinetic profiles. For example, one of the proteins can have an on-pathway kinetic trap, while closely related protein will proceed through the same transition state without the trap.

Larger proteins generally fold in modules, each folding module corresponds to a segment or domain of a protein. The key interactions within each separate domain define local folding, while other specific interactions between completely or partially folded modules ensure that the independent modules subsequently interact to form the correct overall structure.

MISFOLDING AND AGGREGATION

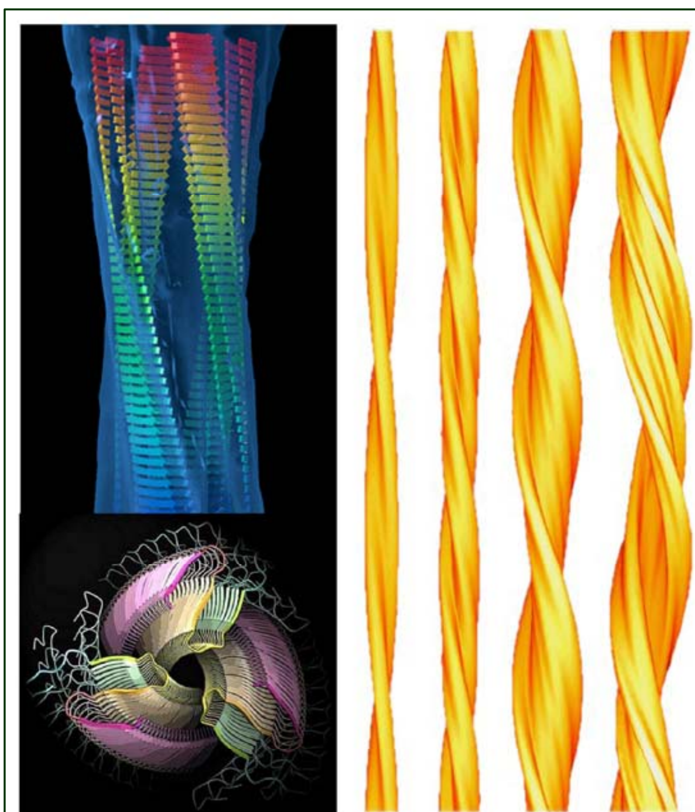


Figure 7.16 Molecular models of the amyloid fibril. A. Side view of fibrils grown from an SH3 domain (REF: Jimenez et al EMBO J 1999). The fibril consists of 4 protofilaments to form a hollow tube with diameter of approximately 6 nm. The protofilaments are thin and flat, so that the beta-sheets must be nearly flat. The cross-beta part of the model is shown inside the EM density (transparent blue surface). **B**. The superpleated beta-structural model for amyloid fibrils of human amylin also known as islet amyloid polypeptide (IAPP), formation of such structures has been correlated with non-insulin dependent (type 2) diabetes mellitus. The structure is viewed along the fibril axis. (Kajava et al. (2005) *J. Mol.Biol*). **C**. A molecular model of the compact, 4-protofilament insulin fibrils.

Most historic and current studies of protein folding has been concentrating on the "well-behaved" systems that can be efficiently unfolded and re-folded *in vitro* resulting in regeneration of the active proteins. Under normal cellular conditions, which feature high temperature and high solute concentration, some proteins fold into "wrong" conformation (usually aggregates). For example, when we boil an egg, the proteins in egg white unfold. But when the egg cools, these proteins do not return to their original shape. This is called ***misfolding***. Misfolding commonly results in exposure of hydrophobic surfaces that should have become buried in the protein interior. When concentration of denatured polypeptides is high, these surfaces can be buried through intermolecular interactions with other partially unfolded chains resulting in protein aggregation.

The term "***misfolding***" is used differently in different contexts. To avoid confusion here, we will use this term to describe ***processes that result in a protein acquiring a sufficient number of persistent non-native interactions to affect its overall architecture and/or its properties in a biologically significant manner***. A non aggregated partially

folded intermediate during the normal folding process will not be described as misfolded even if it may contain some non-native type contacts stabilizing its structure. However, when an intermediate assumes a compact stable structure with non-native topology, its architecture as well as its properties would be distinct from the native state. Such a species will be defined as misfolded. One of the most profound examples of misfolding is the aggregation process that results in incorporation of the protein (or peptide) into an amyloid fibril.

Amyloid fibrils (**Figure 7.16**) are highly organized structures (effectively one dimensional crystals) adopted by an unfolded polypeptide chain when it behaves as a typical polymer and hence the essential features of fibrils are determined by physicochemical properties of the chain; fibril structures can be formed by many types of synthetic polymers.

Propensity to form organized fibrils may be the reason why only a small fraction of the combinatorial space is explored by the known protein sequences. The sequences of modern proteins were selected in evolution to fold in the native structures suitable for performing a multitude of different tasks in the cell.

Both, folding and unfolding of proteins are essential processes in regulation of cellular activities, degradation of the proteins followed by subsequent recycling of the polypeptide building blocks, translocation of the proteins across cellular membranes, trafficking, secretion and even immune response. Consequently, failure of the protein to fold into native structure and remain folded, often gives rise to a wide range of the disease phenotypes in humans. Ever-increasing number of "folding diseases" that have been directly associated with deposition of protein aggregates in form of amyloid fibrils and plaques in tissues include Alzheimer's and Parkinson's diseases, spongiform encephalopathy (mad cow disease) and some forms of diabetes to name just a few.

