CHEMICALLY INDUCED NUCLEAR POLARIZATION

Chemically induced nuclear polarization (CIDNP) can be used to probe the solvent accessibility of certain aromatic residues in proteins. The reactive collision of polarizable amino acids such as tryptophan, tyrosine, and histidine with a photoexcited dye such as flavin mononucleotide (FMN) results in an electron transfer (in the case of Trp and Tyr) or proton transfer (His) reaction forming a pair of radicals. Electron-nuclear hyperfine interactions between the two radicals result in a significant enhancement of NMR signals. The ‘‘photosensitizer’’ flavin molecule can be excited by laser as light source. For the photoreaction to take place, the aromatic side chains must be accessible to the photosensitizer, e.g., located on the surface of the protein molecule. The CIDNP spectrum is recorded immediately after the laser flash and corrected by a ‘‘dark’’ spectrum recorded without irradiation. Besides the equilibrium studies of protein surfaces, the technique can be combined with a stopped-flow apparatus and in this way it can be used to study folding intermediates. Using CIDNP pulse-labeling technique, the exposed tryptophan and tyrosine residues in a molten globule state can be identified.

HIGH-PRESSURE NMR SPECTROSCOPY

When high pressure is applied to a protein solution, it shifts the conformational equilibrium of the protein molecules toward lower volume conformers, thereby decreasing the partial molar volume of the protein. The combination of high pressure with hetero nuclear two-dimensional NMR spectroscopy provides atomic resolution information on the structure of the protein molecule at different stages of the folding process. By varying the pressure, one can explore the conformational space from the folded to the unfolded conformer. In recent years, numerous studies using high-pressure NMR spectroscopy have Protein folding been carried out on locally disordered, molten globule unfolded as well as oligomeric or aggregated states of proteins.

PROTEIN FOLDING AND DYNAMICS STUDIED BY MASS SPECTROMETRY

Mass spectrometry of protein molecules has become a rapidly developing field in the last decade. In comparison with NMR spectroscopy, which provides site-specific information averaged in time, mass spectrometry is capable of detecting different conformers coexisting in the protein solution. This method is especially useful for the study of low- populated intermediate states and is free of the molecular size limitation of NMR spectroscopy. Because

of its high sensitivity, a protein concentration in the femtomolar range is sufficient for analysis. Structural and dynamic properties of various conformational states can be studied by hydrogen/deuterium exchange (HDX) combined with mass spectrometry. Recently,

Kaltashov and coworkers investigated the conformational ensemble of the molten globule state of ubiquitin. Using protein ion fragmentation in the gas phase, they evaluated the stability of various segments of the protein in the molten globular state. By the method of pulse-labeling HDX-MS, it is possible to study the kinetics of folding and to explore complex folding scenarios with parallel pathways. Co-populated protein conformers can be detected and characterized directly by electrospray ionization mass spectrometry (ESIMS). Protein surface areas in solution may be determined by ESIMS.

Limited proteolysis with ESIMS provides site-specific structural information on different conformational states of the protein molecules including protein aggregates and the amyloid state.

MECHANICAL UNFOLDING OF PROTEINS

In the first studies of the mechanical unfolding of single protein molecules using AFM, the giant sarcomeric protein titin, consisting of a large number of immunoglobulin segments, was used. Because of the heterogeneity of titin domains, it was not possible to assign the individual force peaks to specific domains. Using tandem repeats of a single domain, constructed by protein engineering techniques, it was possible to explain the mechanical characteristics of single domains in terms of their specific structures. Using force-measuring optical tweezers, it is possible to induce mechanical unfolding and refolding of individual molecules. In a recent work, Cecconi and coworkers showed that E. coli ribonuclease H molecule unfolds in a two-state manner and refolds through a transient molten globule-like intermediate. We may expect significant progress in the application of other new techniques such as the study of single-molecule folding kinetics by optical techniques in the near future.

