PREPARATION AND CHARACTERIZATION OF BIOLOGICAL SAMPLES FOR STRUCTURAL STUDIES

SOLUBILITY, STABILITY AND AGGREGATION

Marc Ruff

Integrated structural biology department
Chromatin stability and DNA mobility team

IGBMC, Illkirch, France

ruff@igbmc.fr

Preparation of biological sample

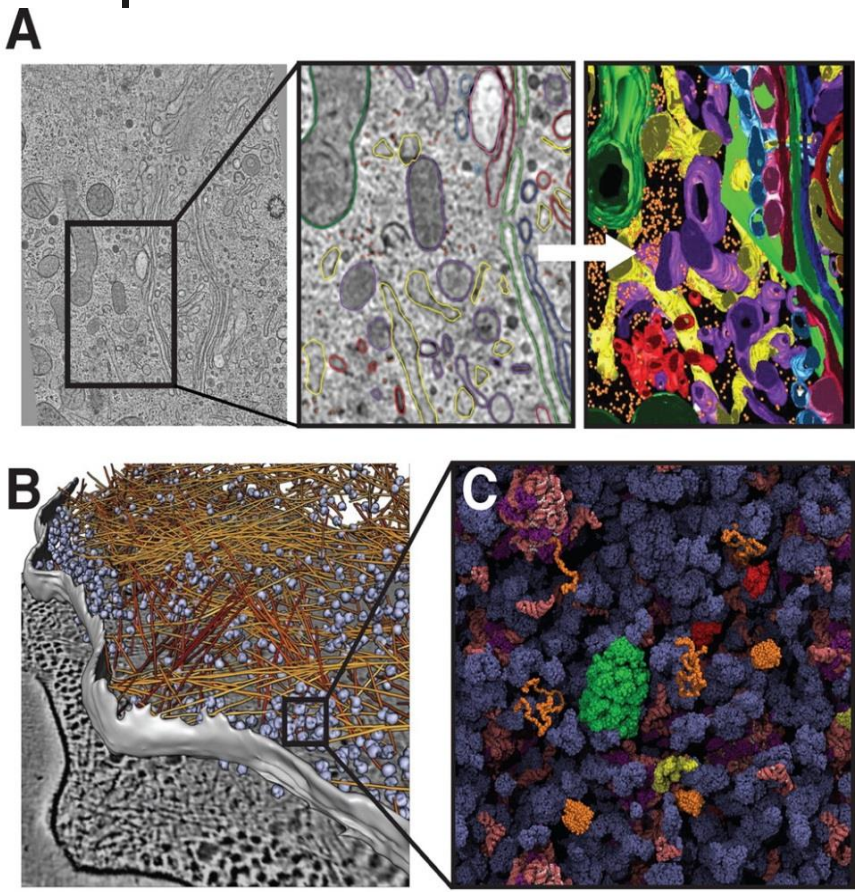
- Extract a single, pure, homogenous, soluble and functional protein / protein complex from a high complex mixture

“Cellular crowding”

Intracellular medium: complex mixture

- Peptides
  - Amino acids chains (< 6 KD)
- Proteins
  - Biological macromolecules (> 6 KD)
  - Nucleic acids
  - Chains of nucleotide triphosphate
- Polysaccharides
- Lipids
- Small molecules
- .....

Intracellular complexity



The intracellular environment is extremely crowded. Estimates show that the concentration of biological macromolecules (proteins, nucleic acids, ribonucleoproteins, polysaccharides, etc.) inside cells is in the range of 80–400 mg/mL. This corresponds to a volume occupancy of 5%–40% and creates a crowded medium, with considerably restricted amounts of free water. Such natural intracellular media, being filled with billions of protein molecules and a myriad of DNA, RNA, and polysaccharide molecules are known as "crowded" rather than "concentrated" environments, as, in general, no individual macromolecular species may be present at high concentration.

Francois-Xavier Theillet; Andres Binolfi; Tamara Frembgen-Kesner; Karan Hingorani; Mohona Sarkar; Ciara Kyne; Conggang Li; Peter B. Crowley; Lila Gierasch; Gary J. Pielak; Adrian H. Elcock; Anne Gershenson; Philipp Selenko; *Chem. Rev.* 2014, 114, 6661-6714

Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140

Macromolecular crowding

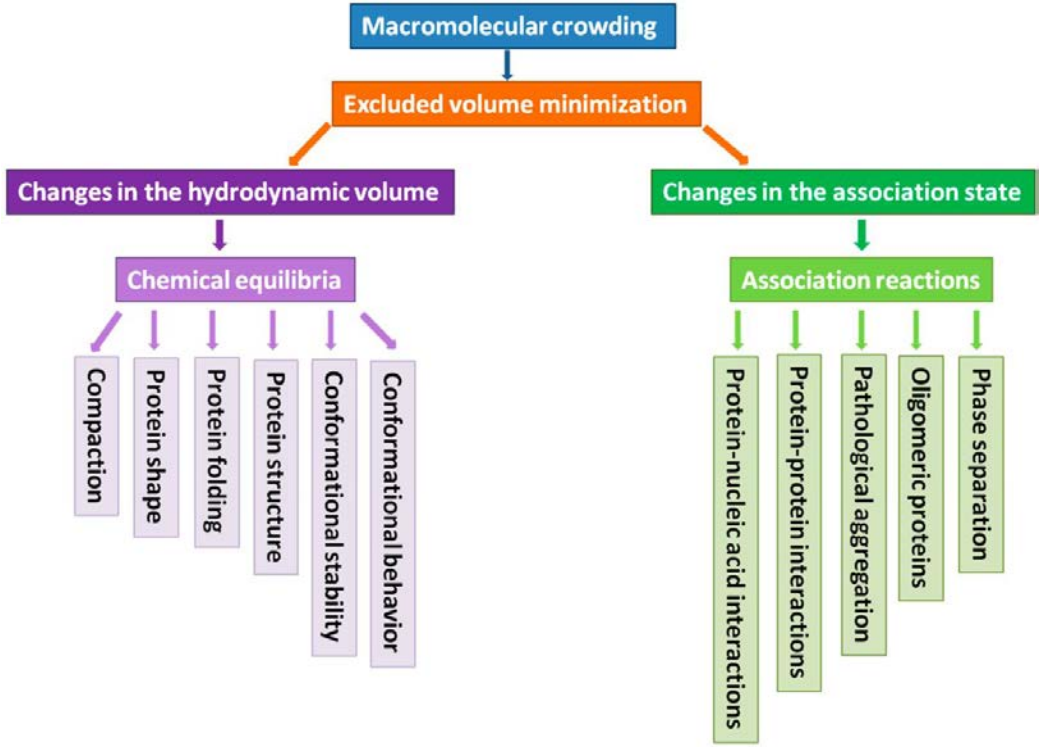
The average spacing between macromolecules in such crowded milieu can be much smaller than the size of the macromolecules themselves. Furthermore, the volume occupied by solutes is unavailable to other molecules because two molecules cannot be in the same place at the same time. As a result, any reactions that depend on available volume can be affected by macromolecular crowding effects.

The thermodynamic consequences of the unavailable volume are called excluded volume effects

Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140

Macromolecular crowding

the fact that two molecules cannot occupy the same space in solution, and that steric hindrance or impediment of a macromolecule is expected to exclude other molecules from its neighborhood give rise to the excluded volume phenomenon



Schematic representation of the potential effects of excluded volume on the behavior of proteins in crowded milieu

Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140

Solubility, Aggregation, Stability, Function

- Stability
  - Folded vs. unfolded state
- Aggregation
  - Protein multimerization
- Solubility
  - Protein dissolved in aqueous solvent
- Function
  - Biological function of the protein

Protein stability

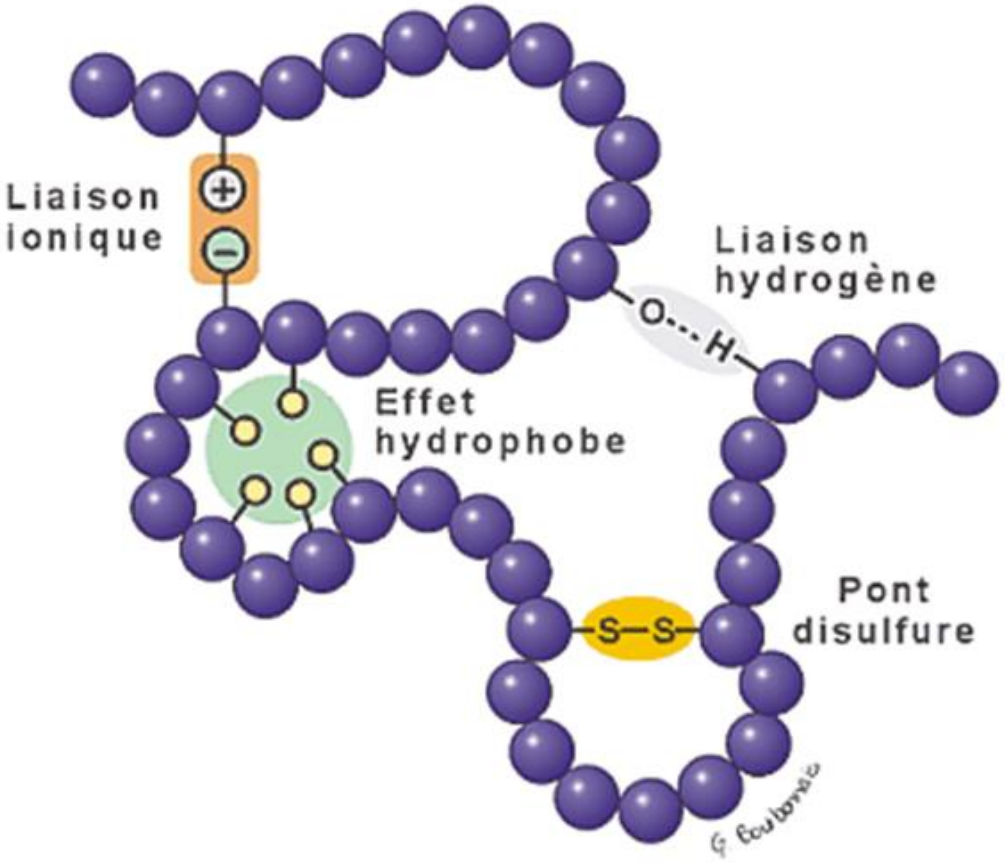
Physico-chemistry properties of proteins

Physicochemistry properties of proteins

Nom	Code		pKa du COOH	pKa du NH_3	pKa de la chaîne latérale	Poids Moléculaire	Occurrence moyenne dans les protéines (%)
Alanine	ALA	A	2,3	9,7	-	89,09	9,0
Arginine	ARG	R	2,2	9,0	12,5	174,20	4,7
Asparagine	ASN	N	2,0	8,8	-	132,12	4,4
Acide Aspartique	ASP	D	2,1	9,8	3,9	133,10	5,5
Cystéine	CYS	C	1,8	10,8	8,3	121,15	2,8
Glutamine	GLN	Q	2,2	9,1	-	146,15	3,9
Acide Glutamique	GLU	E	2,2	9,7	4,2	147,13	6,2
Glycine	GLY	G	2,3	9,6	-	75,07	7,5
Histidine	HIS	H	1,8	9,2	6,0	155,16	2,1
Isoleucine	ILE	I	2,4	9,7	-	131,17	4,6
Leucine	LEU	L	2,4	9,6	-	131,17	7,5
Lysine	LYS	K	2,2	9,0	10,0	146,19	7,0
Méthionine	MET	M	2,3	9,2	-	149,21	1,7
Phénylalanine	PHE	F	1,8	9,1	-	165,19	3,5
Proline	PRO	P	2,0	10,6	-	115,13	4,6
Sérine	SER	S	2,2	9,2	-	105,09	7,1
Thréonine	THR	T	2,6	10,4	-	119,12	6,0
Tryptophane	TRP	W	2,4	9,4	-	204,23	1,1
Tyrosine	TYR	Y	2,2	9,1	10,1	181,19	3,5
Valine	VAL	V	2,3	9,6	-	117,15	6,9
Non identifié		X					

Masse molaire moyenne d'un acide aminé: 110 g/mol

Physico-chemistry properties of proteins



Forces stabilizing proteins

(1) Based on studies of 138 hydrophobic interaction variants in 11 proteins, burying a $-CH_2-$ **group on folding contributes 1.1 ±** 0.5 kcal/mol to protein stability.

(2) The burial of non-polar side chains contributes to protein stability in two ways: first, a term that depends on the removal of the side chains from water and, more importantly, the enhanced London dispersion forces that result from the tight packing in the protein interior.

(3) Based on studies of 151 hydrogen bonding variants in 15 proteins, forming a hydrogen bond on folding contributes 1.1 **±** 0.8 kcal/mol to protein stability.

(4) The contribution of hydrogen bonds to protein stability is strongly context dependent.

(5) Hydrogen bonds by side chains and peptide groups make similar contributions to protein stability.

(6) Polar group burial can make a favorable contribution to protein stability even if the polar group is not hydrogen bonded.

(7) Hydrophobic interactions and hydrogen bonds both make large contributions to protein stability.

C. Nick Pace, J. Martin Scholtz, Gerald R. Grimsley, Forces stabilizing proteins, FEBS Letters 588 (2014) 2177–2184

Proteins stability

- Protein stability is the net balance of forces, which determine whether a protein will be in its native folded conformation or a denatured state.

- Protein stability refers to the physical (thermodynamic) stability or the chemical stability.

Chemical stability

- The chemical stability implies a loss of integrity due to the breaking of chemical bonds.
  - deamination of asparagine and/or glutamine residues,
  - hydrolysis of the peptide bond of Asp residues at low pH,
  - oxidation of the amino acid Met at high temperature,
  - Elimination of the disulfide bonds,
  - Exchange of the disulfide bonds at neutral pH.
- Other processes include thiol-catalyzed disulfide interchange and oxidation of cysteine residues.

The stability of proteins: the thermodynamic stability

- The native state of a protein is achieved when it reaches the state of the lowest free energy. From a thermodynamic point of view, the free energy of a protein depends on enthalpy and entropic contributions.
- The enthalpy contributions include covalent bonds, hydrogen bonds, electrostatic interactions and Van der Waals interactions.
- The entropic contributions involve the hydrophobic effect and conformational entropy

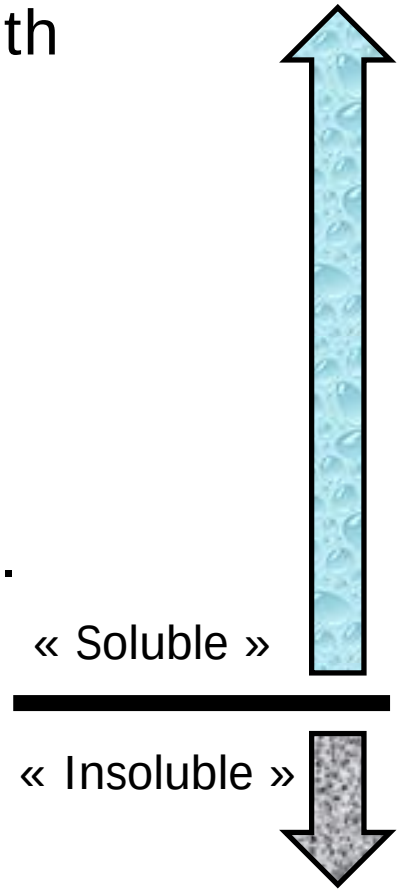
Protein aggregation

Protein aggregation

- Aggregation is a general term that encompasses several types of interactions or characteristics. Aggregates of proteins may arise from several mechanisms and may be classified in numerous ways, including soluble/insoluble, covalent/non-covalent, reversible/irreversible, and native/denatured.
- For protein therapeutics, the presence of aggregates of any type is typically considered to be undesirable because of the concern that the aggregates may lead to an immunogenic reaction (small aggregates) or may cause adverse events on administration (particulates).

Protein aggregation

- The term aggregate covers a broad spectrum of different types and sizes of associated protein states.
  - 1. Small reversible non-covalent oligomers with fast kinetics (dimers, trimers, tetramers...).
  - 2. Irreversible non-covalent oligomers.
  - 3. Covalent Oligomers (eg. disulfides).
  - 4. « Large » aggregates (> 10-mer).
    - May be reversible if non-covalent.
  - 5. « very large » aggregates (~50nm à 3µm).
    - May be reversible if non-covalent.
  - 6. Visible particles.