Each amyloid disease involves aggregation of a specific protein. Although native structures of known proteins involved in characterized amyloidoses differ greatly from one another (from well-structured globular proteins of various α/β contents to largely unstructured peptide molecules), when these proteins assemble into the fibril, they all display similar characteristics (**Figure 7.16**). The fibrillar structures formed *ex vivo* are usually long, un-branched and often twisted; the core of the organized structure is composed of β-sheets having strands positioned perpendicular to the fibril axis. The portion of a polypeptide chain that is incorporated into fibril core may vary substantially for different proteins; in some cases only a handful residues may be involved in the core structure, with the remainder of the chain associated in some other manner with the fibrillar assembly.

Under appropriate conditions fibrillar structures self-assemble *in vitro* without any additional components or facilitators. Recent studies demonstrated that fibrils can be formed by many (if not all) polypeptides including proteins that have never been associated with disease or with function requiring a fibril formation. Synthetic homo-polymers (such as polythreonine or polylysine) can also assemble into fibrils. The cores of fibrils containing large proteins and those assembled from the small peptides have remarkably similar appearances. Thus, the ability of polypeptide chains to form fibrils is common and can be considered a generic feature of the polypeptides. Similar to other highly organized materials, structures of fibrils are stabilized by repetitive long-range interactions (*note: here long-range means interaction between groups spatially separated along the polypeptide chain and does not refer to the physical basis of the interaction*).

Hydrogen bonds between the groups on the polypeptide main chain are the main contributors to stabilization of the fibrils. In contrast, the side chains affect the details of the fibrillar assembly but not its general structure. Such generality of the structural organization profoundly contrasts the global structure of many natural proteins, where the interactions associated with the highly specific packing of the side chains can sometimes overwrite the main chain preferences. If the solution environment around a perfectly folded globular protein changes so that the interactions associated with the side chains are now insufficiently stable, the protein may unfold and then, under right circumstances, it may reassemble into the amyloid fibrils, where its newly acquired structure is defined by the main chain interactions. This notion can also explain why the fibrils formed by the peptides with completely different sequences are so similar in appearances.

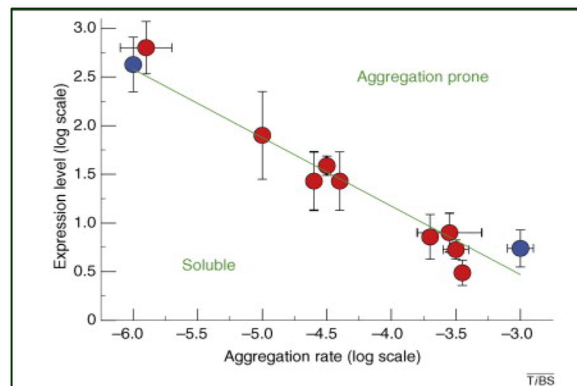


Figure 7.17 **Correlation between expression levels and the measured aggregation rates of the corresponding proteins.** Reference: Tartaglia et al 2007 TIBS.

Although the ability to form amyloid fibrils appears generic, the propensity to do so may vary greatly between proteins. One important determinant of the propensity to form fibrils is the polypeptide solubility. This is because aggregation is generally concentration dependent. The critical concentration of a particular polypeptide that is required to trigger (nucleate) aggregation depends primarily on its solubility and is different for different proteins. Since solubility may prevent toxic aggregation, why haven't proteins evolved to be more soluble? As we discussed earlier, the core of globular proteins is populated primarily by the hydrophobic side chains, which are excluded from the aqueous environment. Thus, increasing polypeptide solubility would be deleterious for its folding in the desired structure and therefore for its function. Thus, protein sequences have co-evolved with their cellular environments to be sufficiently soluble for their biological roles, but not more so. Indeed there is a strong anti-correlation between protein solubility and its cellular concentration (**Figure 7.17**). When proteins are expressed at higher concentrations than their native levels, or modified as result of environmental stress (oxidation damage, aging, excessive glycosylation, etc), their aggregation is inevitable.

The energy landscape for folding of such proteins requires more complex representation than a single-well funnel. Under physiological conditions when the intermolecular collisions are frequent, the off-pathway of aggregation should be taken into consideration along with the main funnel. The funnel shown in **Figure 7.14** depicted the folding behavior of an isolated polypeptide in dilute solution. Although sound for many well-behaved proteins, this depiction does not take into account the possibility of aggregation. To avoid erroneous description of the folding process within a cellular milieu, the folding funnel concept can be extended to include funnels with additional minima. **Figure 7.18** depicts such a funnel for a highly simplified folding/aggregation reaction. It assumes that a partially folded polypeptide chain forms about 50% of native-like contacts ("partially folded state"). From this state, the chain can proceed down the right funnel to form the remaining contacts and to collapse into the native state, or it may collide with another polypeptide chain in an identical conformation

and form additional stabilizing contacts between the two chains. The collision frequency between the chains defines the ratio of the partially folded aggregation prone configurations that will self associate to form the native conformation and those that will polymerize into the aggregates state.