MOLECULAR CHAPERONES

In molecular biology, **molecular chaperones** are proteins that assist the covalent folding or unfolding and the assembly or disassembly of other macromolecular structures. Chaperones are present when the macromolecules perform their normal biological functions and have correctly completed the processes of folding and/or assembly. The chaperones are concerned primarily with protein folding. The first protein to be called a chaperone assists the assembly of nucleosomes from folded histones and DNA and such assembly chaperones, especially in the nucleus, are concerned with the assembly of folded subunits into oligomeric structures.

One major function of chaperones is to prevent both newly synthesised polypeptide chains and assembled subunits from aggregating into nonfunctional structures. It is for this reason that many chaperones, but by no means all, are heat shock proteins because the

tendency to aggregate increases as proteins are denatured by stress. In this case, chaperones do not convey any additional stericinformation required for proteins to fold. However, some highly specific 'steric chaperones' do convey unique structural (steric) information onto proteins, which cannot be folded spontaneously. Such proteins violate Anfinsen's dogma.

Various approaches have been applied to study the structure, dynamics and functioning of chaperones. Bulk biochemical measurements have informed us on the protein folding efficiency, and prevention of aggregation when chaperones are present during protein folding. Recent advances in single-molecule analysis have brought insights into structural heterogeneity of chaperones, folding intermediates and affinity of chaperones for unstructured and structured protein chains.

Properties

- Molecular chaperones interact with unfolded or partially folded protein subunits, e.g. nascent chains emerging from the ribosome, or extended chains being translocated across subcellular membranes.
- They stabilize non-native conformation and facilitate correct folding of protein subunits.
- They do not interact with native proteins, nor do they form part of the final folded structures.
- Some chaperones are non-specific, and interact with a wide variety of polypeptide chains, but others are restricted to specific targets.
- They often couple ATP binding/hydrolysis to the folding process.
- Essential for viability, their expression is often increased by cellular stress.

Main role: They prevent inappropriate association or aggregation of exposed hydrophobic surfaces and direct their substrates into productive folding, transport or degradation pathways.

Location and Function

Many chaperones are heat shock proteins, that is, proteins expressed in response to elevated temperatures or other cellular stresses. The reason for this behaviour is thatprotein folding is severely affected by heat and, therefore, some chaperones act to prevent or correct damage caused by misfolding. Other chaperones are involved in folding newly made proteins as they are extruded from the ribosome. Although most newly synthesized proteins can fold in absence of chaperones, a minority strictly requires them for the same.

Some chaperone systems work as foldases: they support the folding of proteins in an ATP-dependent manner (for example, the GroEL/GroES or the DnaK/DnaJ/GrpE system). Other chaperones work as holdases: they bind folding intermediates to prevent their aggregation, for example DnaJ or Hsp33.

Macromolecular crowding may be important in chaperone function. The crowded environment of the cytosol can accelerate the folding process, since a compact folded protein will occupy less volume than an unfolded protein chain. However, crowding can reduce the yield of correctly folded protein by increasing protein aggregation. Crowding may also increase the effectiveness of the chaperone proteins such as GroEL, which could counteract this reduction in folding efficiency.

More information on the various types and mechanisms of a subset of chaperones that encapsulate their folding substrates (e.g. GroES) can be found in the chaperonins. Chaperonins are characterized by a stacked double-ring structure and are found in prokaryotes, in the cytosol of eukaryotes, and in mitochondria.

Other types of chaperones are involved in transport across membranes, for example membranes of the mitochondria and endoplasmic reticulum (ER) in eukaryotes. Bacterial translocation—specific chaperone maintains newly synthesized precursor polypeptide chains in a translocation-competent (generally unfolded) state and guides them to the translocon.

New functions for chaperones continue to be discovered, such as assistance in protein degradation, bacterial adhesin activity, and in responding to diseases linked to protein aggregation (e.g. see prion) and cancer maintenance.