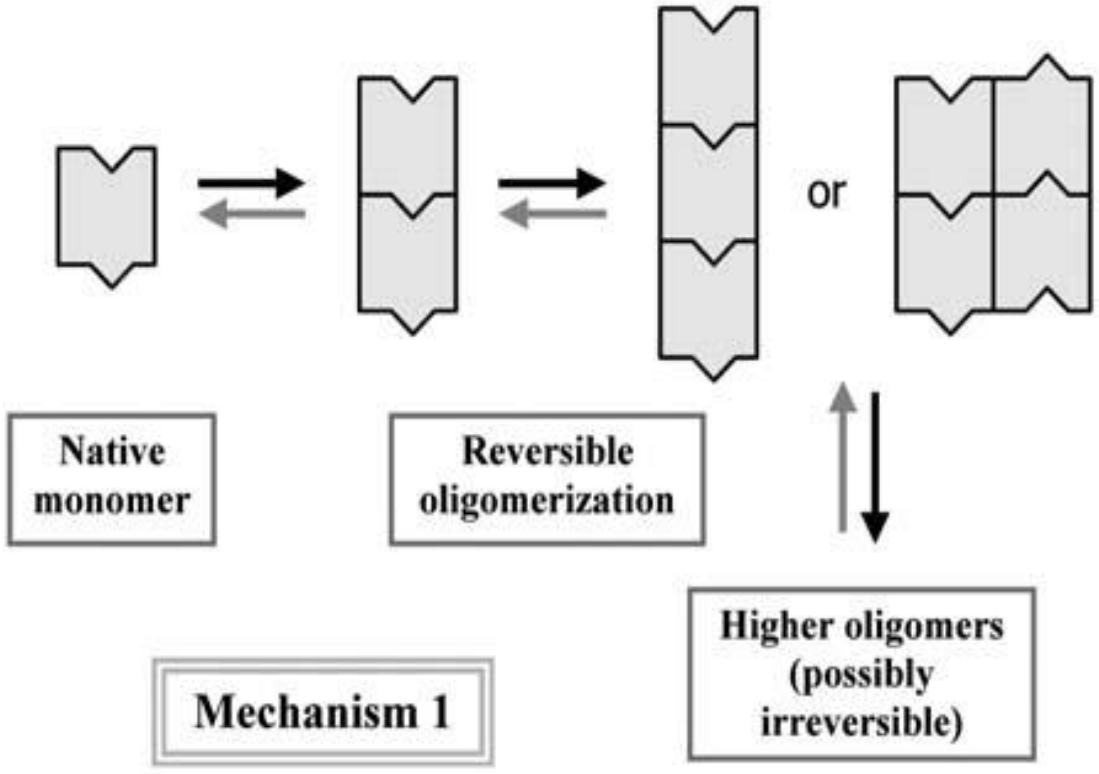


Protein aggregation

- 3 types of protein aggregation.
  - Natural and productive aggregation as the reaction n(G-actin) →(F-actin)n controlling mobility and shape of cells.
  - Aggregation unwanted in biology including α synuclein, amyloid β, polyglutamine and Prion proteins that play an important role in neuro-degenerative disease as Parkinson's, Alzheimer's, Huntington and the prion-related diseases (mad cow).
  - Aggregation unwanted in vitro important for the conservation of proteins for various applications in research and industry.

Protein aggregation

MECHANISM 1: REVERSIBLE ASSOCIATION OF THE NATIVE MONOMER

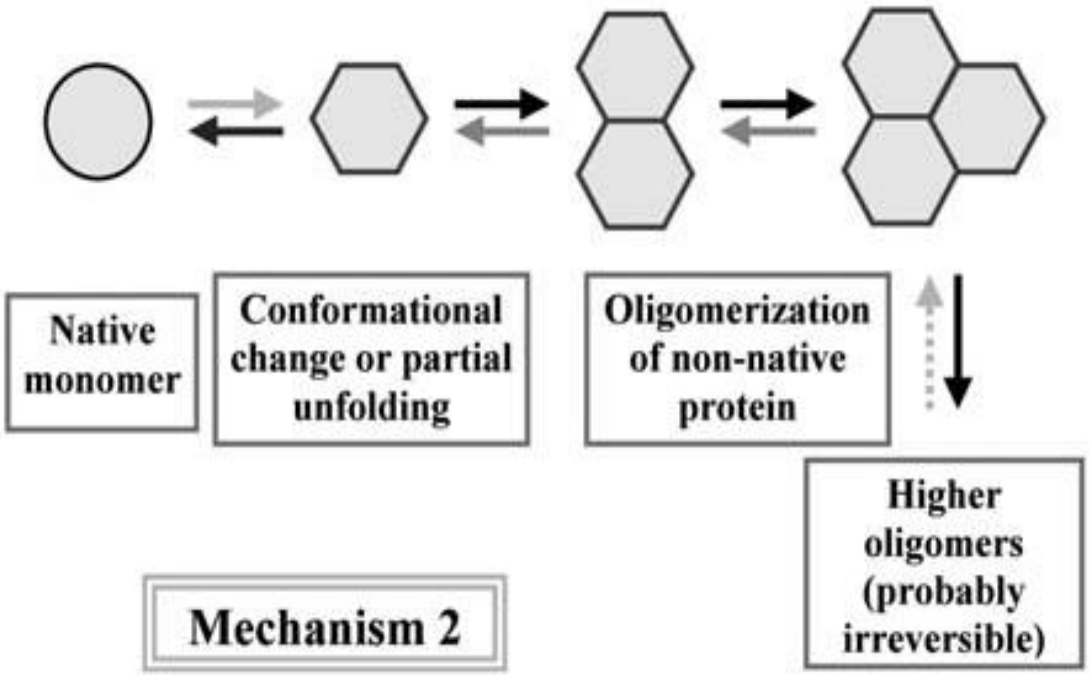


The tendency to reversibly associate (aggregate) is intrinsic to the native form of the protein. The surface of the native protein monomer is self-complementary so it will readily self-associate to form reversible small oligomers. There may be multiple "sticky" or complementary patches on the monomer surface. Those can then produce different types of interfaces, potentially producing multiple conformations for oligomers of the same stoichiometry and different patterns of oligomer growth. As the protein concentration rises and larger and larger oligomers form (driven by the law of mass action), over time these larger aggregates often become irreversible (sometimes through formation of covalent bonds such as disulfide linkages).

John S. Philo and Tsutomu Arakawa, Current Pharmaceutical Biotechnology, 2009, 10, 348-351

Protein aggregation

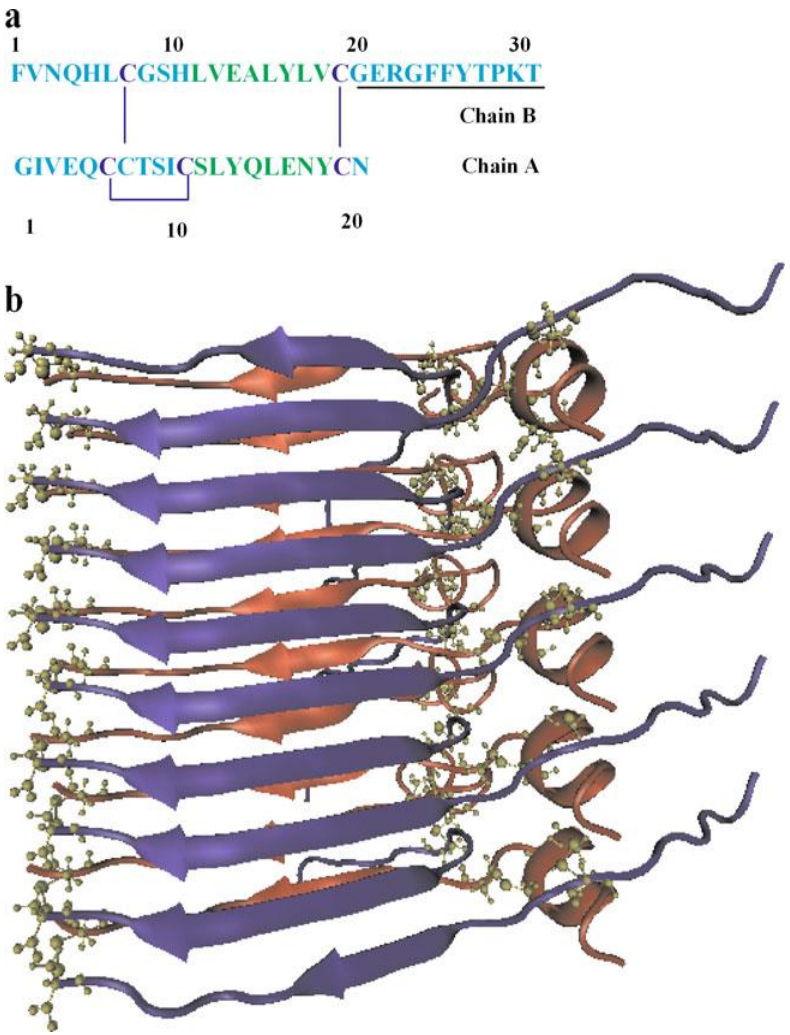
MECHANISM 2: AGGREGATION OF CONFORMATIONALLY-ALTERED MONOMER



In contrast to Mechanism 1, for Mechanism 2 the native monomer has a very low propensity to reversibly associate. However after it transiently undergoes a conformational change or partial unfolding the resultant altered conformation of monomer associates strongly (in a manner similar to Mechanism 1). Thus a key difference between Mechanisms 1 and 2 is that in Mechanism 2 the first step is a conformational change to a non-native state, and at any given time the fraction of protein in that aggregation-prone non-native state will usually be quite small. For Mechanism 2 aggregation will be promoted by stresses such as heat or shear that may trigger the initial conformational change. A further (and important) consequence is that aggregation will be inhibited by excipients or conditions that stabilize the native conformation. This aggregation mechanism does appear to be the dominant one for many proteins.

John S. Philo and Tsutomu Arakawa, Current Pharmaceutical Biotechnology, 2009, 10, 348-351

Protein aggregation: mechanism 2



Amino acid sequences and structural models of double layer insulin oligomers.

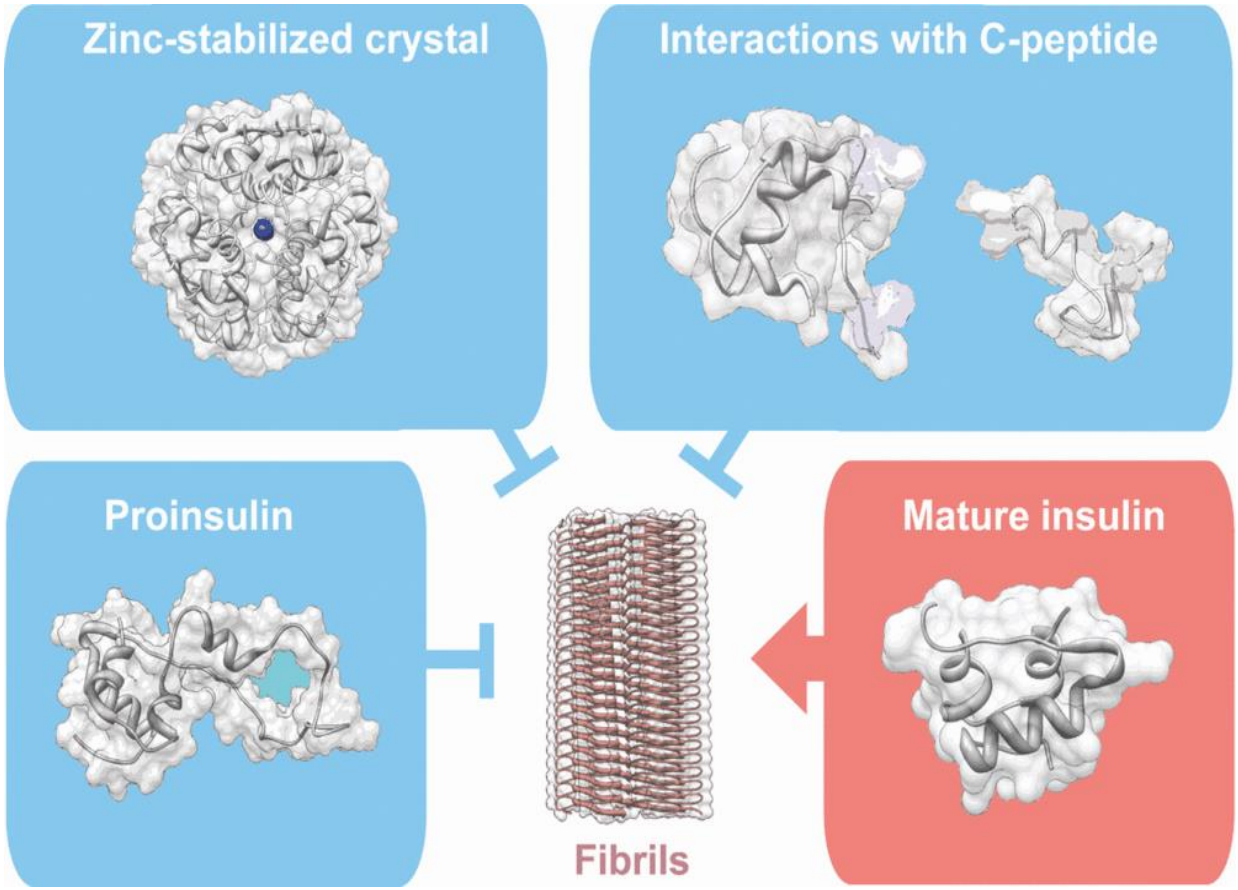
a. Amino acid sequence of insulin. Segments LVEAYLV of chain B and SLYQLENT of chain A are colored green. Disulfide bonds are colored blue. The C-terminal region of chain B (underlined) is not involved in amyloid fibrillation.

b. Single-layered structural models of insulin oligomers (ten-stranded). Two chains are associated via an inter-digitated pair of LYQLENY molecules of chain A and LVEALYL molecules of chain B, which interlock tightly to form the dry steric zipper. Chain A is red and chain B is shown in blue. Disulfide bonds are indicated in yellow

Workalemahu Mikre Berhanu & Artëm E. Masunov, J Mol Model (2012) 18:1129–1142

Protein aggregation: mechanism 2

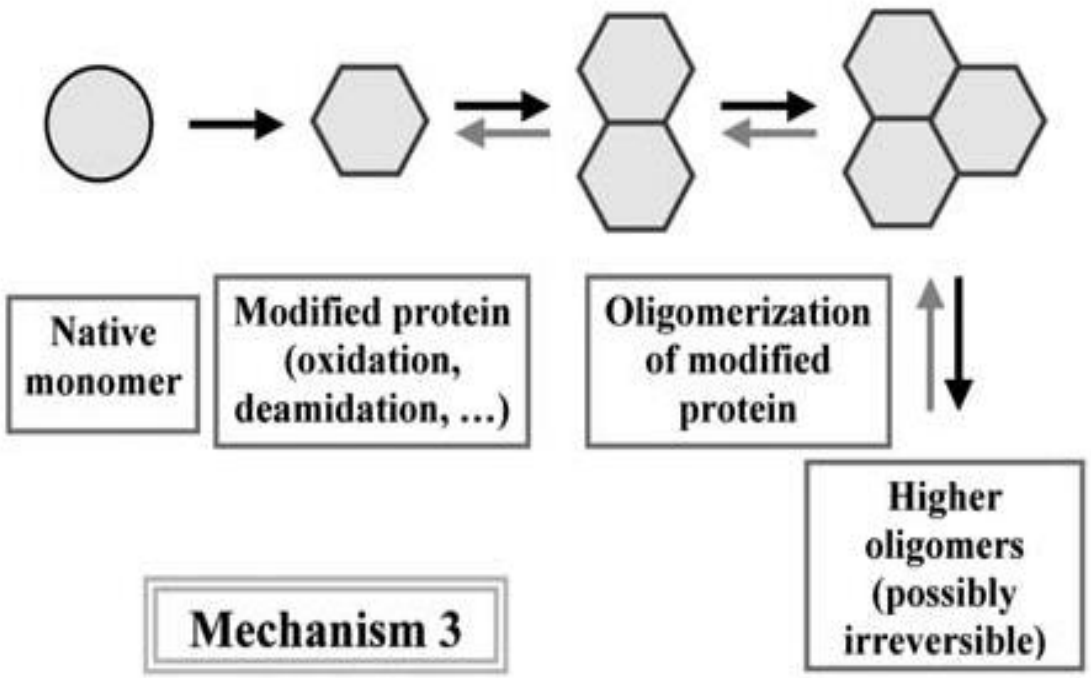
Multiple strategies for the prevention of insulin fibrillation



Biochemical Journal 2012 447, 185-192 - Michael Landreh, Jan Johansson and others.

Protein aggregation

MECHANISM 3: AGGREGATION OF CHEMICALLY MODIFIED PRODUCT



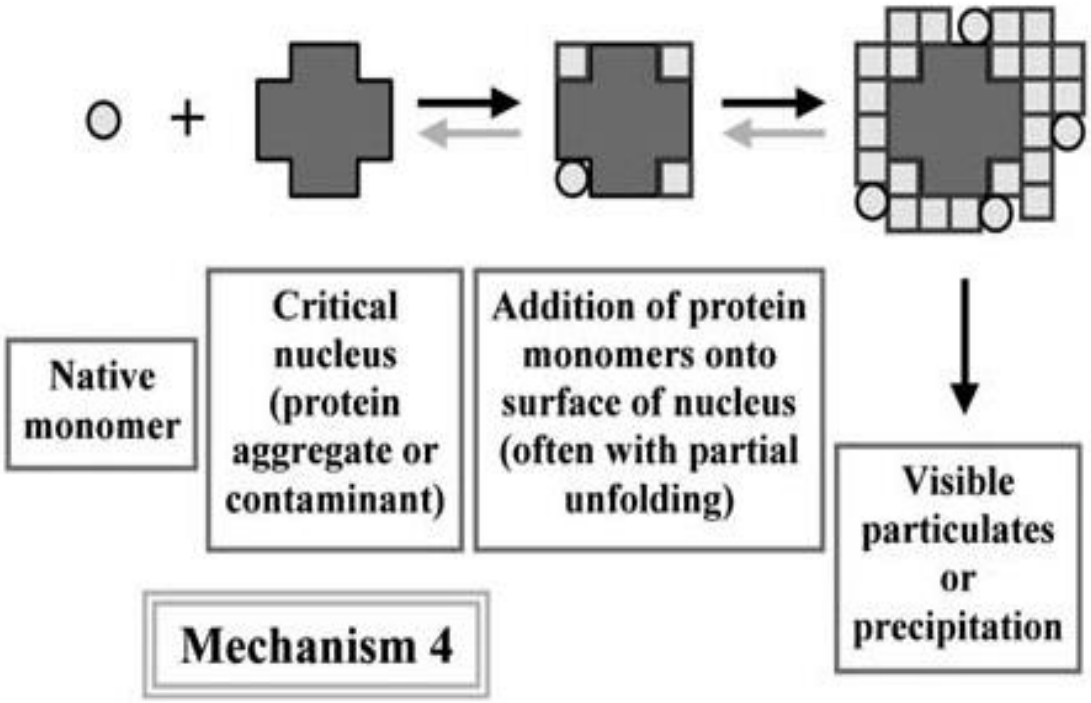
John S. Philo and Tsutomu Arakawa, Current Pharmaceutical Biotechnology, 2009, 10, 348-351

Mechanism 3 is really a variant of mechanism 2 where the change in protein conformation that precedes aggregation is caused by a difference in covalent structure. Usually this difference is caused by chemical degradation such as oxidation of methionine, deamidation or proteolysis. Chemical changes may for example create a new sticky patch on the surface, or change the electric charge in a way that reduces electrostatic repulsion between monomers.