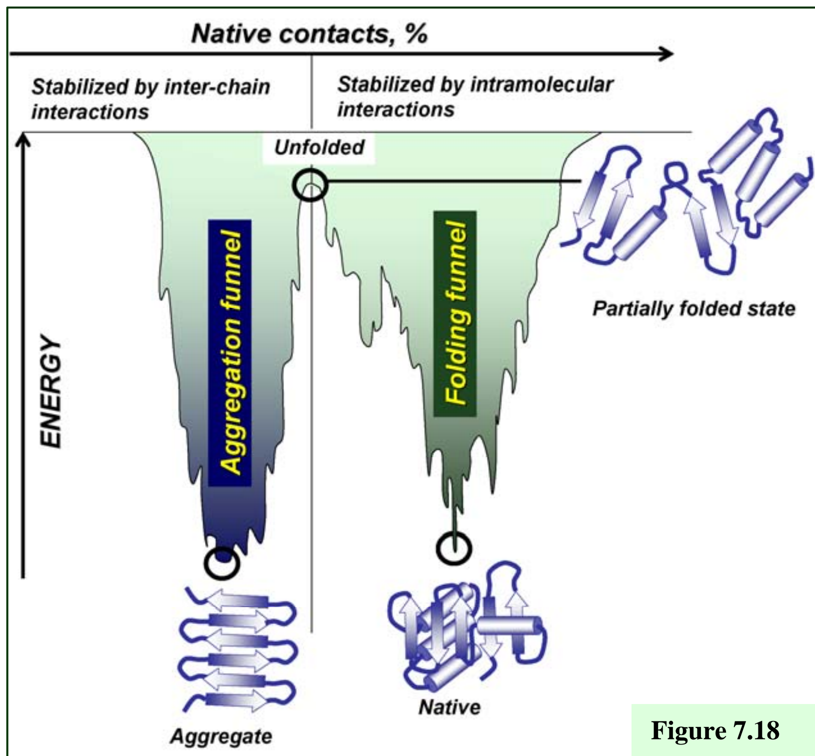


Figure 7.18

For a more realistic depiction, the system should be adjusted to allow formation of long fibrils. Additionally, aggregation pathway may occur through interaction between chains present in the non-identical conformations. The aggregation process will display a "nucleation-polymerization" behavior (**Figure 7.19**). As before, intermolecular collisions and therefore protein concentration will determine the choice between native pathway and formation of a dimer. Dimers will exist in equilibrium with trimers and higher aggregates until a nucleus of a critical size is