CHEPARONINE

Chaperonins are proteins that provide favourable conditions for the correct folding of other proteins, thus preventing aggregation. Newly made proteins usually must fold from a linear chain of amino acids into a three-dimensional form. Chaperonins belong to a large class of molecules that assist protein folding, called molecular chaperones. The energy to fold proteins is supplied by adenosine triphosphate

GroupI Chaperonins

GroupI Chaperonins are found in bacteria as welas organelles of endosymbiotic origin: chloroplasts and mitochondria. The GroEL/GroES complex in *E. coli* is a Group I chaperonin and the best characterized large (~ 1 MDa) chaperonin complex.

1.GroEL is a double-ring 14mer with a greasy hydrophobic patch at its opening and can accommodate the native folding of substrates 15-60 kDa in size.

2.GroES is a single-ring heptamer that binds to GroEL in the presence of ATP or transition state analogues of ATP hydrolysis, such as ADP-AlF_3. It's like a cover that covers GroEL (box/bottle).

GroEL/GroES may not be able to undo protein aggregates, but kinetically it competes in the pathway of misfolding and aggregation, thereby preventing aggregate formation.

Group II Chaperonins

Group II chaperonins, found in the eukaryotic cytosol and in archaea, are more poorly characterized. TRiC (TCP-1 Ring Complex, also called CCT for chaperonin containing TCP-1), the eukaryotic chaperonin, is composed of two rings of eight different though related subunits, each thought to be represented once per eight-membered ring. TRiC was originally thought to fold only the cytoskeletal proteins actin and tubulin but is now known to fold dozens of substrates.

Mm cpn (Methanococcus maripaludis chaperonin), found in the archaea Methanococcus maripaludis, is composed of sixteen identical subunits (eight per ring). It has been shown to fold the mitochondrial protein rhodanese; however, no natural substrates have yet been identified.

Group II chaperonins are not thought to utilize a GroES-type cofactor to fold their substrates. They instead contain a "built-in" lid that closes in an ATP-dependent manner to encapsulate its substrates, a process that is required for optimal protein folding activity.

Mechanism of action

Chaperonins undergo large conformational changes during a folding reaction as a function of the enzymatic hydrolysis of ATP as well as binding of substrate proteins and cochaperonins, such as GroES. These conformational changes allow the chaperonin to bind an unfolded or misfolded protein, encapsulate that protein within one of the cavities formed by the two rings, and release the protein back into solution. Upon release, the substrate protein will either be folded or will require further rounds of folding, in which case it can again be bound by a chaperonin.

The exact mechanism by which chaperonins facilitate folding of substrate proteins is unknown. According to recent analyses by different experimental techniques, GroEL-bound substrate proteins populate an ensemble of compact and locally expanded states that lack stable tertiary interactions. A number of models of chaperonin action have been proposed, which generally focus on two (not mutually exclusive) roles of chaperonin interior: passive and active. Passive models treat the chaperonin cage as an inert form, exerting influence by reducing the conformational space accessible to a protein substrate or preventing intermolecular interactions e.g. by aggregation prevention. The active chaperonin role is in turn involved with specific chaperonin–substrate interactions that may be coupled to conformational rearrangements of the chaperonin.

Probably the most popular model of the chaperonin active role is the iterative annealing mechanism (IAM), which focus on the effect of iterative, and hydrophobic in nature, binding of the protein substrate to the chaperonin. According to computational simulation studies, the IAM leads to more productive folding by unfolding the substrate from misfolded conformations or by prevention from protein misfolding through changing the folding pathway.

HUMAN CHAPERONE PROTEINS

Chaperones are found in, for example, the endoplasmic reticulum (ER), since protein synthesis often occurs in this area.

Endoplasmic reticulum

In the endoplasmic reticulum (ER) there are general, lectin- and non-classical molecular chaperones helping to fold proteins.