In some cases however the chemically different species is not a degradant but rather it is a normal variant within the bulk drug product (for example in glycoproteins there might be a unglycosylated or under glycosylated fraction that is prone to aggregation.

Protein aggregation

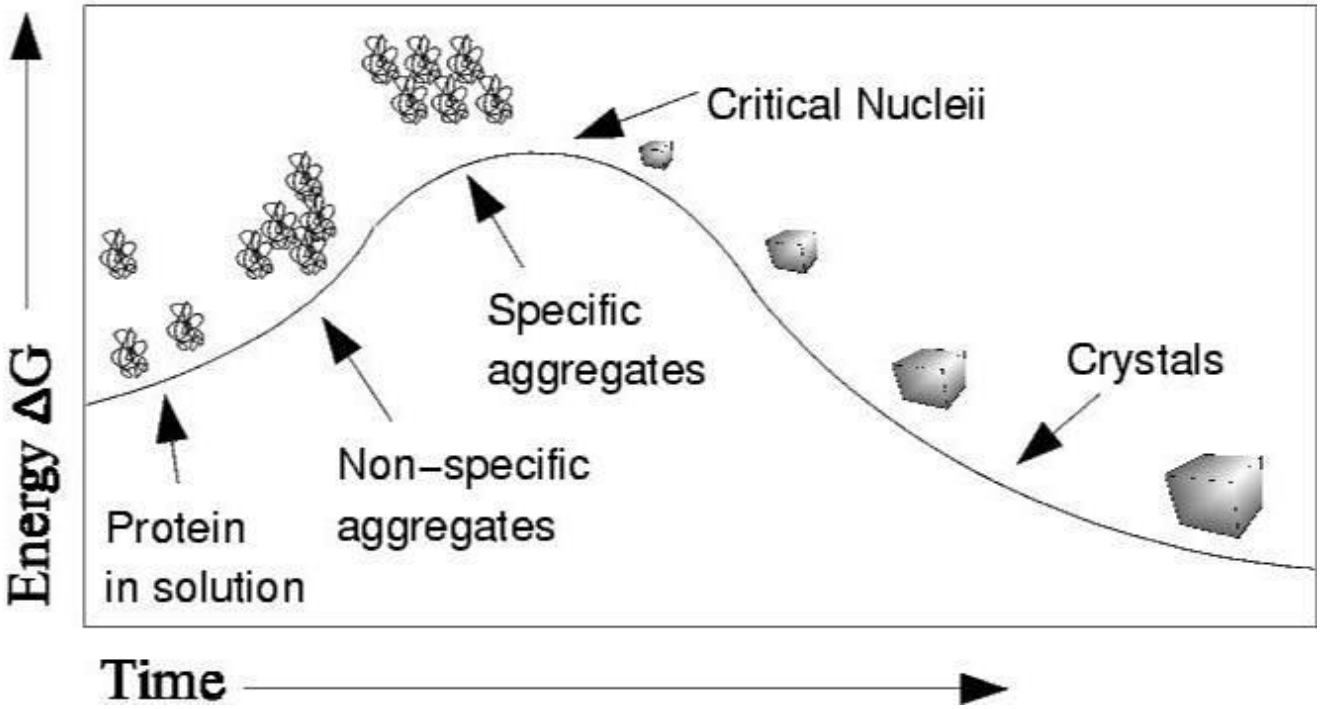
MECHANISM 4: NUCLEATION-CONTROLLED AGGREGATION



Nucleation-controlled aggregation is a common mechanism for formation of visible particulates or precipitates. In this mechanism the native monomer has a low propensity for formation of small and moderately-sized oligomers (the addition of monomers onto these smaller aggregates is not thermodynamically favored). However if an aggregate of sufficient size manages to form, then the growth of this so called "critical nucleus" through addition of monomers is strongly favored and the formation of much larger species is rapid. This type of process is similar to growing large crystals by adding micro-crystal "seeds" to a saturated solution, and thus the critical nuclei are also sometimes called the "seeds" or "templates" for aggregate growth.

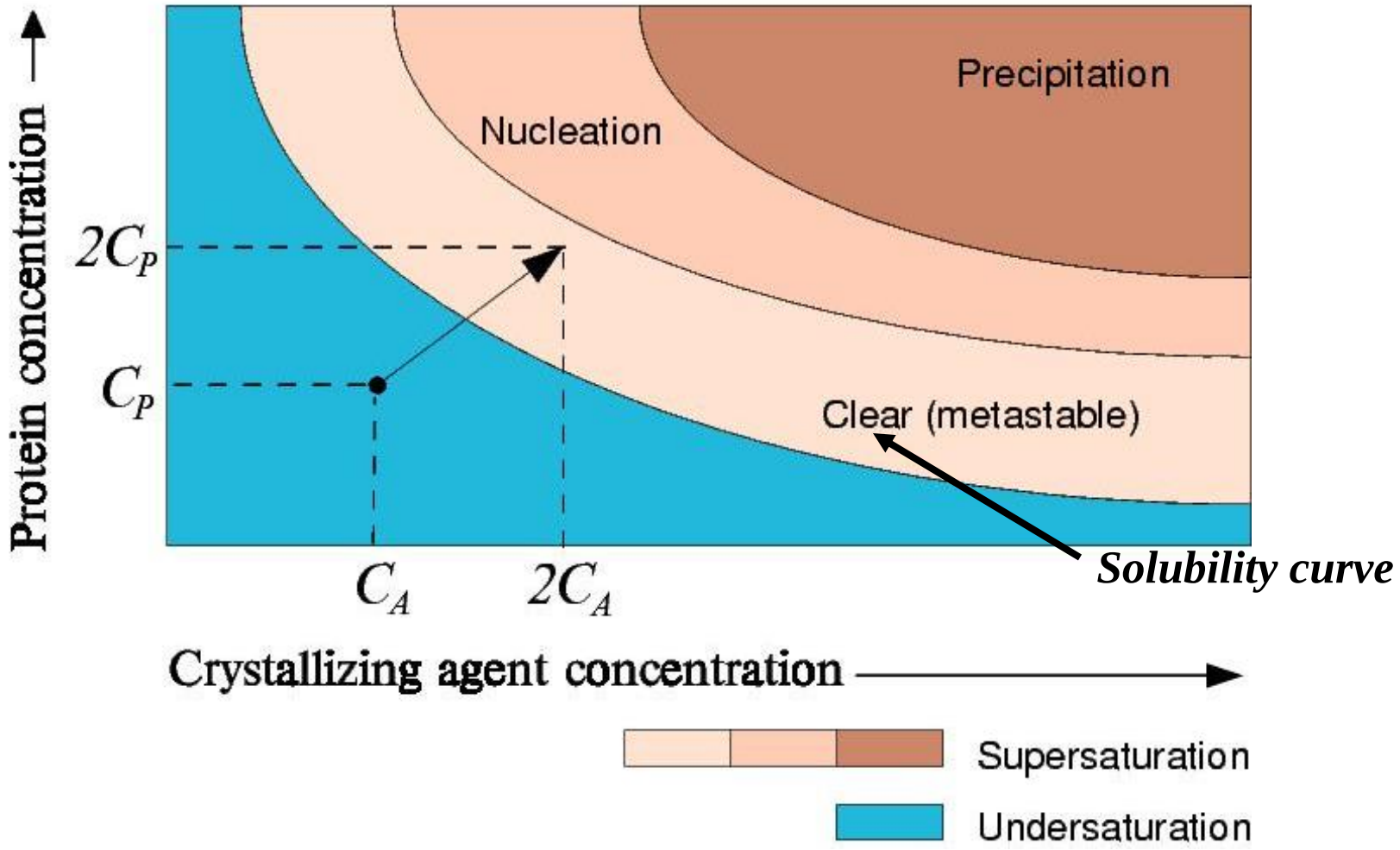
John S. Philo and Tsutomu Arakawa, Current Pharmaceutical Biotechnology, 2009, 10, 348-351

Crystallization



Crystallization

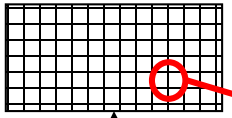
Phase transition



Crystallization assays: general strategy

General screening

Commercial screens
Home-made screens
Total of more than 3000 conditions



Protein characterization

Crystals OK → Data Collection

Poly crystals
Small crystals
Bad diffracting crystals
Twinned crystals
Amorphous solid phases
.....

Refinement

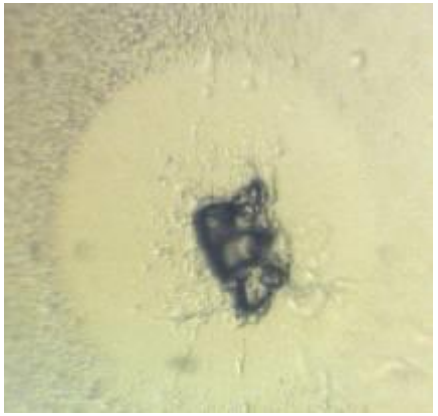
Crystals OK

No crystals

Cloning, expression, purification, …

Crystallization assays: amorphous solid phase

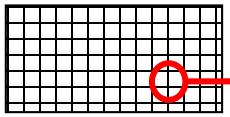


Amorphous solid phase → Crystallization refinement

Crystallization assays: refinement

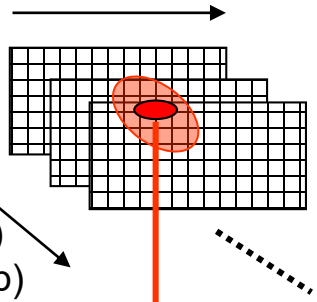
General screening



Parameters refined:
[Precipitating Agent]
pH
Additives

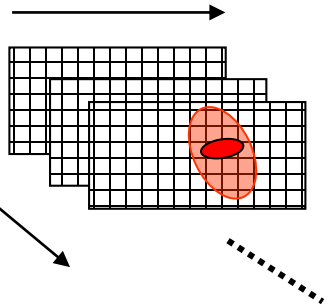
Fbuffer(A,B…)
FpH [Δ=0.5]
(x-1< pH < x+1)

F(prec. Agent)
PEG,AS Δc=5%...



F(add)
F(temp)

F(prec. Agent)
PEG,AS Δc=1%...

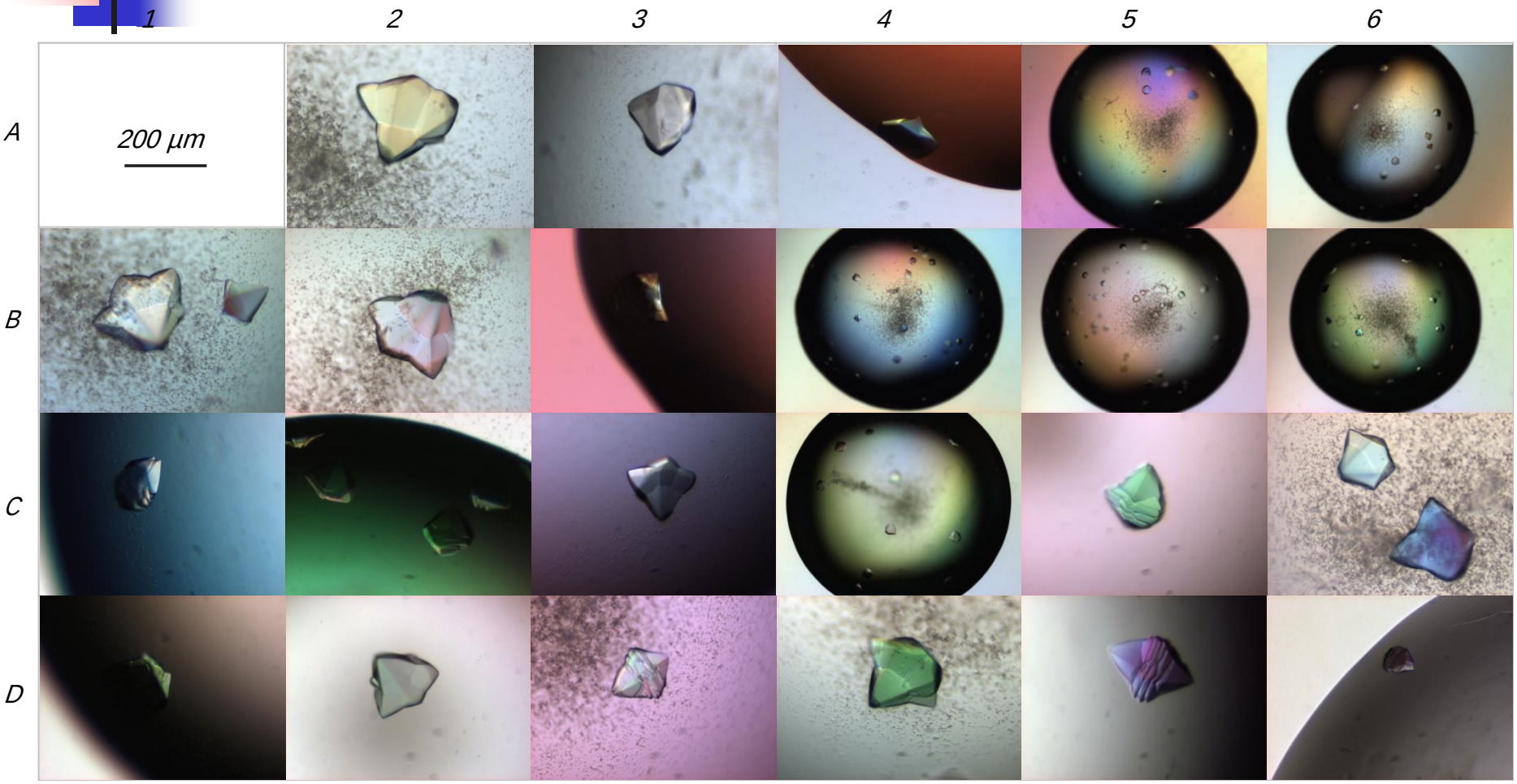


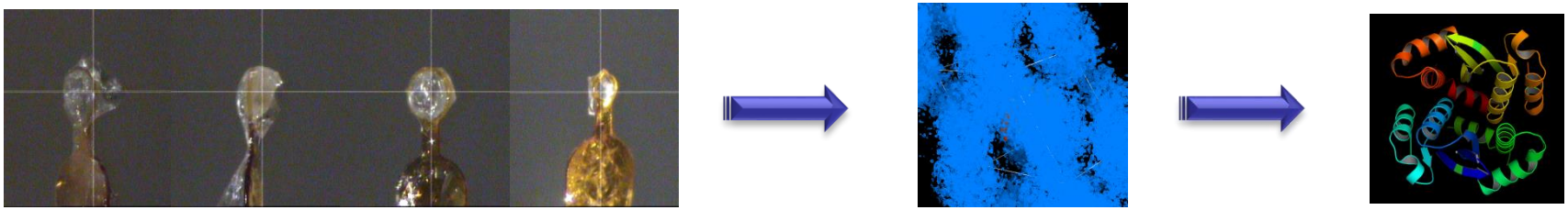
Fbuffer(A,B…)
FpH [Δ=0.1]
(x-0.5< pH < x+0.5)

F[add]

- Increasing drop volume
- Crystallization in gel
- Sitting drops on fluorinert
- Cubic phase crystallization
- ……..

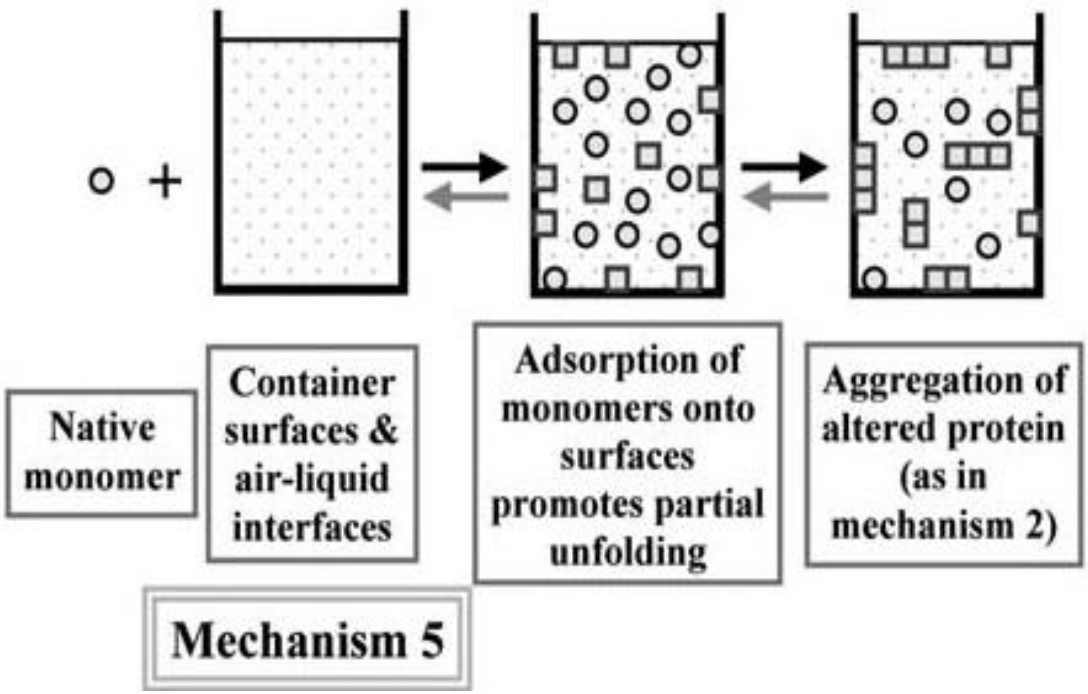
HIV integrase 50 - 212 (F185K), Crystallization





Protein aggregation

MECHANISM 5: SURFACE INDUCED AGGREGATION



John S. Philo and Tsutomu Arakawa, Current Pharmaceutical Biotechnology, 2009, 10, 348-351

This aggregation process starts with binding of the native monomer to a surface. In the case of an air-liquid interface that binding would probably be driven by hydrophobic interactions, but for a container favorable electrostatic interactions might also be involved. After this initial reversible binding event the monomer undergoes a change in conformation (for example to increase the contact area with the surface). Like in Mechanism 2, it is then that conformational altered monomer which aggregates, but in this case that aggregation might occur either on the surface or perhaps after the altered monomer is released back into the solution. Freeze/thaw damage can also arise from aggregation at the surfaces of ice crystals or crystals of excipients, and thus can occur through Mechanism 5, but freeze/thaw damage can also involve other mechanisms such as changes in pH.