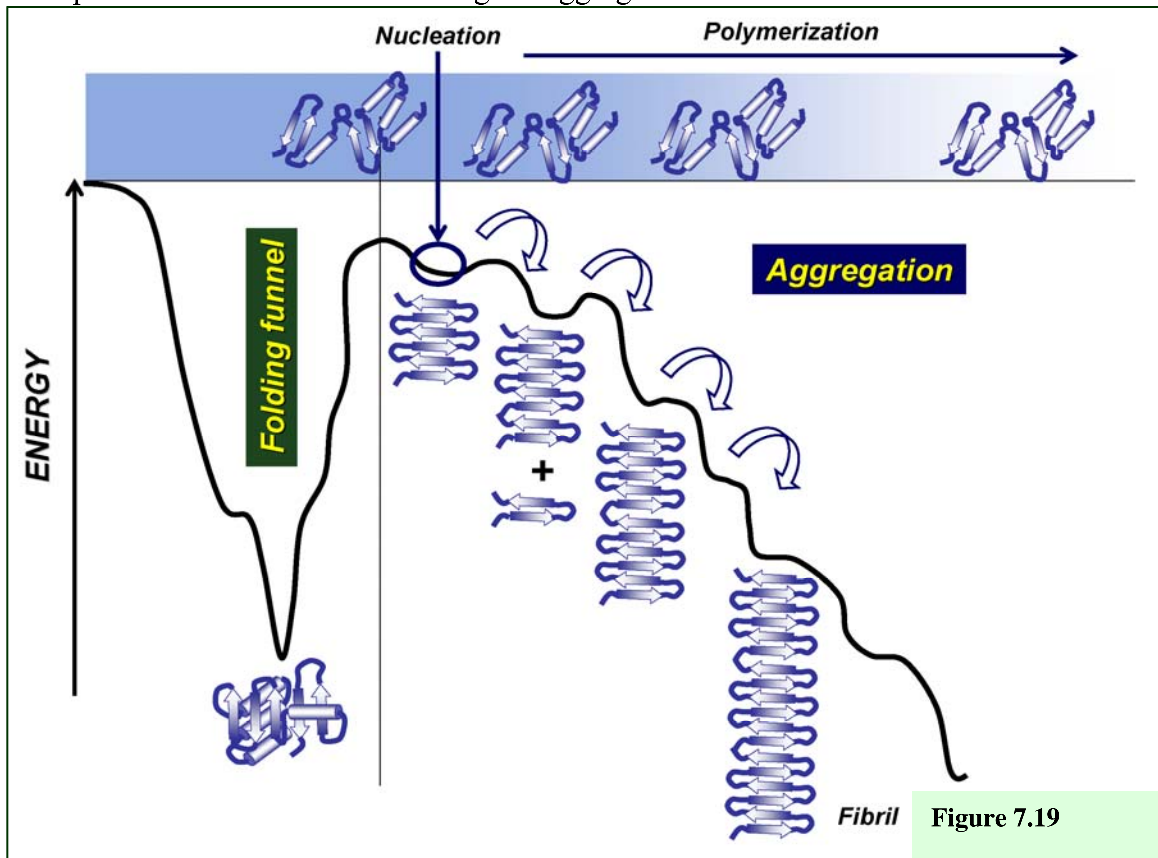


Figure 7.19

formed. Then addition of every subsequent monomer to the end of the growing fibril will bring the overall energy system down. If the concentration of the protein is unlimited, the fibril may grow indefinitely. Under limiting concentrations, fibril formation will proceed until the supply of partially unfolded conformers is depleted so that the protein concentration falls below the critical concentration that triggers aggregation. The number and the length of the fibrils, in this case, will depend on the relative rates of nucleation and polymerization.

CHAPERON ASSISTED FOLDING – CLIMBING OUT OF THE ALTERNATIVE MINIMA

The folding funnel formulation can be also useful in appreciation of the roles played by chaperones, ribosomes and other cellular factors that function to prevent polypeptide chains from populating conformations that self-associate into kinetically trapped states or aggregates *in vivo*.

While folding is viewed as a self-assembly process, in reality it is controlled by the environmental conditions, by the sequential protein synthesis on the ribosome, and in some cases by specific proteins called chaperons. Chaperons prevent random association of partially denatured proteins allowing them to assume proper active forms. They either isolate individual proteins so that their folding is not interrupted by interactions with other proteins (thus preventing aggregation) or help to unfold misfolded proteins, giving them a second chance to refold properly. An example of a chaperon, bacterial GroEL-GroES protein complex is shown in **Figure 7.20**. GroEL forms double heptameric ring that can bind ATP (molecular fuel driving thermodynamically unfavorable activities of various cellular machineries). ATP-binding is cooperative within each ring but displays high negative cooperativity between two rings. Thus, at any given time one of the rings has ATP bound and the other is empty. Binding of a 7 ATP molecules to the GroEL ring triggers conformational change that results in slight twist and tilt in the subunits and in exposure of hydrophobic patches that interact with and help to unfold misfolded protein. ATP-bound GroEL also gains high affinity for GroES, which acts as a "lid". Binding of GroES results in additional conformational change that removes hydrophobic surface of GroEL

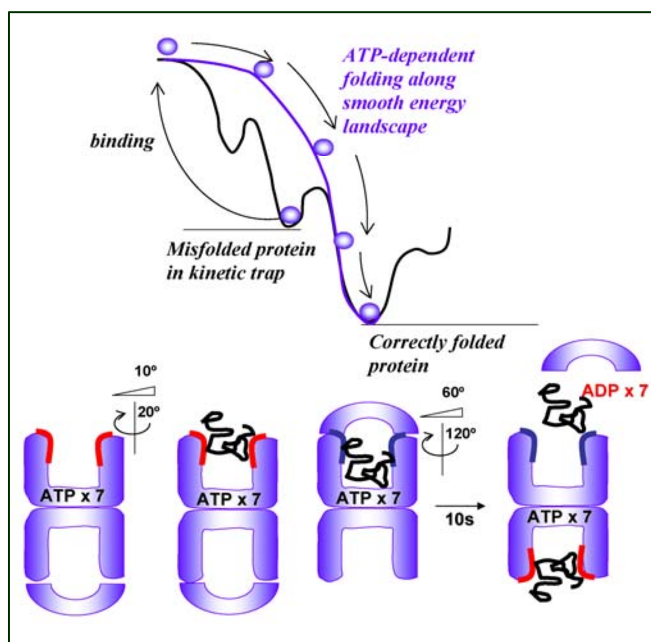


Figure 7.20 Structure and mechanism of bacterial chaperone. Unfolded (misfolded) polypeptide binds in a compact conformation to hydrophobic patches (red) on the apical domains (a) facing the cavity. GroES then binds to the GroEL ring that houses the polypeptide. GroES binding induces cooperative conformational changes in GroEL resulting in the burial of the hydrophobic binding regions of GroEL and to the formation of a large hydrophilic cage of volume ~ 175000 Å^3 where polypeptide substrate up to ~ 60 kDa can fold into native conformation.

and delivers substrate protein into the large hydrophilic cavity. The protein has about 10 s to refold. ATP hydrolysis trigger the next conformational change, release of the GroES lid, release of the protein (whether refolded or not), dissociation of ADP from the GroEL and binding of ATP into the “trans” GroEL ring.

The study of protein folding has been greatly advanced in recent years by the development of fast, time-resolved techniques. These are experimental methods for rapidly triggering the folding of a sample of unfolded protein, and then observing the resulting dynamics. Fast techniques in a widespread use include ultra-fast mixing of solutions, photochemical methods, laser temperature jump spectroscopy, and single-molecule measurements of the protein folding.

this section is for your information, not for testing

Secondary structure elements display characteristic spectroscopic properties.