- General chaperones: GRP78/BiP, GRP94, GRP170.
- Lectin chaperones: calnexin and calreticulin
- Non-classical molecular chaperones: HSP47 and ERp29
- Folding chaperones:
  - Protein disulfide isomerase (PDI),
  - *Peptidyl prolyl cis-trans-isomerase* (PPI)
  - ERp57

Nomenclature and examples of bacterial and archael chaperons.

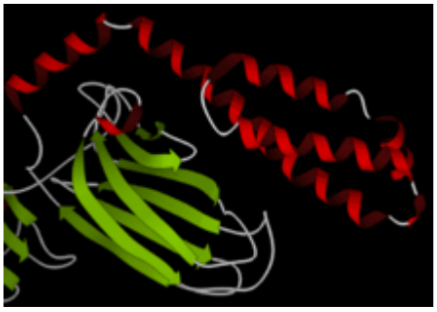
There are many different families of chaperones; each family acts to aid protein folding in a different way. In bacteria like *E. coli*, many of these proteins are highly expressed under conditions of high stress, for example, when the bacterium is placed in high temperatures. For this reason, the term "heat shock protein" has historically been used to name these chaperones. The prefix "Hsp" designates that the protein is a heat shock protein.

Hsp60

Hsp60 (GroEL/GroES complex in *E. coli*) is the best characterized large (~ 1 MDa) chaperone complex. GroEL is a double-ring 14mer with a hydrophobic patch at its opening; it is so large it can accommodate native folding of 54-kDa GFP in its lumen. GroES is a single-ring heptamer that binds to GroEL in the presence of ATP or ADP. GroEL/GroES may not be able to undo previous aggregation, but it does compete in the pathway of misfolding and aggregation.[19] Also acts in mitochondrial matrix as molecular chaperone.

Hsp70

Hsp70 (DnaK in *E. coli*) is perhaps the best characterized small (~ 70 kDa) chaperone.



The Hsp70 proteins are aided by Hsp40 proteins (DnaJ in *E. coli*), which increase the ATP consumption rate and activity of the Hsp70s.

It has been noted that increased expression of Hsp70 proteins in the cell results in a decreased tendency toward apoptosis.Although a precise mechanistic understanding

has yet to be determined, it is known that Hsp70s have a high-affinity bound state to unfolded proteins when bound to ADP, and a low-affinity state when bound to ATP. It is thought that many Hsp70s crowd around an unfolded substrate, stabilizing it and preventing aggregation until the unfolded molecule folds properly, at which time the Hsp70s lose affinity for the molecule and diffuse away. Hsp70 also acts as a mitochondrial and chloroplastic molecular chaperone in eukaryotes.

Hsp90

Hsp90 (HtpG in *E. coli*) may be the least understood chaperone. Its molecular weight is about 90 kDa, and it is necessary for viability in eukaryotes (possibly for prokaryotes as well).Heat shock protein 90 (Hsp90) is a molecular chaperone essential for activating many signaling proteins in the eukaryotic cell.Each Hsp90 has an ATP-binding domain, a middle domain, and a dimerization domain.

Hsp100

Hsp100 (Clp family in *E. coli*) proteins have been studied *in vivo* and *in vitro* for their ability to target and unfold tagged and mis folded proteins. Proteins in the Hsp100/Clp family form large hexameric structures with unfoldase activity in the presence of ATP. These proteins are thought to function as chaperones by processively threading client proteins through a small 20 Å (2 nm) pore, thereby giving each client protein a second chance to fold. Some of these Hsp100 chaperones, like ClpA and ClpX, associate with the double-ringed tetradecameric serine protease ClpP; instead of catalyzing the refolding of client proteins, these complexes are responsible for the targeted destruction of tagged and misfolded proteins. Hsp104, the Hsp100 of Saccharomyces cerevisiae, is essential for the propagation of many yeast prions. Deletion of the HSP104 gene results in cells that are unable to propagate certain prions.