Protein aggregation (misfolding)

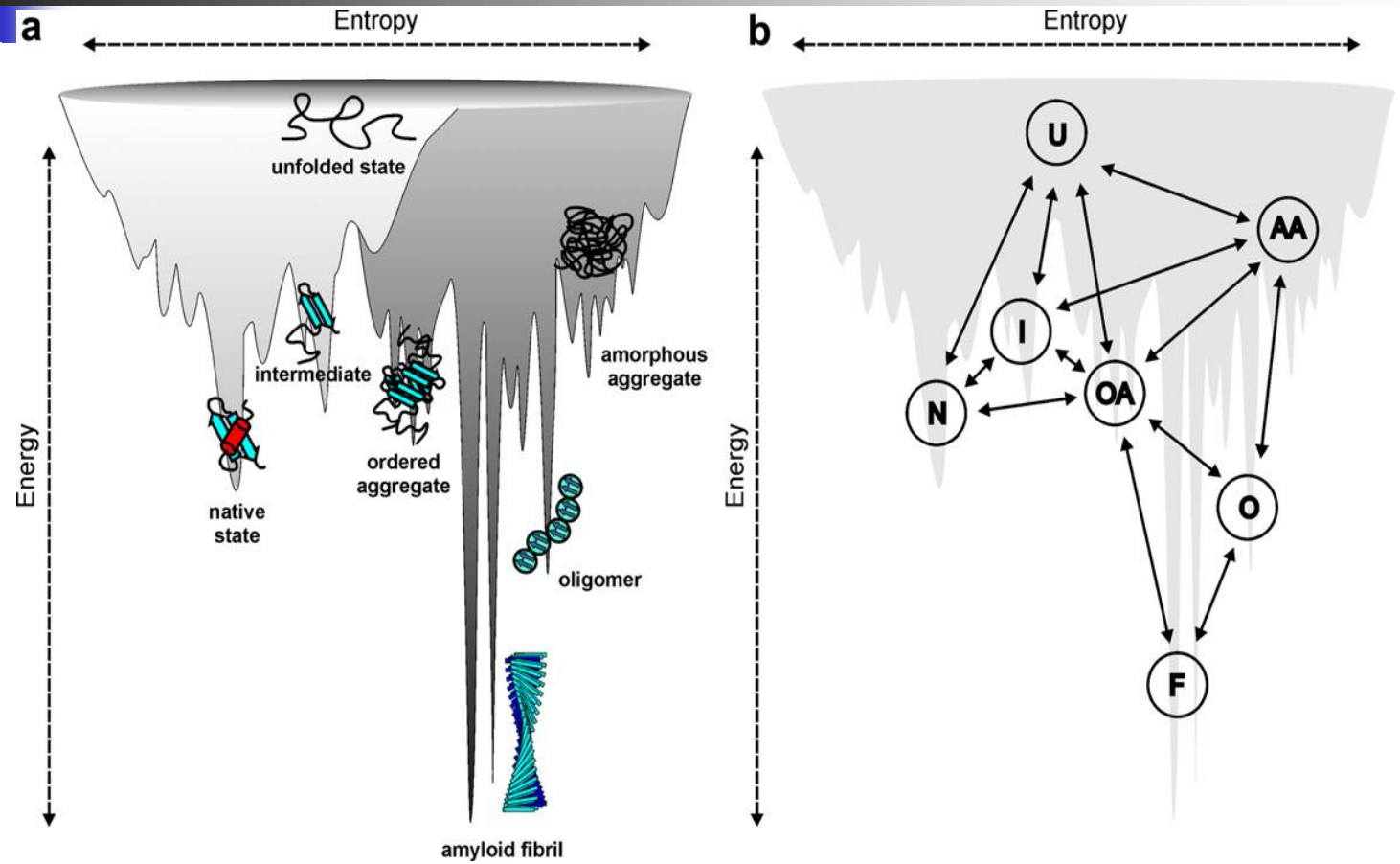
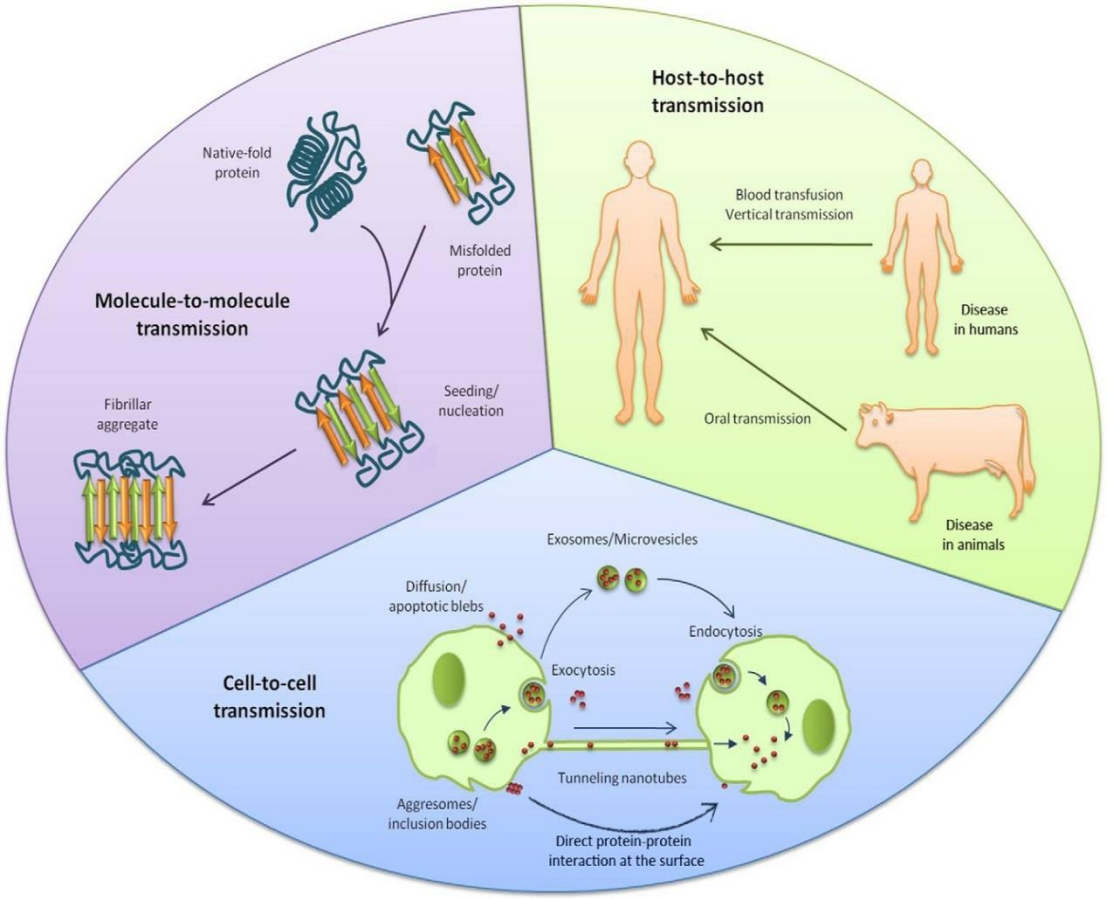


Illustration of a combined energy landscape for protein folding and aggregation. (a) The surface illustrates the roughness of the protein energy landscape, showing the multitude of conformational states available to a polypeptide chain. While rather simple folding funnels (light grey) can describe the conformational search of a single polypeptide chain to a functional monomer, intermolecular protein association dramatically increases ruggedness (dark grey). (b) Proposed pathways linking the conformational states shown in (a) populated on the combined folding and aggregation energy landscape.

T.R. Jahn, S.E. Radford / Archives of Biochemistry and Biophysics 469 (2008) 100–117

Neuro-degenerative diseases



Transmission of protein misfolding between molecules, cells and individuals. Prion-like transmission of protein misfolding may operate at various levels, including molecule-to-molecule, cell-to-cell and host-to-host. Propagation of the pathological conformational changes and downstream effects to cells, tissues and the entire individual appears to be a universal property of misfolded protein aggregates.

Ines Moreno-Gonzalez, Claudio Soto, Seminars in Cell & Developmental Biology 22 (2011) 482– 487

Protein aggregation: sequence and gatekeepers

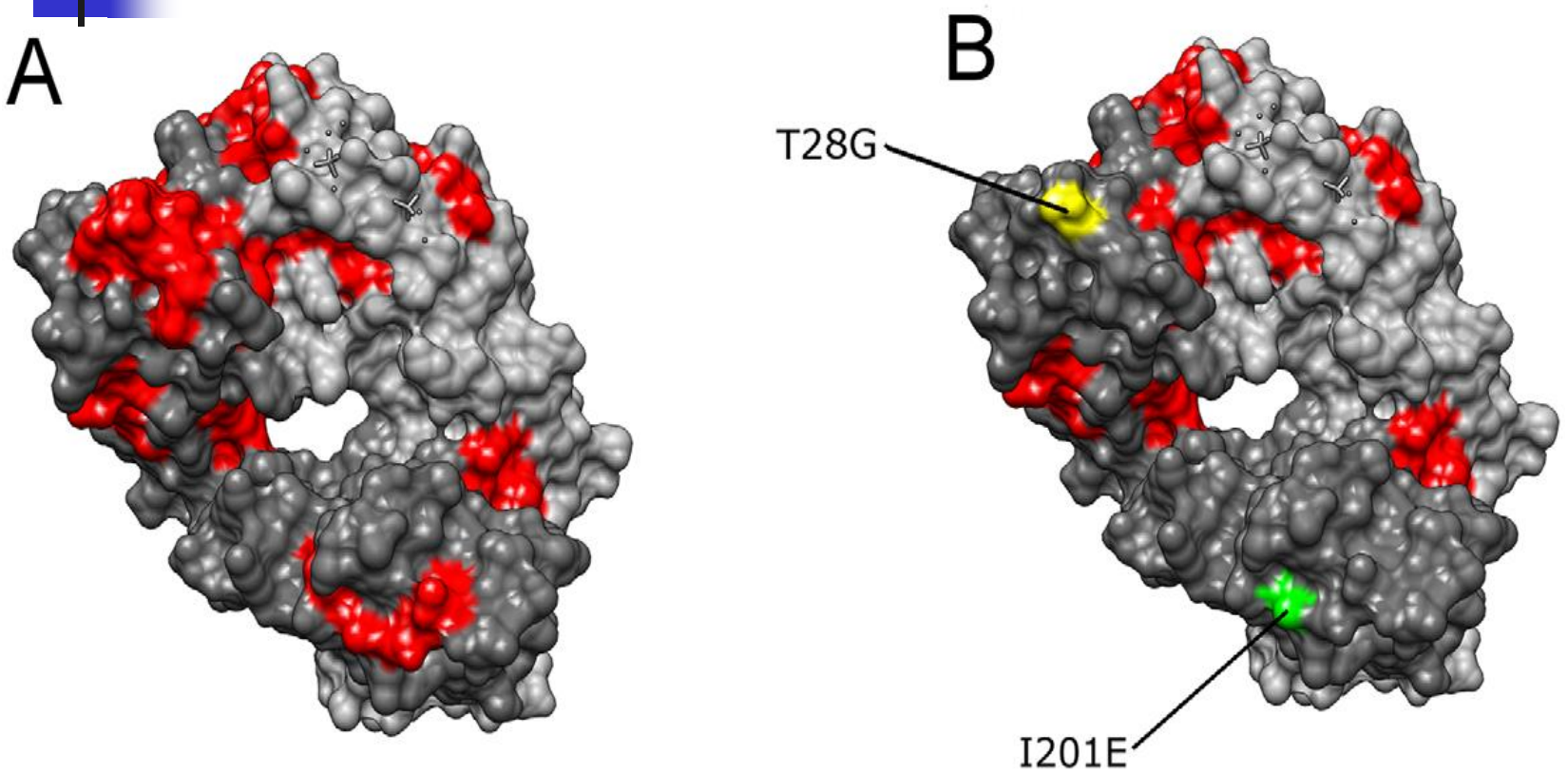
Evolutionary pressure against protein aggregation also results in the placement of amino acids that counteract aggregation at the flanks of protein sequences that are aggregation-prone. These so-called aggregation gatekeepers reduce aggregation by opposing nucleation of aggregates.

This disruption is achieved using the repulsive effect of charge (arginine [R], lysine [K], aspartate [D], glutamate [E]), the entropic penalty on aggregate formation (R and K) or incompatibility with β-structure backbone conformation (proline [P]).

The evolutionary enrichment of charged amino acids on the flanks of aggregating regions is coupled to chaperone specificity: studies have shown that chaperones recognize the pattern of charged residues followed by a hydrophobic region. As gatekeeper residues are enriched at the flanks of strongly aggregating hydrophobic sequences, chaperone binding occurs on average more tightly to strongly aggregating than to weakly aggregating sequences

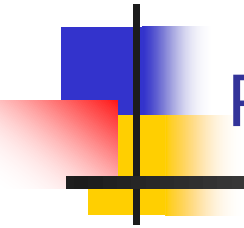
Joke Reumers et al., HUMAN MUTATION, Vol. 30, No. 3, 431–437, 2009

Protein aggregation: sequence and gatekeepers



The crystal structure (space-filling model) of the anti- ErbB2 Fab2C4 (PDB code: 1L7I) is shown. (A). This is a humanized monoclonal antibody fragment that binds to the extracellular domain of the human oncogene product ErbB2 (ErbB2 has been shown to play an important role in the pathogenesis of certain aggressive types of breast cancer). Computationally predicted 'aggregation-prone' regions by AMYLPRED2 are colored red. Performing only two single amino acid substitutions (T28G and I201E), the AMYLPRED2 output suggests that the antibody has 'lost' two crucial 'aggregation-prone' regions and may, therefore, be more soluble, not forming aggregates (B).

Antonios C. et al, 2013, PLoS ONE 8(1): e54175. doi:10.1371/journal.pone.0054175



Proteins solubility

Proteins solubility

The solubility of a protein in aqueous solution varies from almost completely insoluble to hundreds of milligrams per milliliter. For instance, crambin has been reported to be completely insoluble in water, and serum albumins have solubilities of >500 mg/mL.

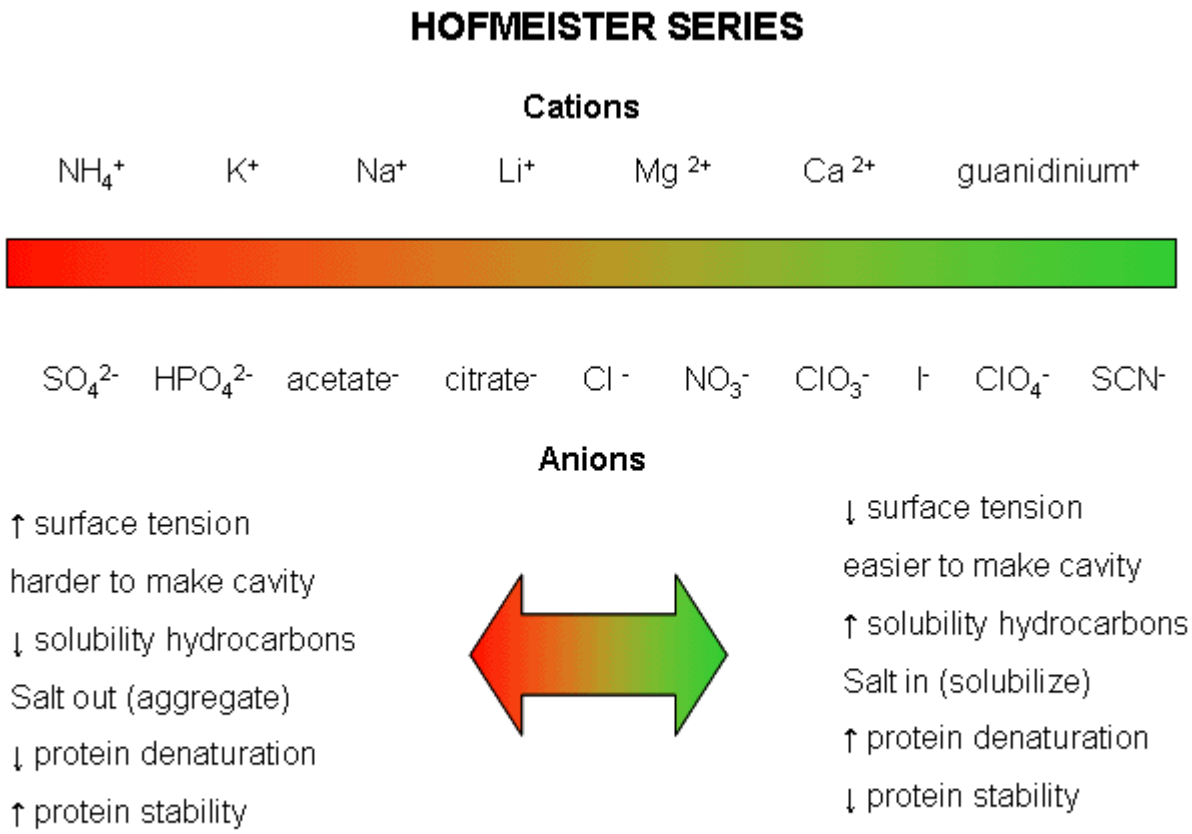
Low protein solubility has also been implicated in a number of human diseases. The P23T mutation in human gD-crystallin shows a markedly decreased solubility and leads to childhood onset of cataracts.