CD:

Circular dichroism (CD) refers to the differential absorption of the two circularly polarized components of the plane polarized light: namely, L (rotating counter-clockwise) and R (rotating clockwise). If L and R components are absorbed to different extents, the resulting radiation is said to possess an elliptical polarization. A CD signal can be observed when a chromophore is chiral (optically active) for one of the following reasons: (a) it is intrinsically chiral because of its structure, (b) it is covalently linked to a chiral center in the molecule, or (c) it is placed in an asymmetric environment by virtue of the 3-dimensional structure adopted by the molecule.

CD signals only arise where absorption of radiation occurs. Thus spectral bands are easily assigned to distinct structural features of a molecule. The information on the secondary structure of the protein can be readily obtained from the far-UV region of the CD spectra (240 nm and below): absorption in this region is due principally to the peptide bond; there is a weak but broad $n \rightarrow \pi^*$ transition centered around 220 nm and a more intense $\pi \rightarrow \pi^*$ transition around 190 nm. The different types of regular secondary structure found in proteins give rise to characteristic CD spectra in the far UV. The spectrum of an actual protein can be de-convoluted into the component spectra representing α-helixes, β-sheets, and random coils. Therefore, one can estimate the secondary structure composition of the protein.

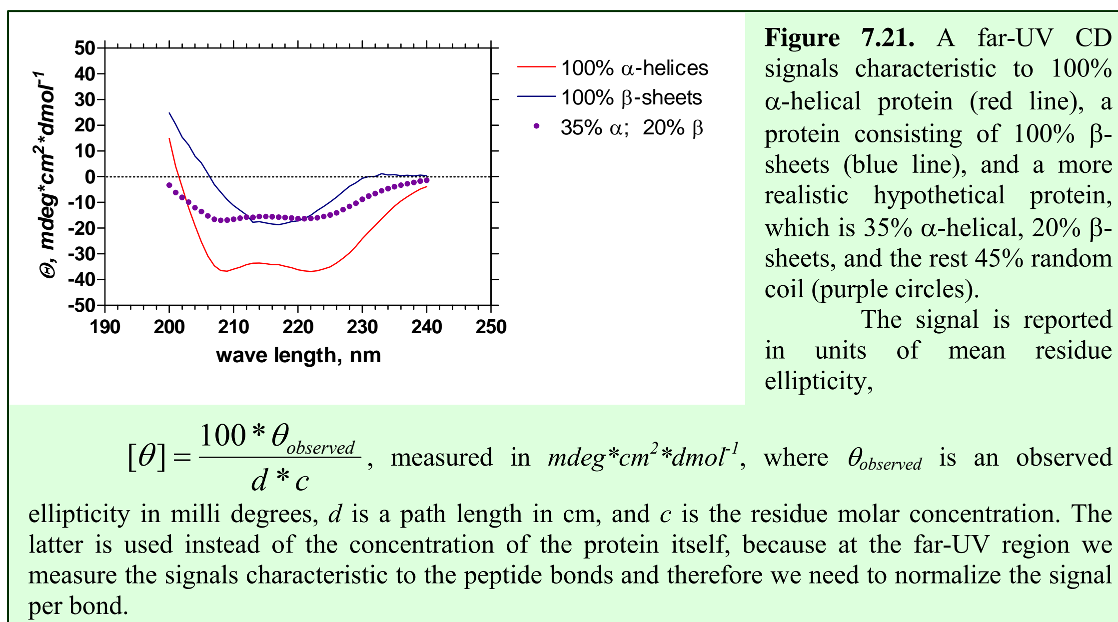


Figure 7.21. A far-UV CD signals characteristic to 100% α-helical protein (red line), a protein consisting of 100% β-sheets (blue line), and a more realistic hypothetical protein, which is 35% α-helical, 20% β-sheets, and the rest 45% random coil (purple circles).

The signal is reported in units of mean residue ellipticity,

$$[\theta] = \frac{100 * \theta_{observed}}{d * c}$$

, measured in $mdeg*cm^2*dmol^{-1}$, where $\theta_{observed}$ is an observed ellipticity in milli degrees, d is a path length in cm, and c is the residue molar concentration. The latter is used instead of the concentration of the protein itself, because at the far-UV region we measure the signals characteristic to the peptide bonds and therefore we need to normalize the signal per bond.

Figure 7.21 shows far-UV spectra characteristic to defined secondary structures. Typical α-helix gives a characteristic double minima spectrum (min at 208 and 222 nm), while a β-sheets give a spectrum with a single minimum at about 215 nm. Both types of

structure display characteristic positive ellipticity below 200 nm as well (not shown). However, obtaining a high signal to noise signal at such low wavelengths is challenging and usually is impossible without high intensity synchrotron radiation (especially below 190 nm).