PROTEOSOME MEDIATED PROTEIN DEGRADATION

The ubiquitin/proteasome system (UPS) is the main eukaryotic cytosolic and nuclear proteolytic pathway serving for selective degradation of cellular proteins. By influencing protein abundance, the proteasome contributes to the dynamic state of cells, which allows a tight control of many biochemical pathways and cellular responses upon changes of the environment.

Ubiquitination and targeting

Proteins are targeted for degradation by the proteasome with covalent modification of a lysine residue that requires the coordinated reactions of three enzymes. In the first step, a ubiquitin-activating enzyme (known as E1) hydrolyzes ATP and adenylylates a ubiquitin molecule. This is then transferred to E1's active-site cysteine residue in concert with the adenylylation of a second ubiquitin.This adenylylated ubiquitin is then transferred to a cysteine of a second enzyme, ubiquitin-conjugating enzyme (E2). In the last step, a member

of a highly diverse class of enzymes known as ubiquitin ligases (E3) recognizes the specific protein to be ubiquitinated and catalyzes the transfer of ubiquitin from E2 to this target protein. A target protein must be labeled with at least four ubiquitin monomers (in the form of a polyubiquitin chain) before it is recognized by the proteasome lid.It is therefore the E3 that confers substratespecificity to this system. The number of E1, E2, and E3 proteins expressed depends on the organism and cell type, but there are many different E3 enzymes present in humans, indicating that there is a huge number of targets for the ubiquitin proteasome system.

The mechanism by which a polyubiquitinated protein is targeted to the proteasome is not fully understood. Ubiquitin-receptor proteins have an N-terminal ubiquitin-like (UBL) domain and one or more ubiquitin-associated (UBA) domains. The UBL domains are recognized by the 19S proteasome caps and the UBA domains bind ubiquitin via three-helix bundles. These receptor proteins may escort polyubiquitinated proteins to the proteasome, though the specifics of this interaction and its regulation are unclear.

The ubiquitin protein itself is 76 amino acids long and was named due to its ubiquitous nature, as it has a highly conserved sequence and is found in all known eukaryotic organisms.The genes encoding ubiquitin in eukaryotes are arranged in tandem repeats, possibly due to the heavy transcription demands on these genes to produce enough ubiquitin for the cell. It has been proposed that ubiquitin is the slowest-evolving protein identified to date. Ubiquitin contains seven lysine residues to which another ubiquitin can be ligated, resulting in different types of polyubiquitin chains. Chains in which each additional ubiquitin is linked to lysine 48 of the previous ubiquitin have a role in proteasome targeting, while other types of chains may be involved in other processes.

Unfolding and translocation

After a protein has been ubiquitinated, it is recognized by the 19S regulatory particle in an ATP-dependent binding step. The substrate protein must then enter the interior of the 20S particle to come in contact with the proteolytic active sites. Because the 20S particle's central channel is narrow and gated by the N-terminal tails of the α ring subunits, the substrates must be at least partially unfolded before they enter the core. The passage of the unfolded substrate into the core is called *translocation* and necessarily occurs after deubiquitination. However, the order in which substrates are deubiquitinated and unfolded is not yet clear. Which of these processes is the rate-limiting step in the overall proteolysis reaction depends on the specific substrate; for some proteins, the unfolding process is rate-limiting, while deubiquitination is the slowest step for other proteins. The extent to which substrates must be unfolded before translocation is not known, but substantial tertiary structure, and in particular nonlocal interactions such as disulfide bonds, are sufficient to inhibit degradation.

The gate formed by the α subunits prevents peptides longer than about four residues from entering the interior of the 20S particle. The ATP molecules bound before the initial

recognition step are hydrolyzed before translocation. While energy is needed for substrate unfolding, it is not required for translocation.The assembled 26S proteasome can degrade unfolded proteins in the presence of a non-hydrolyzable ATP analog, but cannot degrade folded proteins, indicating that energy from ATP hydrolysis is used for substrate unfolding. Passage of the unfolded substrate through the opened gate occurs via facilitated diffusion if the 19S cap is in the ATP-bound state.