Kramer et al., 2012, Biophysical Journal, 102, 1907-1915

Proteins solubility

- Protein solubility is a thermodynamic parameter defined as the concentration of protein in a saturated solution that is in equilibrium with a solid phase, either crystalline or amorphous, under a given set of conditions
- Solubility can be influenced by a number of extrinsic and intrinsic factors.
  - Extrinsic factors that influence protein solubility include pH, ionic strength, temperature, and the presence of various solvent additives
  - The intrinsic factors that influence protein solubility are defined primarily by the amino acids on the protein surface

Solubility: ionic strength, Hofmeister series

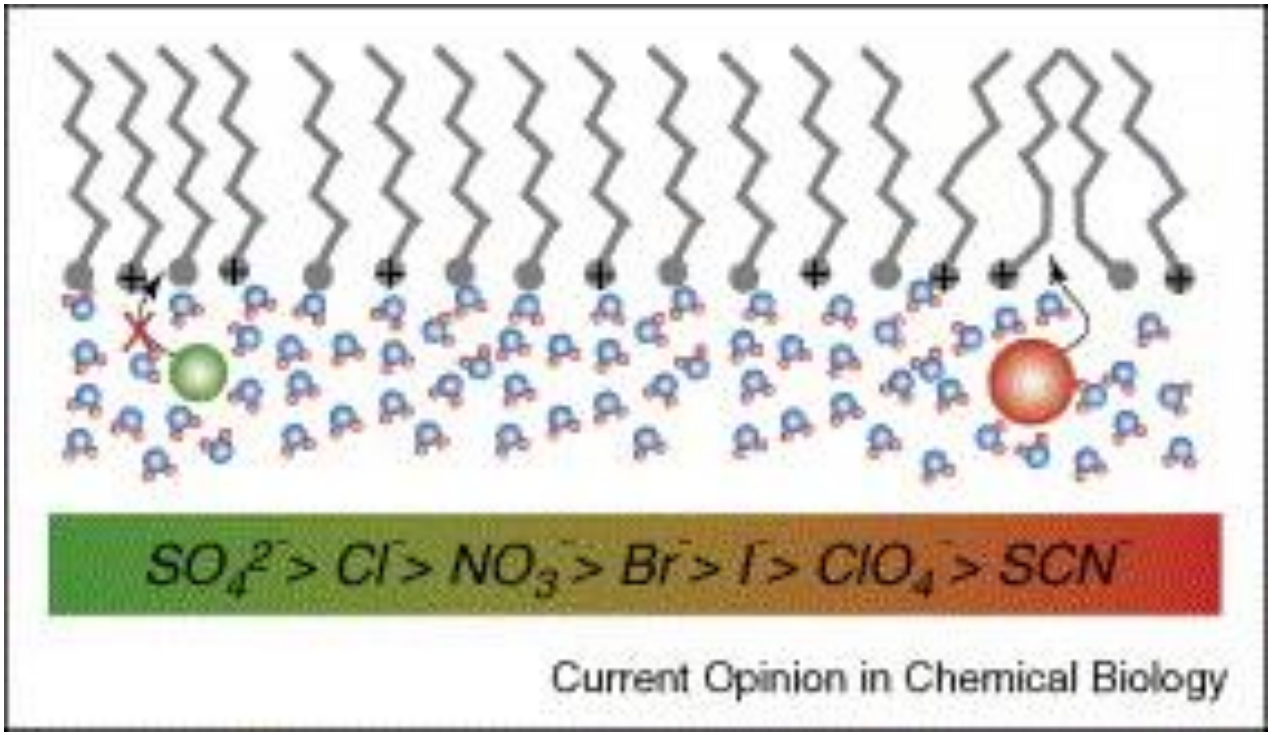


Kosmotropes « water structure makers »

Chaotropes « water structure brakers »

Yanjie Zhang , Paul S Cremer, Current Opinion in Chemical Biology Volume 10, Issue 6 2006 658 - 663

Proteins solubility



An octadecylamine monolayer at the air/salt solution interface. The arrows compare the penetration of kosmotropic (green) and chaotropic (red) anions into the headgroup region of the monolayer. The red x indicates that the kosmotrope doesn't effectively penetrate into the headgroup region.

Yanjie Zhang , Paul S Cremer, Current Opinion in Chemical Biology Volume 10, Issue 6 2006 658 - 663

Solubility: ionic strength

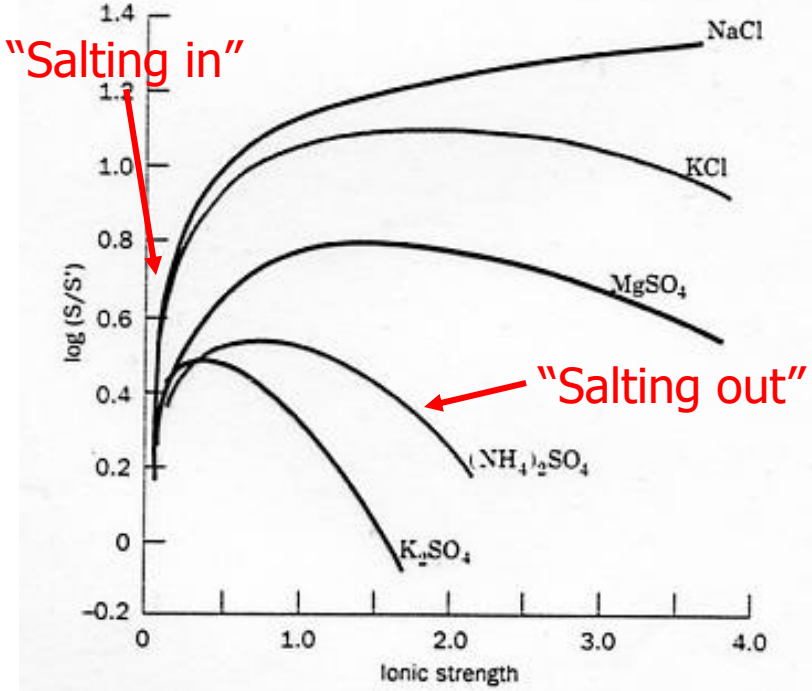


Figure 5-3
Solubility of carboxy-hemoglobin at its isoelectric point as a function of ionic strength and ion type. Here S and S′ are, respectively, the solubilities of the protein in the salt solution and in pure water. The logarithm of their ratios is plotted so that the solubility curves can be placed on a common scale. [After Green, A. A., *J. Biol. Chem.* **95**, 47 (1932).]

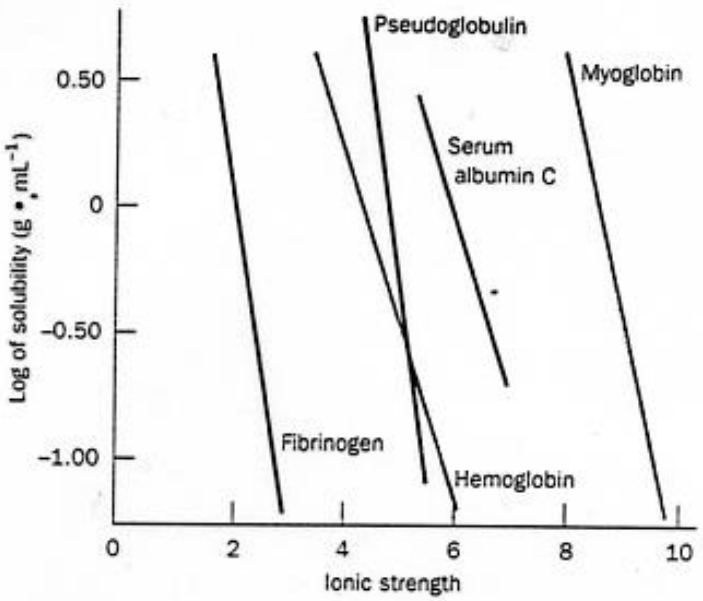
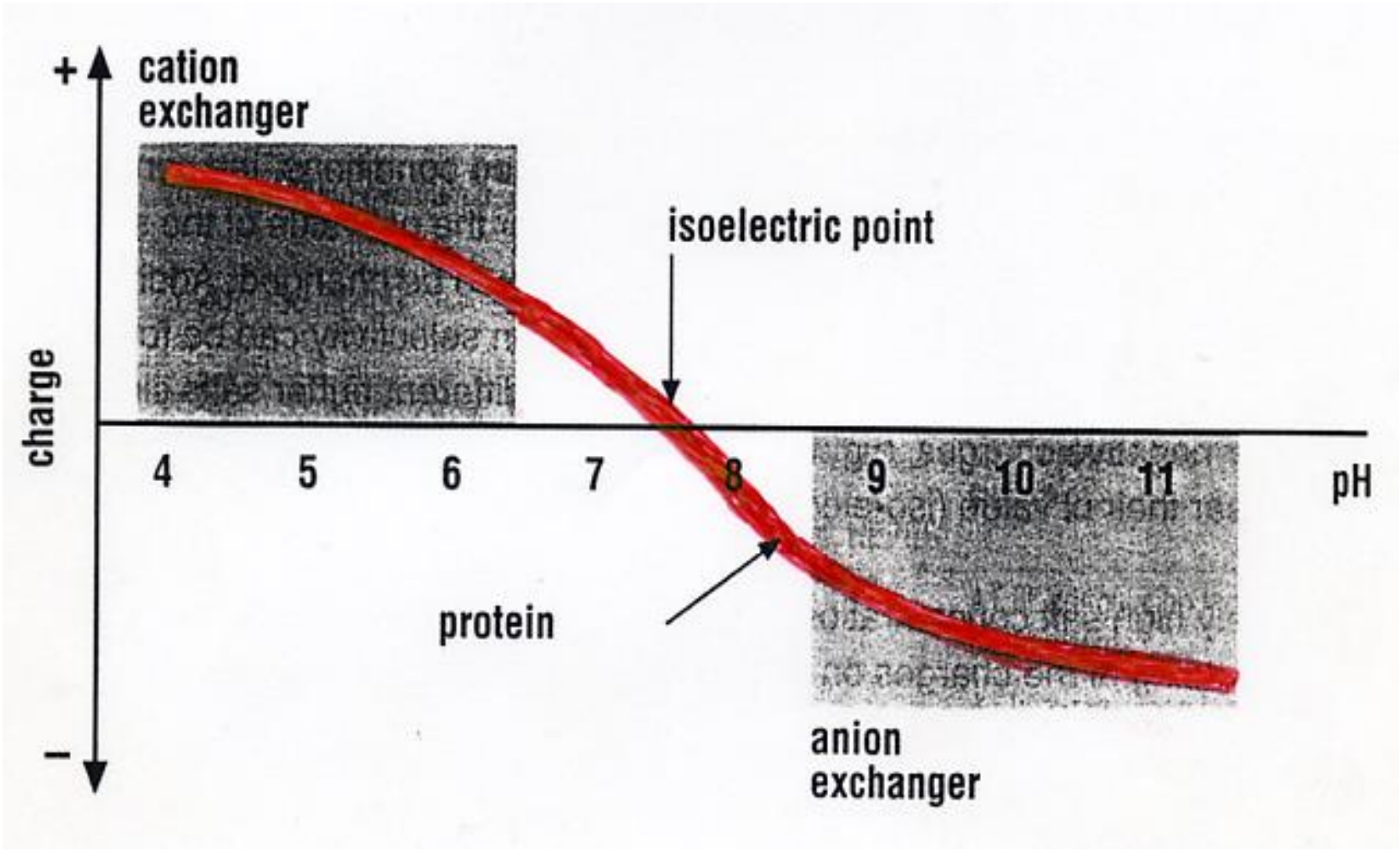


Figure 5-2
Solubilities of several proteins in ammonium sulfate solutions. [After Cohn, E. J. and Edsall, J. T., *Proteins, Amino Acids and Peptides, p.* 602, Academic Press (1943).]

Proteins solubility

Protein net charge in function of pH



Solubility: pH

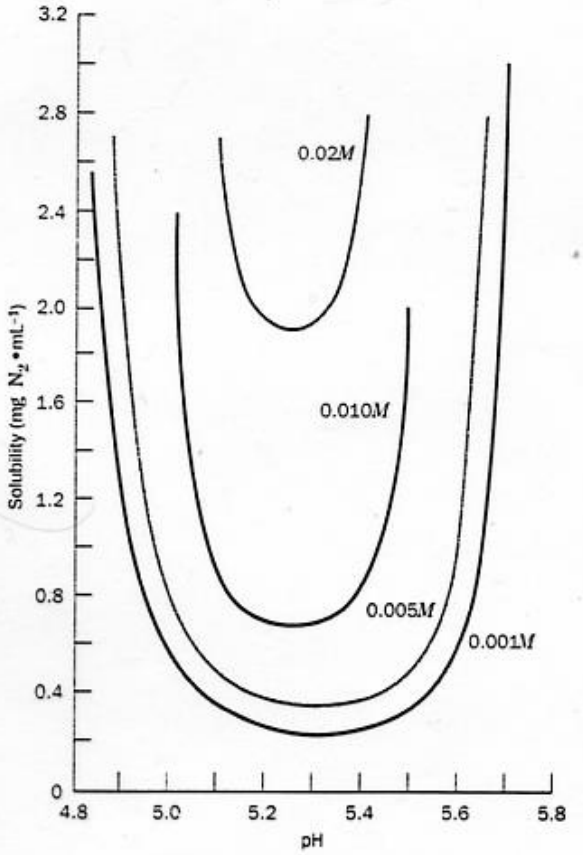


Figure 5-4
Solubility of β-lactoglobulin as a function of pH at several NaCl concentrations. [After Fox, S. and Foster, J. S., *Introduction to Protein Chemistry, p.* 242, Wiley (1957).]

The Isoelectric Points of Several Common Proteins

Protein	Isoelectric pH
Pepsin	<1.0
Ovalbumin (hen)	4.6
Serum albumin (human)	4.9
Tropomyosin	5.1
Insulin (bovine)	5.4
Fibrinogen (human)	5.8
γ-Globulin (human)	6.6
Collagen	6.6
Myoglobin (horse)	7.0
Hemoglobin (human)	7.1
Ribonuclease A (bovine)	7.8
Cytochrome *c* (horse)	10.6
Histone (bovine)	10.8
Lysozyme (hen)	11.0
Salmine (salmon)	12.1

Solubility: hydrophobicity

Capacity of non-polar areas of one or several molecules to come together to minimize the exposure of the hydrophobic area to solvent

TABLEAU I — PROPRIÉTÉS DE CERTAINS DÉTERGENTS

Propriétés	Cholate de sodium	CHAPS	Octyl glucoside	Triton X100	Lubrol PX
Poids moléculaire des micelles	1 700	6 150	8 000	90 000	64 000
Concentration micellaire critique % (p/v)	0,36	0,49	0,73	0,02	0,006
"Dialysabilité"	oui	oui	oui	non	non
Présence de charges	oui	non	non	non	non

Protein denaturation

Denaturation is a process in which proteins or nucleic acids lose the tertiary structure and secondary structure which is present in their native state, by application of some external stress or compound such as a strong acid or base, a concentrated inorganic salt, an organic solvent (e.g., alcohol or chloroform), or heat.

If proteins in a living cell are denatured, this results in disruption of cell activity and possibly cell death. Denatured proteins can exhibit a wide range of characteristics, from loss of solubility to communal aggregation.

Protein denaturation

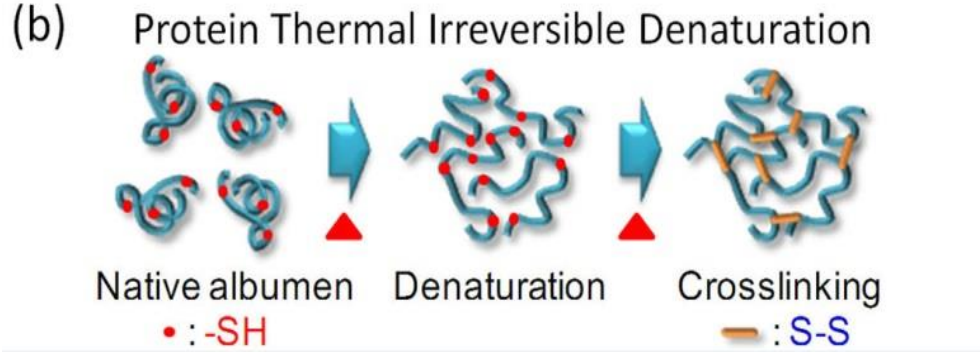
- The denaturing agents are numerous and can be either of physical nature, or be chemical agents.
- Physical agents
  - The increase in temperature causes a thermal agitation of the atoms in the molecule which causes the break of weak interactions such as hydrogen bonds, that stabilize the spatial structure.
  - Change in pH: it leads to a modification of the charges brought by the ionizable groups, and therefore alters the ionic bonds and hydrogen stabilizing the spatial structure.
- Chemical agents
  - Chaotropic agents as urea or guanidine chloride. At high concentrations, these compounds greatly weaken hydrogen bonds (main links of low energies responsible for the maintenance of secondary, tertiary and quaternary protein structures).
  - Thiols reducing agents like 2-mercaptoéthanol or DTT (Dithiothreitol). They allow the reduction (rupture) of the disulfide bonds and can thus contribute to destabilize the tertiary or quaternary protein structure.
  - Bases and acids, by alteration of pH.
  - Detergents, by modification of the interaction with the aqueous solvent.

Protein denaturation

- Heat can be used to disrupt hydrogen bonds and non-polar hydrophobic interactions. This occurs because heat increases the kinetic energy and causes the molecules to vibrate so rapidly and violently that the bonds are disrupted. The proteins in eggs denature and coagulate during cooking. Other foods are cooked to denature the proteins to make it easier for enzymes to digest them. Medical supplies and instruments are sterilized by heating to denature proteins in bacteria and thus destroy the bacteria.

(a)

(b)



Protein denaturation

- Alcohol Disrupts Hydrogen Bonding:

- Hydrogen bonding occurs between amide groups in the secondary protein structure. Hydrogen bonding between "side chains" occurs in tertiary protein structure in a variety of amino acid combinations. All of these are disrupted by the addition of another alcohol.