Real proteins usually contain combinations of secondary structure elements (α-helices, β-sheets, random coils, and rarely polyproline II helices and other unconventional structures). The far-UV CD signal of a real protein can be deconvoluted using characteristic spectra for the secondary structure elements to access quantitatively the overall secondary structure content. Figure 11.8 shows an example of a spectra of a protein containing 35% α-helices and 20% β-sheets.

CD measurements in the far UV can give quantitative estimates of secondary structure which can be compared with those from X-ray crystallography or NMR. CD spectra in the different spectral regions are invaluable for assessing the structural relationships between native and recombinant protein, and between wild-type and mutant proteins. CD data can be used to confirm the integrity of expressed domains of a multi-domain protein, an essential prerequisite before detailed structural studies, e.g., by X-ray crystallography are undertaken. In addition the loss of CD signals either on addition of denaturing agents (such as urea or guanidinium hydrochloride) or by an increase in temperature can be used to provide quantitative estimates of the stability of the folded state of the native protein. CD can also been used to measure the rate of acquisition of secondary and tertiary structure. Continuous- or stopped-flow CD methods can be used to detect events happening on the ms time scale; such information allows exploring the mechanism of protein folding.

The main advantage of the CD techniques is that one can rapidly get estimations of the secondary structure content under a broad range of experimental conditions (much broader than FTIR). It is also much less labor-intensive than FTIR spectroscopy. Additionally, commonly used denaturing agents do not absorb in the far-UV region and thus can be used in the analysis of the protein stability.

There are, however, several limitations of this method:

1. Measurements can be obtained only in solution.
2. Accurate prediction of the secondary structure content requires high signal-to-noise ratio in the far-UV CD region. Conventional CD instruments allow obtaining spectra starting from about 180 nm. Below that, the radiation from the Xe lamp is insufficient and both N_2 gas used to purge air from the lamp and H_2O in the sample absorb the light. The highest quality spectra can be obtained using synchrotron radiation CD.
3. Although, the secondary structure composition can be obtained from the spectra recorded between 190 and 240 nm, it requires (i) the buffer and the ligand should be optically clear in the far-UV region; (ii) the concentration of the protein has to be accurately determined.
4. The rapid changes in the secondary structure can be obtained from monitoring the change in the CD signal at 220 nm; recording of the full spectra and therefore the complete information on the secondary structure content is slow.

FTIR

Molecular bonds vibrate at various frequencies depending on the elements and the type of bonds. For any given bond, there are several specific frequencies at which it can vibrate. According to quantum mechanics, these frequencies correspond to the ground state (lowest frequency) and several excited states (higher frequencies). One way to cause the frequency of a molecular vibration to increase is to excite the bond by having it absorb light energy. For any given transition between two states the light energy (determined by the wavelength) must exactly equal the difference in the energy between the two states, which for many bonds correspond to the IR region.

FTIR spectroscopy is another versatile method for monitoring structure of proteins. It stands for Fourier Transform Infra-Red, the preferred method of infrared spectroscopy. In the infrared spectroscopy, IR radiation is passed through a sample. Some of the infrared radiation is absorbed by the sample and some is transmitted. The resulting spectrum represents the molecular absorption and transmission, creating a molecular fingerprint of the sample.

One of the advantages of infrared spectroscopy is that it can be used with proteins that are either in solution or in the solid state. In addition, specialized infrared methods allow studying of membrane proteins and oriented samples, follow reaction kinetics and ligand binding in proteins, as well as protein dynamics.

Employing the interferometer, which simultaneously records all desired frequencies, and the Fourier Transform of the signal, result in the fast collection of spectra that provide signal from which the detailed information on the protein secondary structure can be readily extracted. Thus, the strongest point of the FTIR approach as a structure-specific probe is that a complete spectrum is available for each time point of measurement. In this way, several spectral windows are accessible simultaneously for the observation of the unfolding or the formation of different secondary structure elements and also events that can be attributed to changes in tertiary structure. Stopped-flow mixing coupled with time-resolved FTIR spectroscopy allows to monitor protein folding at the ms scale.

FTIR covers the frequency range from 4000 to 400 cm^{-1} (from 2.5 to 25μm). This wavelength region includes bands that arise from three conformationally sensitive vibrations within the peptide backbone (amide I, II and III). Of these vibrations, amide I is the most widely used and can provide information on secondary structure composition and structural stability. Of the nine vibrational amide bands of proteins, the most intense and most useful as a structural probe is the amide I band, which represents primarily the C=O stretching motion of the amide groups and occurs in the region 1600–1700 cm^{-1}. The amide I band contour is an established indicator of protein secondary structure because of its sensitivity to hydrogen-bonding pattern, dipole–dipole interaction, and the geometry of the polypeptide backbone. Typically, the amide I band of proteins consists of a series of overlapping component bands which occur as a result of the secondary structure present in such molecules, such as α-helices, β-sheets, turns, and irregular structures. Once the corresponding component bands can be identified, they provide a powerful probe of conformational changes in proteins. The α helical and irregular structures show bands very close together (1645–1660 cm^{-1}), which may cause complications in the analysis. On the other hand, amide groups in β-sheet structures give rise to major diagnostic bands between 1615 and 1640 cm^{-1} and weaker bands near 1670–1695 cm^{-1}. This makes time-resolved FTIR spectroscopy, in contrast to CD spectroscopy, very useful to indicate the presence of, and to monitor directly the kinetics of processes involving changes in, β-sheet structures.