The mechanism for unfolding of globular proteins is necessarily general, but somewhat dependent on the amino acid sequence. Long sequences of alternating glycine andalanine have been shown to inhibit substrate unfolding, decreasing the efficiency of proteasomal degradation; this results in the release of partially degraded byproducts, possibly due to the decoupling of the ATP hydrolysis and unfolding steps. Such glycine-alanine repeats are also found in nature, for example in silk fibroin; in particular, certain Epstein-Barr virus gene products bearing this sequence can stall the proteasome, helping the virus propagate by preventing antigen presentation on the major histocompatibility complex.

Proteolysis

The mechanism of proteolysis by the β subunits of the 20S core particle is through a threonine-dependent nucleophilic attack. This mechanism may depend on an associated water molecule for deprotonation of the reactive threonine hydroxyl. Degradation occurs within the central chamber formed by the association of the two β rings and normally does not release partially degraded products, instead reducing the substrate to short polypeptides typically 7–9 residues long, though they can range from 4 to 25 residues, depending on the organism and substrate. The biochemical mechanism that determines product length is not fully characterized. Although the three catalytic β subunits have a common mechanism, they have slightly different substrate specificities, which are considered chymotrypsin-like, trypsin-like, and peptidyl-glutamyl peptide-hydrolyzing (PHGH)-like. These variations in specificity are the result of interatomic contacts with local residues near the active sites of each subunit. Each catalytic β subunit also possesses a conserved lysine residue required for proteolysis.

Although the proteasome normally produces very short peptide fragments, in some cases these products are themselves biologically active and functional molecules. Certain transcription factors regulating the expression of specific genes, including one component of the mammalian complex NF-κB, are synthesized as inactive precursors whose ubiquitination and subsequent proteasomal degradation converts them to an active form. Such activity requires the proteasome to cleave the substrate protein internally, rather than processively degrading it from one terminus. It has been suggested that long loops on these proteins' surfaces serve as the proteasomal substrates and enter the central cavity, while the majority of the protein remains outside. Similar effects have been observed in yeast proteins; this mechanism of selective degradation is known as *regulated ubiquitin/proteasome dependent processing* (RUP).

Ubiquitin-independent degradation

Although most proteasomal substrates must be ubiquitinated before being degraded, there are some exceptions to this general rule, especially when the proteasome plays a normal role in the post-translational processing of the protein. The proteasomal activation of NF-κB by processing p105 into p50 via internal proteolysis is one major example. Some proteins that are hypothesized to be unstable due tointrinsically unstructured regions, are degraded in a ubiquitin-independent manner. The most well-known example of a ubiquitin-independent proteasome substrate is the enzyme ornithine decarboxylase. Ubiquitin-independent mechanisms targeting key cell cycle regulators such as p53 have also been reported, although p53 is also subject to ubiquitin-dependent degradation. Finally, structurally abnormal, misfolded, or highly oxidized proteins are also subject to ubiquitin-independent and 19S-independent degradation under conditions of cellular stress.

Protein folding errors

Proteins can miss function for several reasons. When a protein is miss folded it can lead to denaturation of the protein. Denaturation is the loss of protein structure and function. The miss folding does not always lead to complete lack of function but only partial loss of functionality. The miss functioning of proteins can sometimes lead to diseases in the human body.

Alzheimer's disease

Alzheimer's disease (AD) is a neurological degenerative disease that affects around 5 million Americans, including nearly half of those who are age 85 or older. The predominant risk factors of AD are age, family history, and heredity. Alzheimer’s disease typically results in memory loss, confusion of time and place, misplacing places, and changes in mood and behavior. AD results in dense plaques in the brain that are comprised of fibrillar β-amyloid proteins with a well-orders β-sheet secondary structure. These plaques visually look like voids in the brain figure matter and are directly connected to the deterioration of thought processes. It has been determined that AD is a protein misfolding disease, where the misfolded protein is directly related to the formation of these plaques in the brain.