- A 70% alcohol solution is used as a disinfectant on the skin. This concentration of alcohol is able to penetrate the bacterial cell wall and denature the proteins and enzymes inside the cell. A 95% alcohol solution merely coagulates the protein on the outside of the cell wall and prevents any alcohol from entering the cell.

Protein denaturation

- Heavy Metal Salts:

- Heavy metal salts act to denature proteins in much the same manner as acids and bases. Heavy metal salts usually contain Hg^{+2}, Pb^{+2}, Ag^{+1} Tl^{+1}, Cd^{+2} and other metals with high atomic weights. Since salts are ionic they disrupt salt bridges in proteins. The reaction of a heavy metal salt with a protein usually leads to an insoluble metal protein salt.

- This reaction is used for its disinfectant properties in external applications. For example $AgNO_3$ is used to prevent gonorrhea infections in the eyes of new born infants. Silver nitrate is also used in the treatment of nose and throat infections, as well as to cauterize wounds.

Protein denaturation

- Heavy Metal Salts

- Mercury salts administered as Mercurochrome or Merthiolate have similar properties in preventing infections in wounds.

- This same reaction is used in reverse in cases of acute heavy metal poisoning. In such a situation, a person may have swallowed a significant quantity of a heavy metal salt. As an antidote, a protein such as milk or egg whites may be administered to precipitate the poisonous salt. Then an emetic is given to induce vomiting so that the precipitated metal protein is discharged from the body.

- Heavy metals may also disrupt disulfide bonds because of their high affinity and attraction for sulfur and will also lead to the denaturation of proteins.

Protein denaturation

Thermal denaturation

Ordered molecular systems are usually more stable at low temperatures, where thermal fluctuations are suppressed, but native proteins tend to be most stable at a temperature, T^*, near room temperature.

As a consequence, the denaturated state population can be increased either by heating (for $T > T^*$) or by cooling (for $T < T^*$). At sufficiently high or low temperatures, the native protein is thus denatured.

These two ways of disrupting the native protein conformation are known as heat denaturation and cold denaturation, respectively.

Monika Davidovic, Carlos Mattea, Johan Qvist, and Bertil Halle, J. AM. CHEM. SOC. 2009, *131*, 1025–1036

Protein denaturation

Overall, protein stability depends on the balance between enthalpy and entropic changes.

For globular proteins, the free energy of unfolding is commonly found to be positive between about 0 °C and 45 °C. It decreases through zero when the temperature becomes either hotter or colder, with the thermodynamic consequences of both cold and heat denaturation.

Heat denaturation is primarily due to the increased entropic effects of the non-polar residues in the unfolded state.

The hydration of the internal non-polar groups is mainly responsible for cold denaturation as their energy of hydration is greatest when cold. Thus, it is the increased natural structuring of water at lower temperatures that causes cold destabilization of proteins in solution.

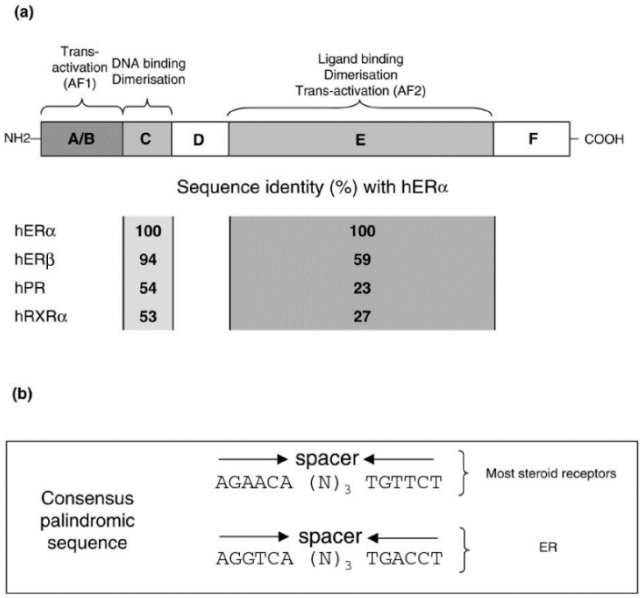
Protein families

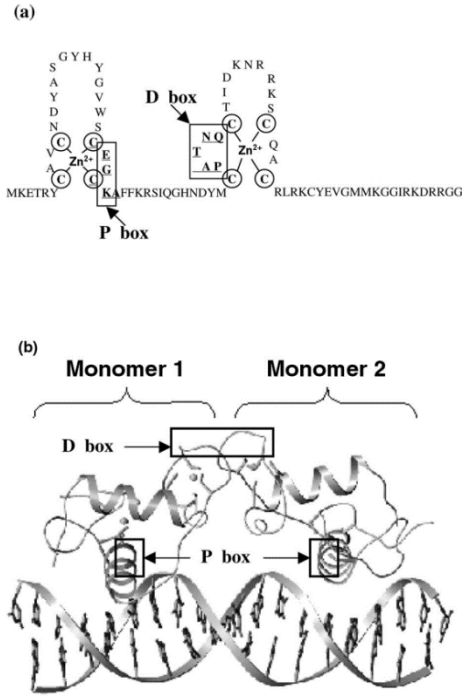
- Protein organization in domains
- Globular proteins
- Fibrous proteins
- Membrane proteins
- Intrinsically disordered proteins
- Multi-protein complexes

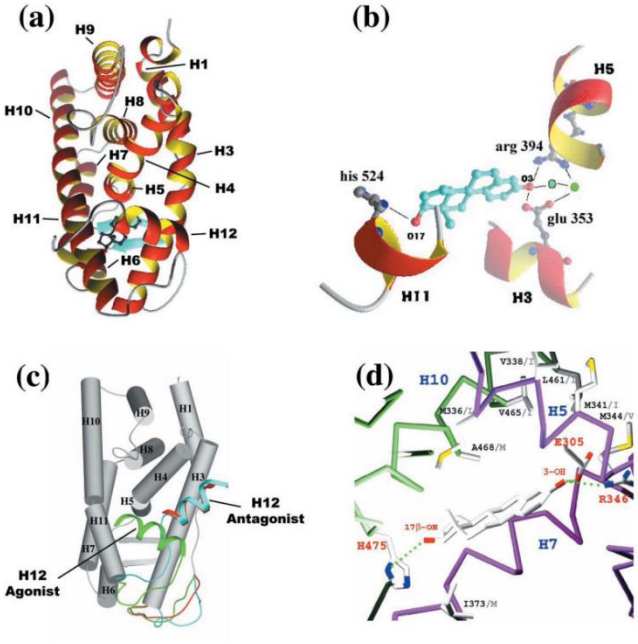
Proteins organization into domains

- Protein organization in structural and functional domains
  - Conserved part of a protein sequence forming an independent structure that can evolve, operate and exist independently of the rest of the protein chain.
  - Each domain forms a compact three-dimensional structure and is often independently stable and folded.
  - Many proteins consist of several structural domains.
  - A single domain can appear in a variety of different proteins.
  - Molecular evolution uses domains as building blocks, and these can be recombined in different arrangements to create proteins with different functions.
  - The domains length varies between about 25 to 500 amino acids. The shorter domains like zinc fingers are stabilized by metal ions or disulfide bridges.
  - Molecular evolution uses domains as building blocks, and these can be recombined in different arrangements to create proteins with different functions.
  - Domains often form functional units.

Proteins organization into domains: nuclear receptors





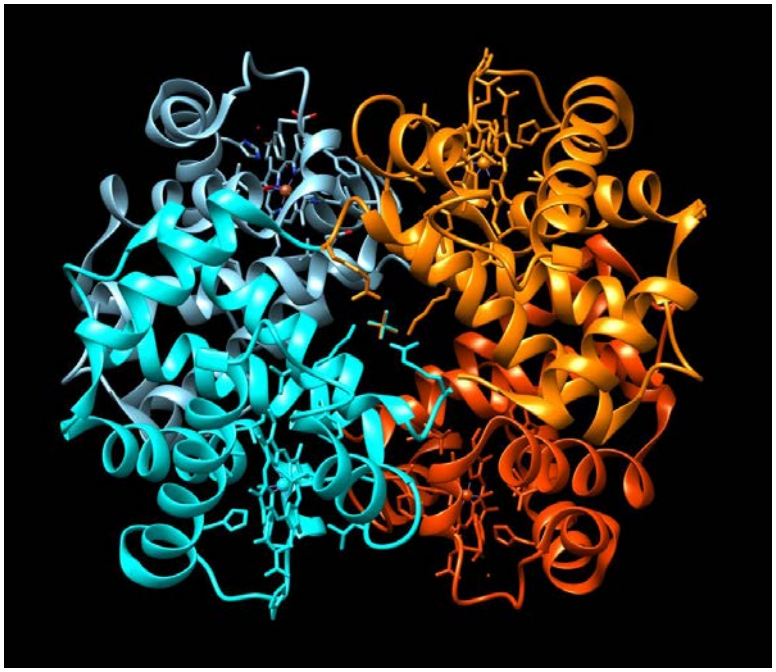


Marc Ruff, Monique Gangloff, Jean Marie Wurtz and Dino Moras. Estrogen receptor transcription and transactivation. Structure–function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res* 2000, 2:353–359

Globular Proteins

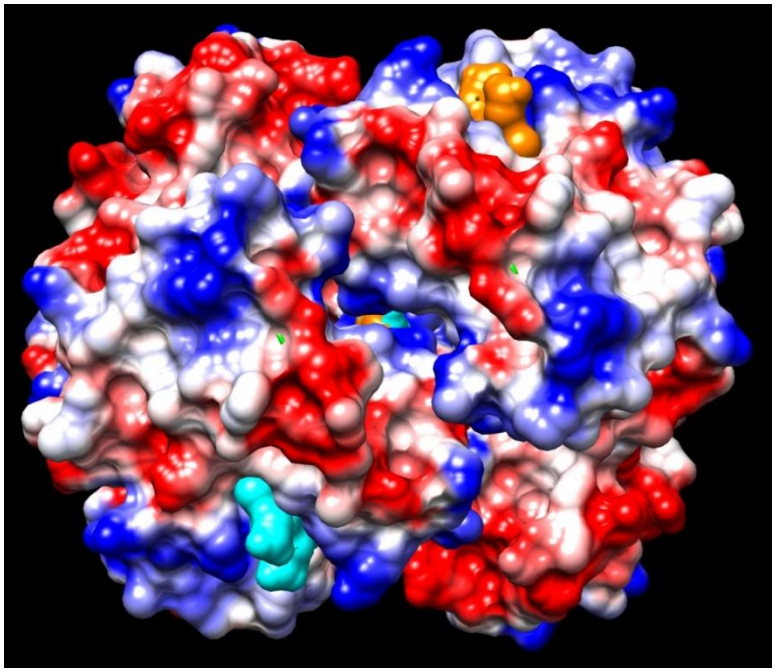
- Globular proteins are generally soluble spherical proteins in water (they actually form a colloid) in contrast to fibrous or membrane proteins.
- The spherical structure is induced by the tertiary structure of the protein. Hydrophobic amino acids are directed to the Interior of the molecule, while the hydrophilic are directed outwards, allowing dipole-dipole interactions with the solvent, which explains the solubility of the molecule.
- Compact structure.
- Soluble in water.
- The secondary structure is complex and contains a mixture of alpha helices, beta sheets and loops.
- The Quaternary structure is maintained by non-Covalent bonds.
- Function in all aspects of the metabolism (enzymes, transport, immune protection, hormones...).

Globular proteins



Human Hemoglobin
Sheet representation
Hetero tetramer α2β2

B. Shaanan, J. Mol. Biol., 1983



Human Hemoglobin
Representation of surface
electrostatic potential
Red = negatives charges
Blue= positives charges

Fibrous proteins

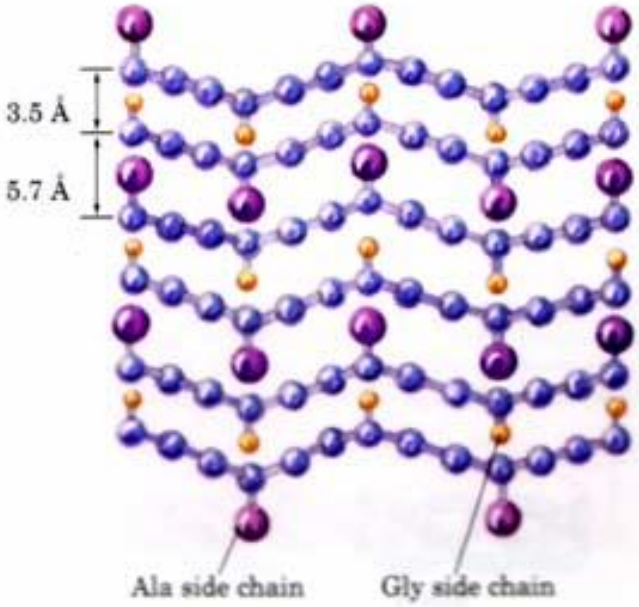
- Extended structure
- Insoluble in aqueous solvents or in Lipid Bilayers
- One type of secondary structure
- The quaternary structure is usually maintained by covalent bridges
- Structural function in the body or cell (tendons, bone, muscle, ligaments, hair, skin)
- Fibrous protein families
  - Collagen (the most abundant proteins in vertebrates, connective tissues: cartilage)
  - Keratin (hair, nails, feathers…)
  - Elastin (ligaments, blood vessels)
  - Fibroid (silk, spider webs)

Fibrous proteins: fibroid

figure 6–14

Structure of silk. The fibers used to make silk cloth or a spider web are made up of the protein fibroin. **(a)** Fibroin consists of layers of antiparallel β sheets rich in Ala (purple) and Gly (yellow) residues. The small side chains interdigitate and allow close packing of each layered sheet, as shown in this side view. **(b)** Strands of fibroin (blue) emerge from the spinnerets of a spider in this colorized electron micrograph.

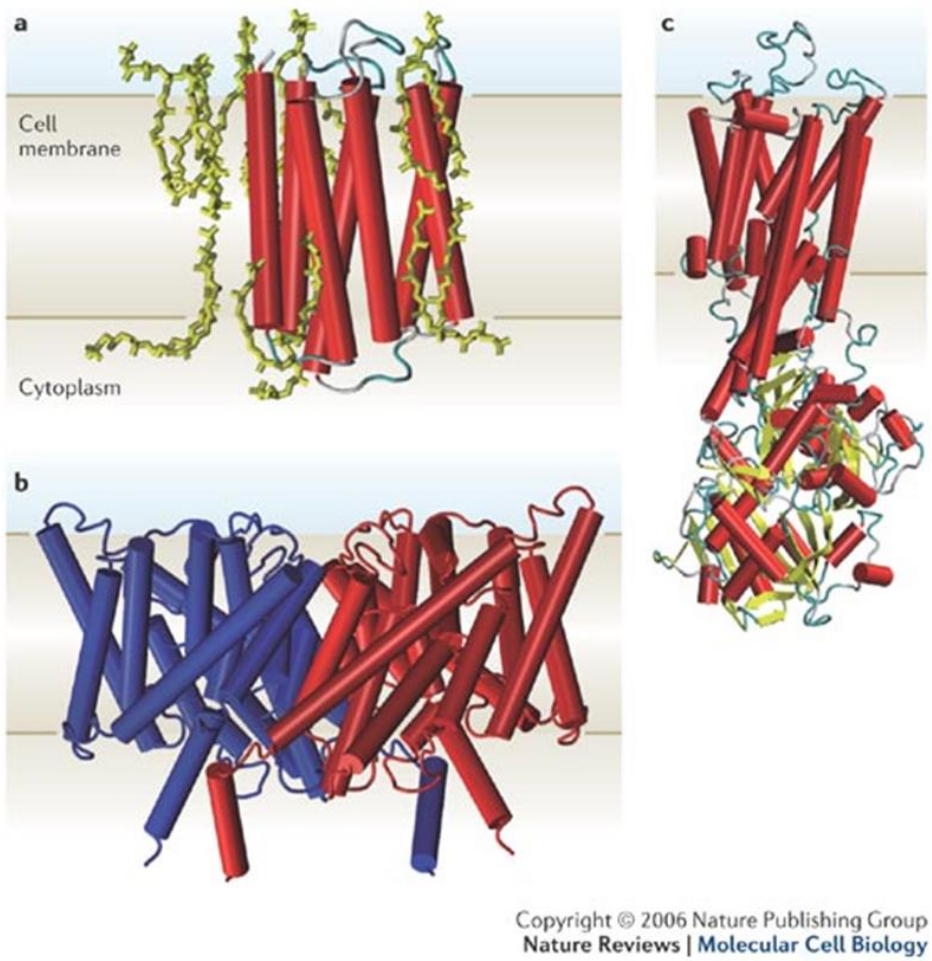
SOURCE: Nelson, D.L. and Cox, M.M. 2003. Lehninger Principles of Biochemistry, 3rd ed. Worth Publishers, New York, NY. p. 174



Membrane proteins

- Membrane proteins are attached or associated with cell membranes or organelles
- They are the target of more than 50% of the modern medical drugs
- 20 to 30% of the genes of most genomes encode membrane proteins.
- Function of membrane proteins
  - Membrane receptors relay signals between the outside and inside of the cells or organelles
  - Transport proteins move molecules and ions across the membrane
  - Membrane enzymes
  - Molecules for cell adhesion allow cells to identify themselves (immune response…)
- Topology of membrane proteins
  - Integral membrane proteins are permanently attached to the membrane and can be separated using detergents, nonpolar solvents or denaturing agents. Polytopic integral proteins (transmembrane proteins) cross the membrane and the monotopic integral proteins are only at one side of the membrane.
  - Peripheral membrane proteins are attached temporarily to the membrane to the lipid bilayer or integral proteins by hydrophobic and Ionic interactions. They can be separated by changes in pH or ionic strength.