Limitations of this method include:

1. Due to the H-O-H bending vibration, H_2O absorbs strongly near 1640 cm^{-1}. Therefore, to avoid total absorption of all light, one has to use a very short path length of the cell (6 – 8 μm) and to achieve a high signal to noise ratio, one has to use high protein concentration (≥ 10 mg/ml). Alternatively, one can replace H_2O with D_2O, which absorbs less light in the desired region due to the downshifted vibrations of heavier deuterium atom.

2. Commonly used chemical denaturants (such as urea and guanidine hydrochloride) produce strong IR bands of their own. Therefore, analysis of protein stability requires using isotopically labeled denaturants.

PROTEIN DYNAMICS (various spectroscopic techniques):

The tertiary structures of many proteins have been determined by x-ray crystallography or assigned by NMR. NMR also provides information regarding the flexibility of the protein regions and the permeability of a protein to solvent or other small molecules.

The vibrational motions with small angular rotations are amendable to NMR relaxation studies (for some side chains). Such studies indicate that surface side chains have considerable rotational freedom. Much less freedom, if any, is detectable for internal residues. These results are consistent with crystallographic studies, in which many surface side chains appear disordered, implying rotational motions.

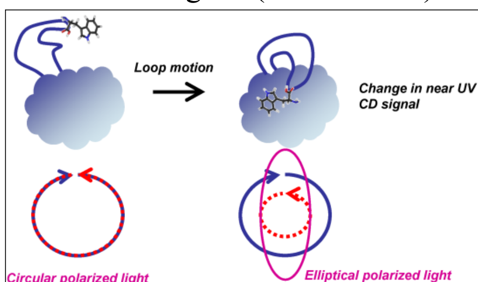
Often proteins undergo significant intermolecular motions of rearrangements as result of interaction with their substrates, co-factors, other partner macromolecules, or as result of changing an environment. A number of physical techniques are available to study such motions. Here are some commonly used examples:

Hydrogen Exchange Methods:

Hydrogen exchange (HX) methods in protein folding usually focus on the main chain amide hydrogen [-C(O)-N**H**-]. These hydrogen atoms are present at every amino acid except for proline and participate in hydrogen bonding interactions that stabilize secondary structure elements of the folded protein. When not involved in hydrogen bonding, the backbone (main chain) amide hydrogens can freely exchange with the hydrogens in solvent water. If instead of H_2O, the solvent is composed of "heavy water" (D_2O or T_2O), the hydrogens in the protein may exchange with deuterium or tritium from solution. Chemical factors that control HX rates for any given protein (such as pH, T, effects of neighboring side chains and isotope effects) have been accounted for and can be calculated. The rates of HX provide a measure of the fraction of folded protein and are usually measured by NMR spectroscopy or mass spectrometry. The HX rates for unprotected amide hydrogens at 0ºC and neutral pH are about 1 s^{-1} and are greatly slowed in folded proteins. This rate increases by an order of magnitude per pH unit or per 22ºC change in temperature. By varying simple experimental conditions one can adjust HX rates between ms and months.

Circular Dichroism Spectroscopy:

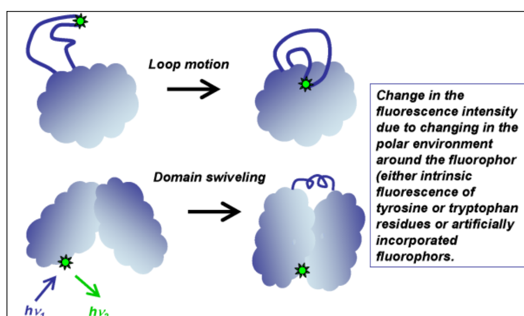
Aromatic side chains in proteins (tryptophan, tyrosine and phenylalanine) absorb light in a near-UV region (250-290 nm). Moreover, they differentially absorb the left and right components of circularly polarized light, which can be monitored by CD spectroscopy (usually at 295 nm) (left). The tertiary folding of the polypeptide can place these side chains in chiral environments, thus giving rise to CD spectra which can serve as characteristic fingerprints of the native structure.



The CD signal is very sensitive to the change in the polar environment of these side chains. Therefore, if the conformational change in protein results in the alteration of the environment of one or several of the aromatic side chains, such change can be readily detected. Moreover, relatively small conformational changes can be detected in this manner. The drawback of this approach is that it is not obvious which of the aromatic residues is responsible for the signal.

Fluorescence methods:

Besides CD signal, changes in the environment of aromatic side chains result in the change in their intrinsic fluorescence (left). ***Fluorescence*** is the process by which an electron that has been promoted to a higher energy level as result of the interaction with a photon of a particular wavelength returns to its ground state by dissipating the acquired energy via emission of a photon of lower energy. Excitation of the fluorophore occurs at a time scale of 10^{-15} s while energy dissipation through fluorescence takes about 10^{-8} s. Because this relaxation phenomenon is very sensitive to various environmental factors, it serves as a convenient probe for macromolecular dynamics and conformational transitions. Similar to CD experiments, it is necessary to assign the change in the fluorescence signal to a particular residue (when the signal is an intrinsic fluorescence of aromatic residues within the protein molecule). This is usually done by site-directed mutagenesis of the residue suspected to be responsible for the signal.