Mad Cow

Diseases caused by prions, like Mad Cow / Creutzfeldt-Jacob are also, in essence, protein folding disorders. These are caused by a certain protein, named PrP, that will stay in a mis-folded conformation (PrPsc) if encouraged to go into it in the first place. In most people, the PrP protein folds normally, leaving the person healthy. Rarely, a mutation in the PrP gene will allow the protein to be made incorrectly, and it will fold incorrectly, making a PrPsc prion. These prions, when exposed to PrP which is in the process of folding, will encourage that PrP to fold badly too, thus creating another PrPsc. While PrP can be processed and cleaned out of a cell once it has been used, PrPsc is shaped differently enough that it can't be, so it never goes away. PrPsc, much more quickly than with Ab in Alzheimer's, builds up

into plaques, handily destroying whatever nervous tissue it's building up in. See the writeups under prion for more on this.

Cystic Fibrosis

Besides building up un-processable plaques, protein folding errors can leave behind too little of the effective conformation for it to do its job. This is the case with diseases like Cystic Fibrosis, and many other hereditary diseases. Cystic Fibrosis results from lack of a protein that regulates chloride ion transport through a cell membrane. Findings show that while this protein seems to be forming correctly, there is a problem with one of its associated chaperone proteins. Chaperone proteins help encourage unfolded proteins to fold in the right way by surrounding them and protecting their movement. In Cystic Fibrosis, the chaperone doesn't pull away from the transport protein smoothly, leaving it partially mis-folded and useless. The broken chaperone protein then moves on to do the same thing to another transport protein, and so forth.

Protein stability

Protein stability is another common problem in protein expression. It is also an important topic in purification, formulation, and storage. Here we will discuss about protein stability in expression only. Properly folded proteins are usually stable during expression and purification. Sufficient amount of intact protein should be obtained. However some proteins appear to be unstable during expression and purification. Some of them are so unstable that sufficient amount of protein cannot be obtained. Many factors such as amino acid sequence of the protein, protein construction, host cell strain, expression and purification conditions may affect protein stability.

Amino acid sequence of a protein itself may be susceptible to degradation. Certain amino acids at the N-terminus of a protein can lead the protein to degradation. These are Arg, Lys, Leu, Phe, Tyr, and Trp residues. Replacing these amino acids with others can greatly increase the protein half-life (N-end rule). Many recombinant proteins are expressed with tags or fusion partners. Amino acid sequences at N-termini of these tags and fusion partners are often optimized for protein yield and stability. Therefore amino acids at N-terminus are not a problem in protein stability for these tagged or fusion proteins.

It is reported that regions containing Pro (P), Glu (E), Ser (S), and Thr (T) termed PEST are prone to degradation. It is generally observed that flexible hydrophilic sequences with protease cleavage sites are easily degraded. These sequences may be integral part of a protein. In most cases these sequences cannot be deleted or mutated. Strategies for improving protein stability are needed for these proteins.

Strategies to improve protein Stability:

- Perform expression in special media containing trace metals, minerals, and vitamins. These chemicals may not be needed for host cell growth, but they may serve as co-factor, prosthetic groups or ligands for recombinant proteins. Therefore they may be critical for correct protein folding and stability. Medium pH should also be balanced near neutral to improve protein stability. There will be no protein degradation caused by nutrition

exhaustion in our special media.