Membrane proteins

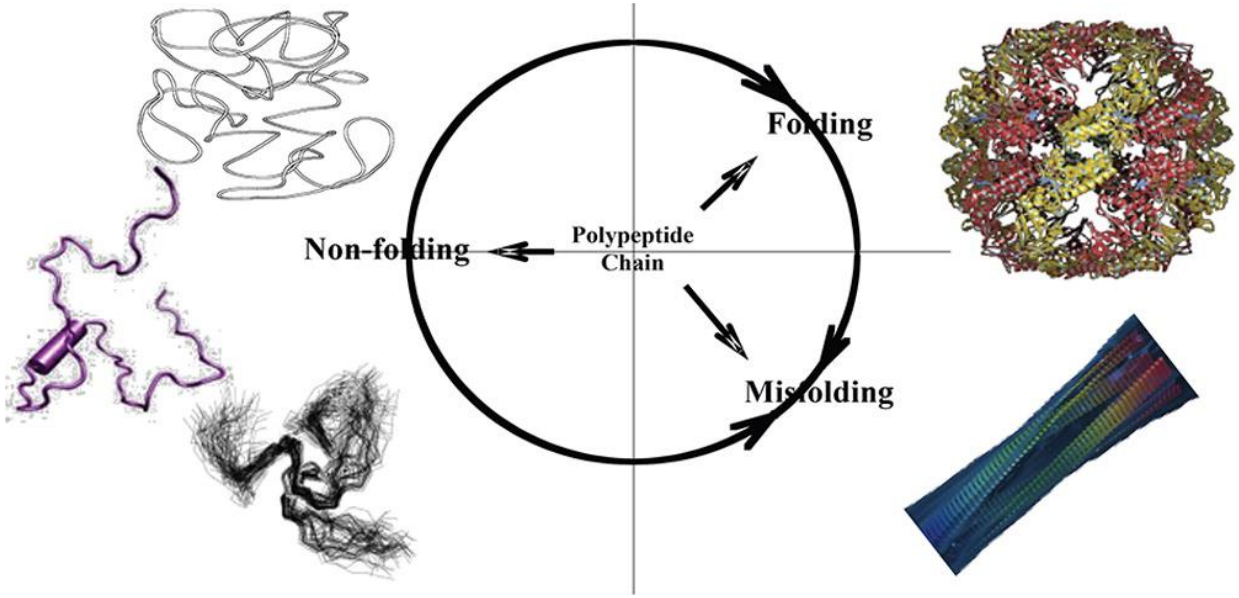


Helix-bundle membrane proteins.
a. Bacteriorhodopsin83 (Protein Data Bank (PDB) accession code 2BRD). The seven transmembrane helices are shown in red and co-crystallizing membrane phospholipids are shown in yellow.
b. The Escherichia coli ClC Cl–/H+ antiporter6 (PDB accession code 1KPK) is a homodimer (one subunit is shown in blue and the other in red). There is one Cl– channel in each monomer.
c. Bovine $Ca2^{+}$-ATPase10 (PDB accession code 1T5S) colored according to secondary structure (helices are red; β-strands are yellow).

Heijne *Nature Reviews Molecular Cell Biology* **7**, 909–918 (December 2006) | doi:10.1038/nrm2063

Protein production in heterologous organism

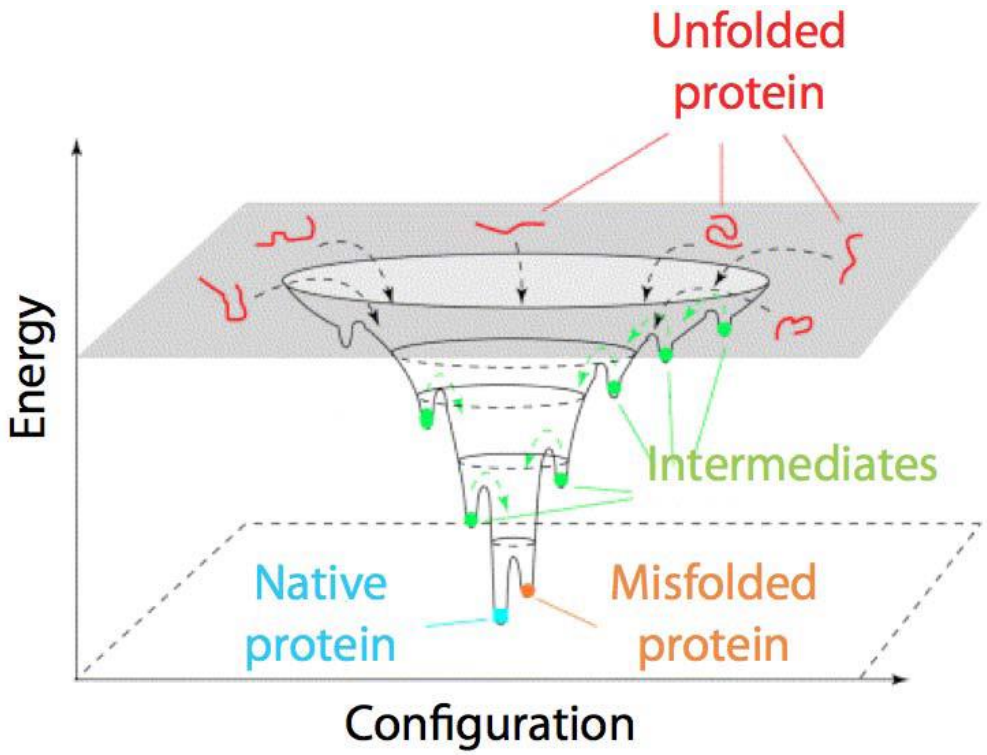
The protein folding



Folding, nonfolding, or misfolding of a polypeptide chain. A polypeptide chain can either gain a well-defined 3-D structure or stay nonfolded in the three different states (i.e., Random Coil, Pre Molten Globule, and Molten Globule, or it also can misfold to form amyloid-like fibrils.

Johnny Habchi; Peter Tompa; Sonia Longhi; Vladimir N. Uversky; *Chem. Rev.* **2014,** 114, 6561-6588.

Protein folding: the folding funnel



- There is no single route for folding, but a large number of structures that follow a multidimensional funnel up to the native structure
- Progress in the funnel is accompanied by an increase of the native structures during the process of folding

The folding of proteins in prokaryotes

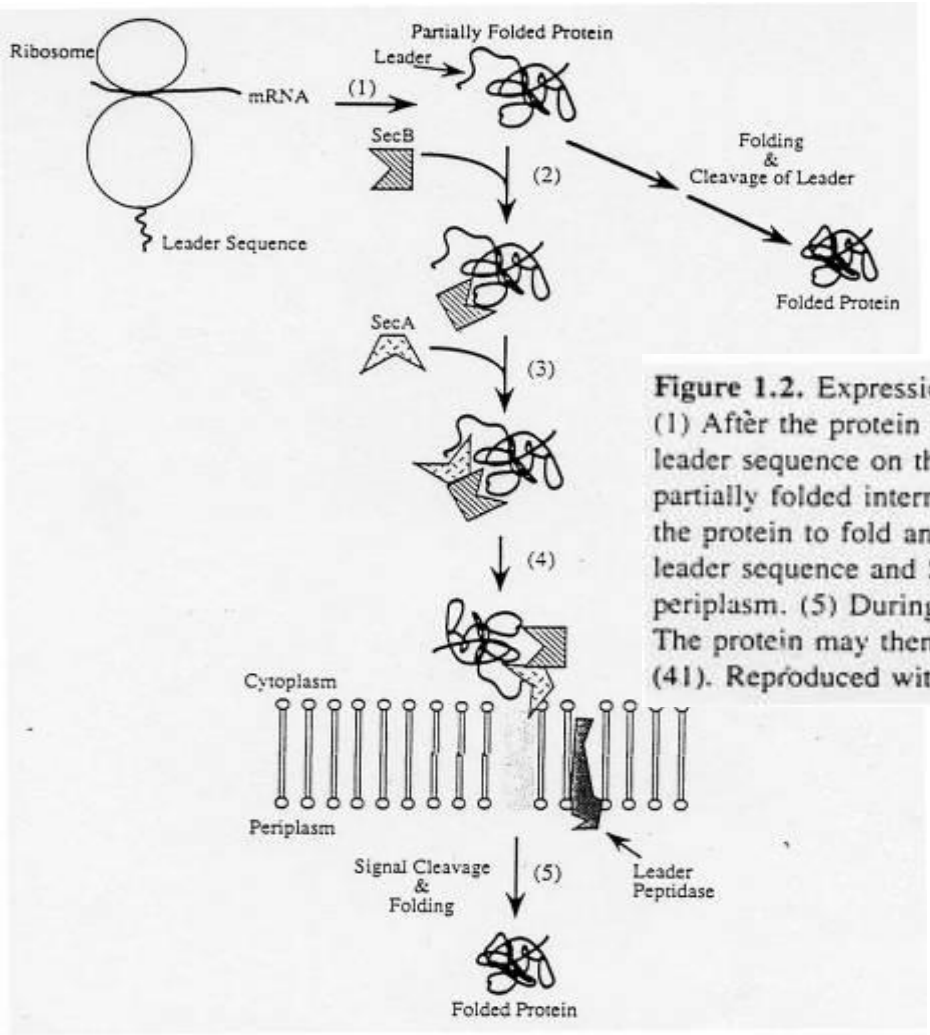


Figure 1.2. Expression, translation, and folding in prokaryotes [Wickner et al. (48)]. (1) After the protein is translated by the cytoplasmic ribosomes, it slowly folds. (2) The leader sequence on the protein interferes with the folding, allowing SecB to bind to a partially folded intermediate. Alternatively, the leader sequence can be removed, allowing the protein to fold and not be exported. (3) For exported proteins, SecA binds to the leader sequence and SecB, and (4) facilitates translocation of the protein into the periplasm. (5) During translocation, the leader sequence is cleaved by a leader peptidase. The protein may then fold or remain in a partially folded state. [From Cleland and Wang (41). Reproduced with permission from VCH Publishers, Inc.]

The folding of proteins in eukaryotes

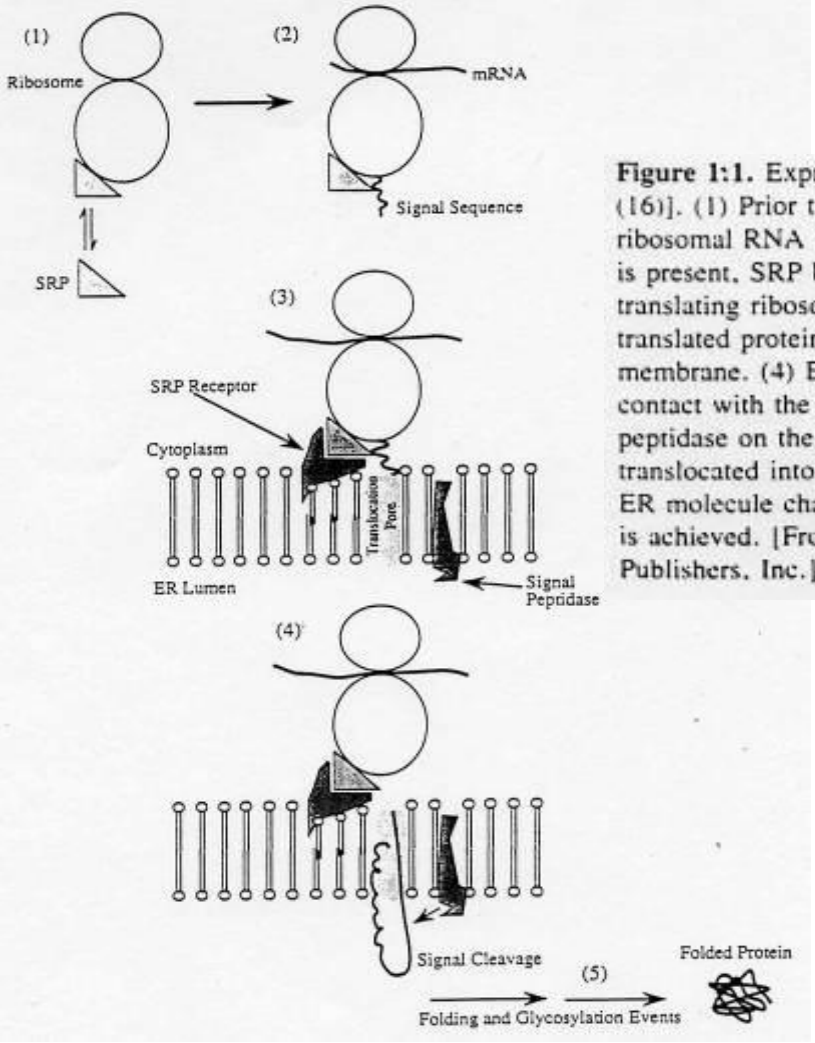


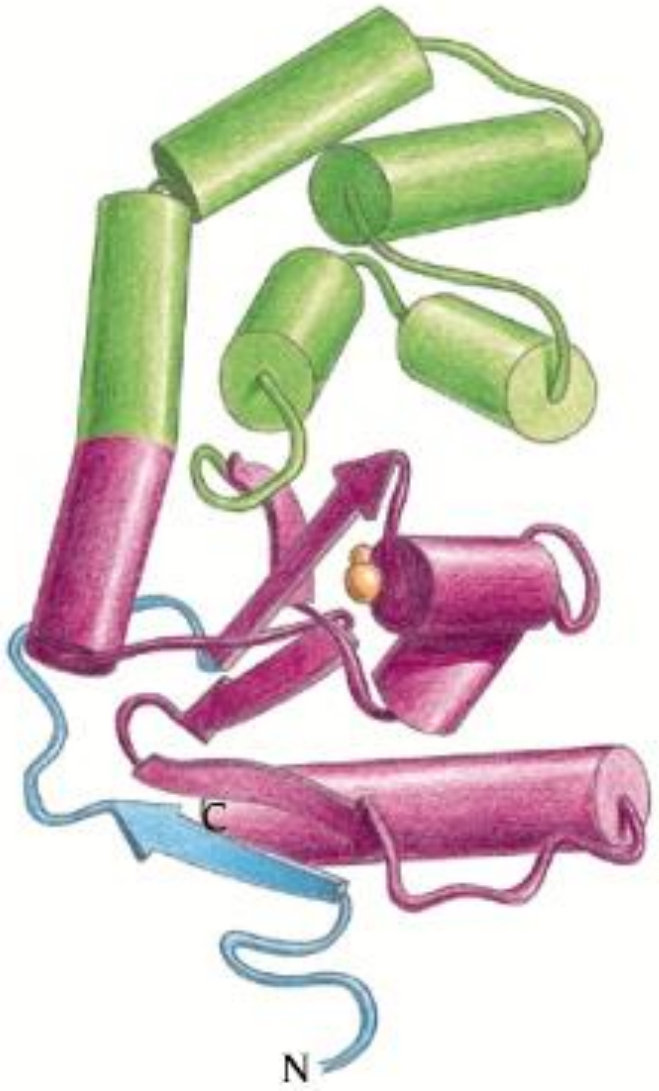
Figure 1:1. Expression, translation, and folding in eukaryotes [Nothwehr and Gordon (16)]. (1) Prior to translation, the signal receptor particle (SRP) is in equilibrium with ribosomal RNA in the cytoplasm. (2) After translation has begun and the signal sequence is present, SRP binds to the signal sequence and the ribosome. (3) SRP then guides the translating ribosome to the ER membrane where it binds to the SRP receptor. The translated protein with signal sequence is translocated through a pore in the ER membrane. (4) Eventually, the signal sequence forms a helical structure that remains in contact with the ER membrane. The signal sequence is then removed by a signal peptidase on the lumen side of the ER membrane. (5) The remaining protein is translocated into the ER lumen as it is produced. The resulting protein may interact with ER molecule chaperones and glycosylation enzymes. Finally, the native protein structure is achieved. [From Cleland and Wang (41). Reproduced with permission from VCH Publishers, Inc.]

Proteins involved in folding

- The chaperones proteins are required if a protein does not fold correctly. They help in the correct folding of proteins.
  - Proteins Disulfides Isomerases (PDI)
  - Peptidyl Prolyl Isomerases (PPI)
  - Much of chaperones are "Heat Shock Proteins", proteins expressed in response to elevations of temperatures or other chemical stress. They are hydrophobic in the interior and hydrophilic towards the solvent.

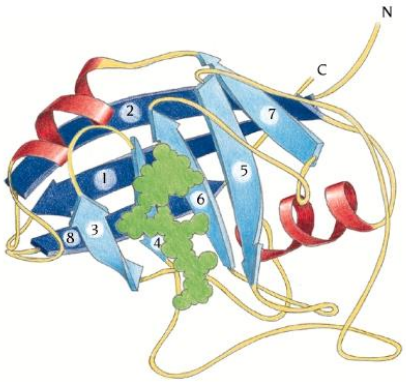
Protein Disulfides Isomerases (PDI)

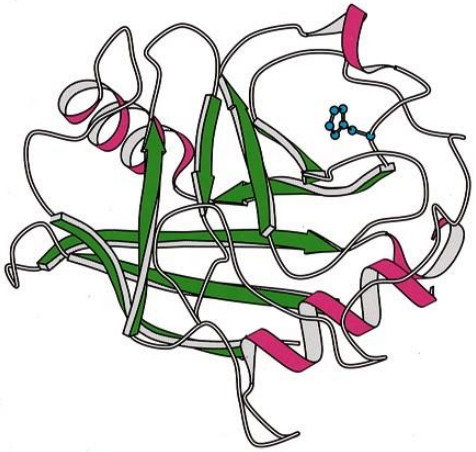
- The formation of correct disulfide bonds in proteins during folding is catalyzed by the PDIs.
- The PDIs bind preferentially to the cysteine-containing peptides. They have a broad specificity for the folding of various proteins containing disulfide bridges.
- By interchanging the disulfide bridges, the PDIs allow that the proteins rapidly find the thermodynamically more stable disulfide bridges.