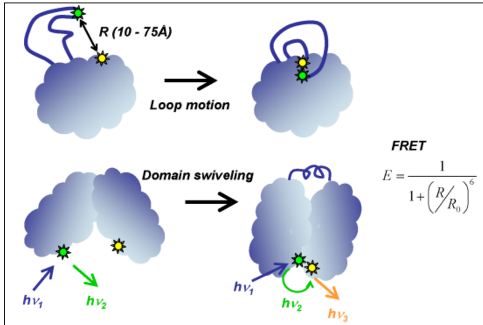


In contrast to CD, however, both excitation and emission spectra for tyrosine and tryptophan differ. Hence the input of the two types of aromatic residues can be distinguished.

Exogenous fluorescent label can be site-specifically incorporated into the protein of interest in a number of ways. For example, a protein can be expressed and purified with a fluorescent protein tag fused to it. A unique cysteine residue can be conjugated to a fluorescent dye of desired spectral properties using maleimide chemistry. The major drawback here is that this involves a laborious process of removing all solution exposed cysteine (if they are present), introducing a new cysteine in a desired location, and ensuring that the properties and activities of the protein were not affected by the mutagenesis and modification. The primary amines (lysine side chains and N-terminus of the polypeptide chain) are the other potential targets for modification. The protein can be

also expressed with certain amino acid sequence motifs amendable for modification with designer fluorescent labels.

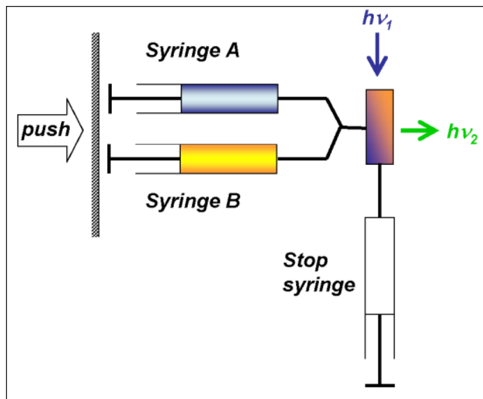
FRET: Conformational changes in a protein also can be measured by fluorescence resonance energy transfer or Föster resonance energy transfer (FRET). FRET is a powerful spectroscopic technique for measuring distances in the 10 – 75 Å range. It occurs when an excited fluorophore (donor) non-radiatively transfers energy to an acceptor fluorophore.



Excitation energy of the donor fluorophore can be transferred to the acceptor fluorophore (whose excitation spectrum should overlap with the emission spectrum of the donor) via an induced dipole – induced dipole interaction. The efficiency of FRET is $E = \frac{1}{1+\left(R/R_0\right)^6}$, where R is distance between two fluorophores and R_0 depends on the chemical structure of the fluorophore and is a distance at which 50% energy emitted by donor is transferred to the acceptor. Two fluorophores can be placed on the two parts of the protein molecule that are expected to come together. The FRET signal then will indicate whether they did.

FRET has been used very successfully as a probe for conformational dynamics in many bulk and single-molecule studies. The reason for this popularity lays in the inherent sensitivity of this method, on the steep dependence of the FRET efficiency on the distance between the two fluorophores within biologically relevant length scale. Additionally, relative photostability of modern commercially available fluorophores allows monitoring single-molecule FRET trajectories for extended periods of time.

Rapid mixing: If used under the equilibrium conditions, all of the techniques described above will provide only limited information about the protein flexibility, interactions and domain movements. These will include: (1) the observation that something happens and produces a measurable signal, and (2) in the case of binding, one can obtain the equilibrium binding constants for the labeled protein and its interacting partner or a substrate, which causes a conformational change.



The information on the protein dynamics cannot be obtained under the equilibrium conditions: the domain motions within the individual proteins comprising the ensemble will not be synchronized and effectively cancel each other. So, if we need to analyze the dynamics of protein motions in response to the external stimuli (such as addition of a substrate), we need to carry out the experiments under the pre-steady state conditions. Often the conformational changes in the protein happen very fast. The need to analyze such fast processes led to development of a number of rapid mixing techniques. All spectroscopic approaches described above can be combined with the stopped-flow apparatus to achieve the millisecond-range time intervals between mixing and beginning of observation.

Single-molecule techniques: Recently developed so-called single-molecule techniques provide a completely new view on many biological processes including protein folding, molecular associations, enzymatic catalysis and molecular motors. These new types of approaches provide access to stochastic events, transient intermediates and macromolecular dynamics not observable in more conventional methods. Among these techniques, single-molecule fluorescence spectroscopy and microscopy allow observation and analysis of the individual fluorescently labeled macromolecules or macromolecular complexes while single-molecule force spectroscopy (by AFM, optical and magnetic tweezers) probes energy landscapes of macromolecular folding, assembly, and catalytic properties of enzymes.

Computational approaches: Finally, advances in the modern computational techniques allow reasonable modeling of the macromolecular dynamics.