- Induce the protein at lower temperature and/or for shorter induction time.
- Fuse the protein with a tag or fusion partner. A tag can change the N-terminal sequence of the protein and therefore increase the yield and stability. In addition to N-terminal sequence change, a relative large fusion partner can further stabilize the protein compared with that the protein was expressed alone or with a small tag.
- Design the protein construct with intact domain or structure. A full-length protein, a part of a protein with intact domains, or an intact domain of a protein can be stably expressed. An integral folding unit of a protein cannot be truncated. Otherwise it may not be structured and therefore will not be stable. A domain is often a folding unit. Truncations at either terminus may disrupt protein folding. A domain of a protein can be determined by homologous alignment of the protein with other proteins. The more proteins are used in the alignment, the more accurate boundaries of a domain can be determined. Including more amino acids at the boundaries is often better than trimming off some residues. If an intact domain has to be truncated, fusing it with a large partner such as GST will make it more stable. Small tags such as his-tag will not improve protein stability in these cases.
- Change the host cell strain. Some cell strains are deficient in some proteases. For example, BL21 lacks cytoplasmic ion and periplasmic ompT proteases. Using these cell strains will lead to enhanced protein stability. Sometimes simply changing a host strain will increase recombinant protein stability.
- Change the location of expression. Some proteins are not stable if they are expressed in cytoplasm. When they are expressed in the periplasmic region, it becomes stable. Periplasmic expression may lead to correct folding of a protein. Periplasmic region may also lack the protease to degrade the protein.
- Express the protein in cell strains containing molecular chaperones. Molecular chaperones may facilitate protein folding and increase it stability

It is clear that strategies such as using special media, cell strains or growth conditions are easy to implement. Making fusion protein, designing protein construct, and changing expression location will involve DNA manipulation. In the case that a truncated protein domain is unstable, the choices may be to re-design the construct and express the protein as an intact domain or to fuse it with large protein partner.

Misfolding and cancer

Whereas too much of an incorrectly folded protein can cause amyloidoses, another group of protein folding diseases is caused by lack of a correctly folded protein. This form of protein folding defect is thought to be involved in diseases such as cystic fibrosis, but mainly affects a protein called p53, which occupies the most important position in the bodys cancer resistance network. Normally, the p53 system is switched off or, at most, is in stand-by mode. It is activated inside a cell if the cell becomes excessively stressed or damaged, which can lead to genetic mutations in DNA that can cause the uncontrolled division and proliferation of cells that is the hallmark of tumour formation. p53 is so good at its job that even a single break in the DNA strand is enough to activate it. It rushes into the cell nucleus and induces the production of other proteins that stop uncontrolled cell division or trigger the programmed death of a cell.

This tumour-suppressing function of p53 is so important that the protein has been described as the guardian of the genome. So, it is no surprise that faults in the p53 gene can be disastrous. Even a mutation in one of the letters (nucleotides) in the gene can be enough to lead to the expression of p53 proteins that do not fold correctly. Half crippled, they cannot carry out their job properly, so the damage to the DNA that would normally be repaired goes unnoticed, allowing the abnormal cell to grow in an uncontrolled manner. This type of mutation in p53 is thought to occur in 50% of all cases of cancer and as many as 95% of all cases of lung cancer.

Priondiseases

Transmissable spongiform encephalopathies (TSEs), which include mad cow disease (bovine spongiform encephalopathy; BSE) and Creutzfeld﷿Jakob disease (CJD) in humans, are special forms of amyloidosis in which the victim﷿s brain degenerates to a structure that looks like a porous sponge. These conditions seem to occur when normal human protein particles called prions misfold. The normal human prion is a component of the membrane of healthy nerve cells (called PrP^{c}), which folds properly, remains soluble and is disposed of without problem. It can, however, misfold in a particular way, which allows it to take on an infectious, incorrectly folded three-dimensional form (called PrP^{sc}), presumably due to a genetic mutation. The infectious prion, which can be transmitted in the diet, triggers a domino effect in healthy prions, forcing them to adopt its incorrectly folded form.

References

1. Libre-texts chemistry - https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Book%3A_Basic_Principles_of_Organic_Chemistry_(Roberts_and_Caserio)/27%3A_More_about_Spectroscopy/27.05%3A_Chemically_Induced_Dynamic_Nuclear_Polarization_(CIDNP)
2. Wikipedia - https://en.wikipedia.org/wiki/CIDNP