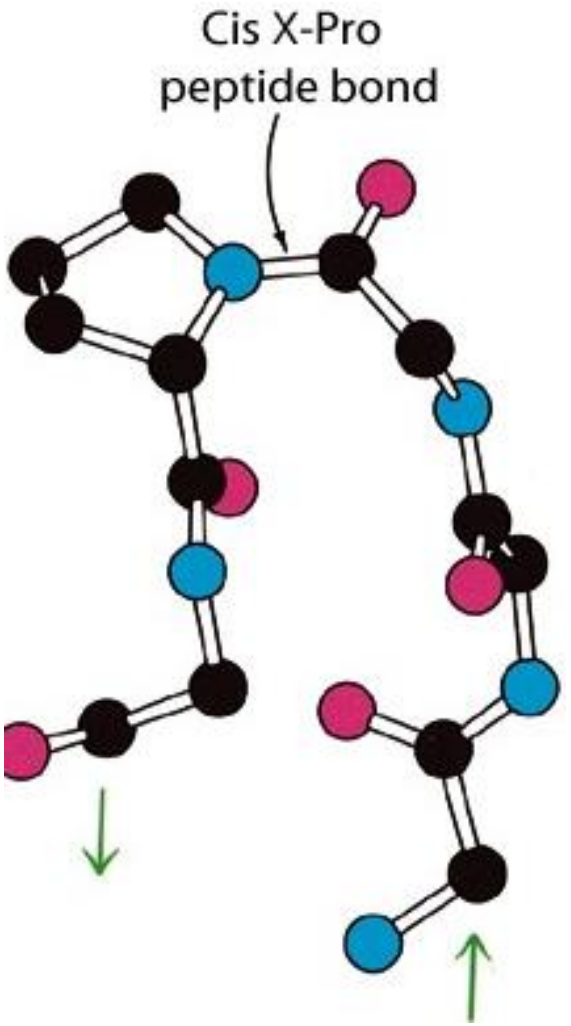


Peptidyl Prolyl Isomerases (PPI)

- The peptide bond in proteins is almost always in trans conformation, but 6% of X - pro peptide bonds are cis.
- The isomerization of proline is limiting for the folding of proteins *in vitro*.
- The PPIs accelerate more than 300 times the cis-trans isomerization by distorting the peptide bond so that C, O and N atoms are no more plan.







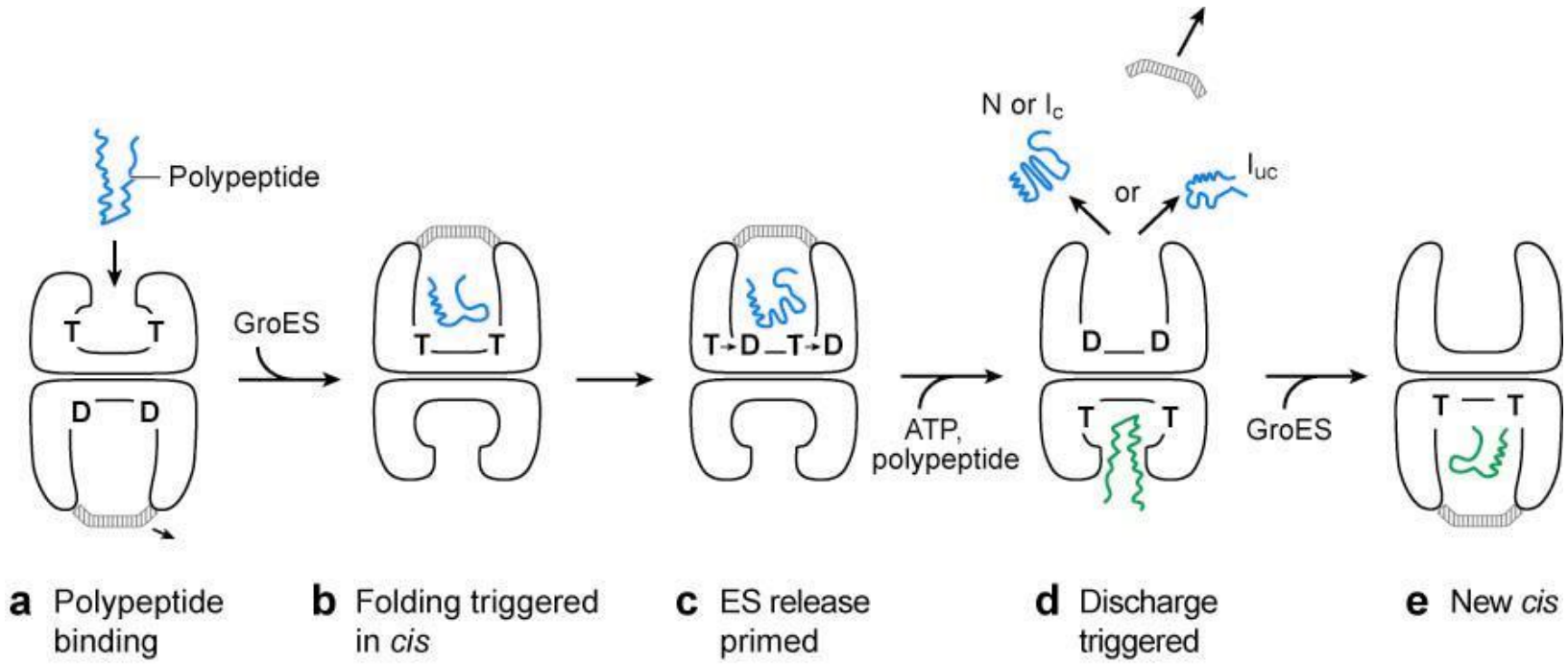
Molecular chaperones

- During synthesis on the ribosome, polypeptides fold spontaneously with molecular chaperones which are involved in their folding in vivo and are related to heat shock proteins.
- Group I chaperonins are found in bacteria as well as organelles of endosymbiotic origin: chloroplasts and mitochondria
- Group II chaperonins are found in the eukaryotic cytosol and in archaea
- The main families of HSPS are:
  - « small HSPs » from 10,000 to 30,000 MW (hsp26/27)
  - hsp40
  - hsp60 (e.g. GroEL chez *E. coli,* CCT in eukaryotes)
  - hsp70 (DnaK in *E. coli*)
  - hsp90
  - hsp100

Function of molecular chaperones

- Minimize denaturation of proteins due to the heat or stress (renaturation/degradation)
- Facilitates the correct folding of proteins by minimizing the aggregation and other folding problems
- Bind to nascent polypeptide synthesis to prevent a premature folding.
- Facilitates membrane translocation/import by preventing folding prior membrane translocation
- Facilitates the assembly/disassembly of multiprotein complexes

Chaperonin: GroEL



Horwich AL, et al. 2007.
Annu. Rev. Cell Dev. Biol. 23:115–45

T=ATP, D=ADP

Protein purification: how to maintain a protein soluble and functional

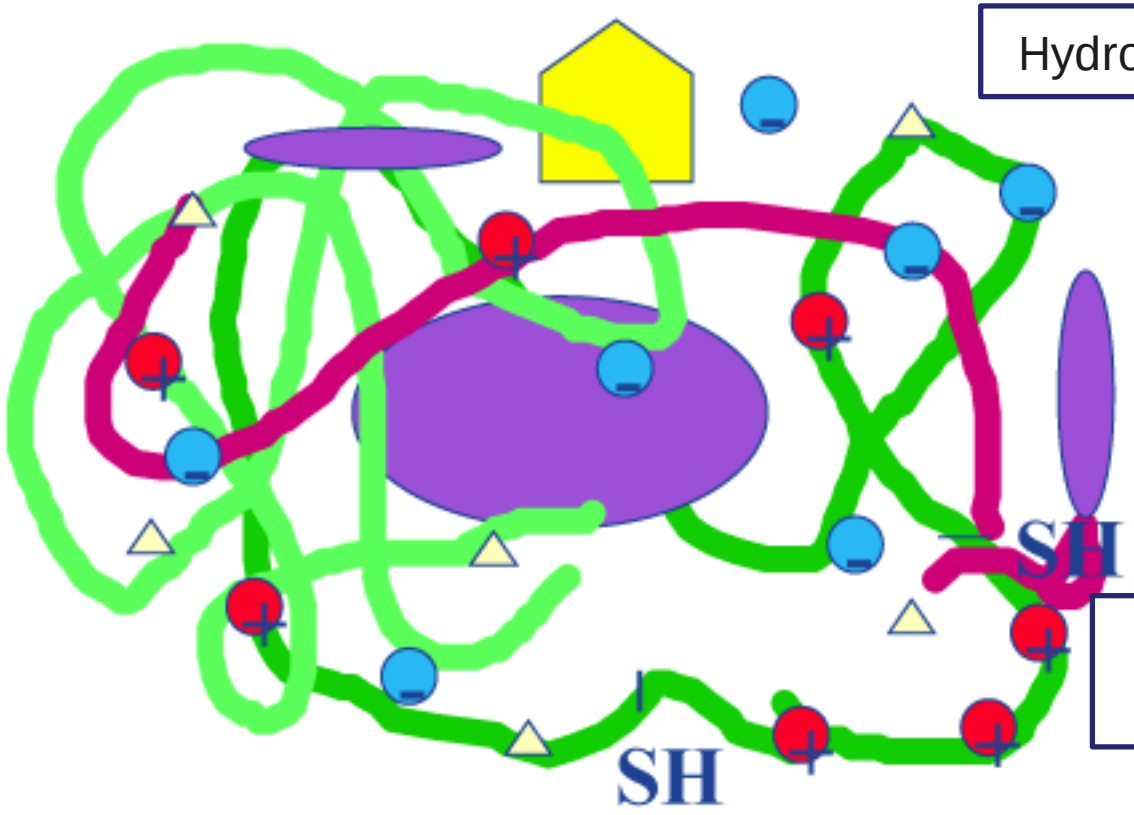
Purification

Size

Charge

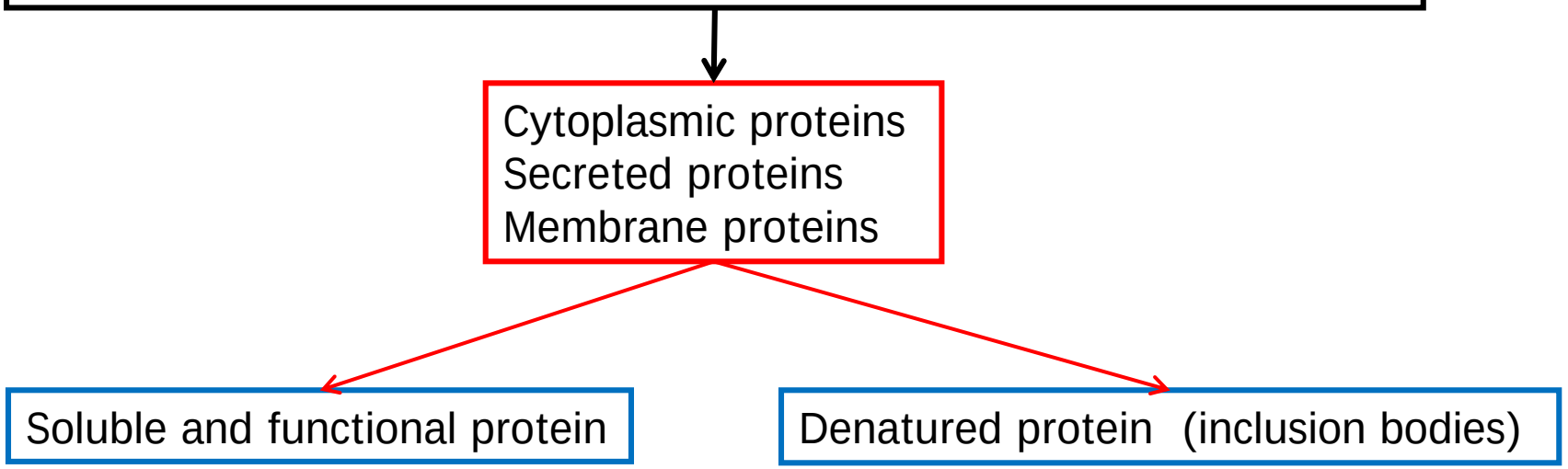
Hydrophobicity

Biological activity

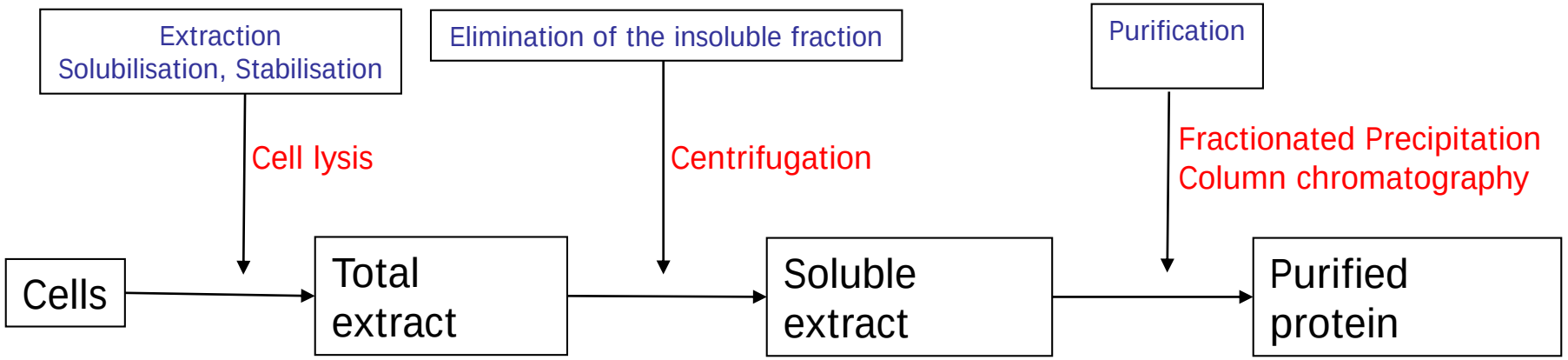


Purification

Purification of a protein from a homologous or heterologous organism is to separate the protein of interest from a complex mixture consisting of all cellular components (fatty acids, lipids, carbohydrates, other proteins...)



Purification



Measuring the amount of:
Total protein
Specific protein

Purity analysis

Buffer exchange
Concentration and dilution

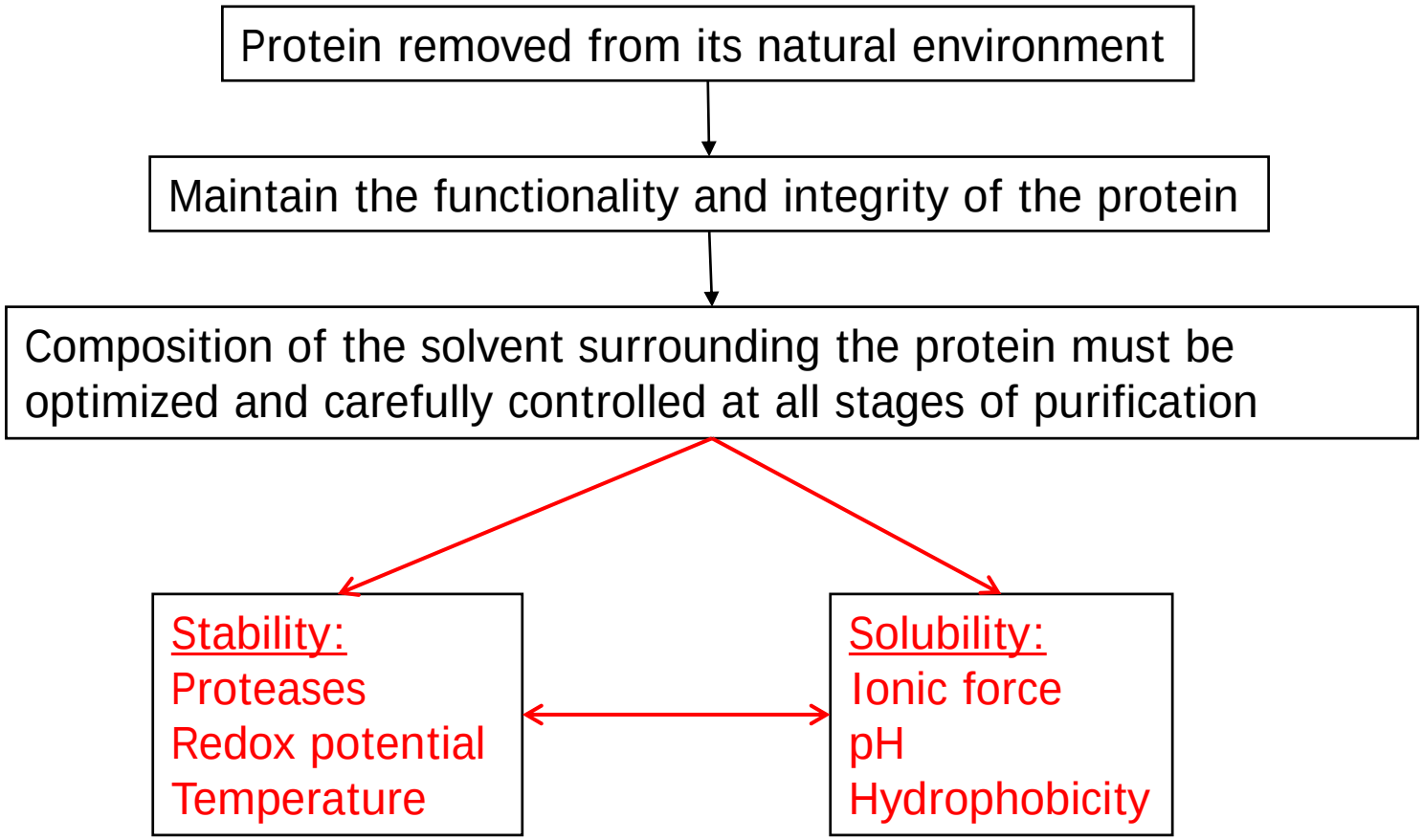
- Adapt the purification to the protein
- Adapt the protein to the purification: fusion proteins for affinity purification (HIS, GST, MBP, biotinilated peptide, STREP, FLAG, ...)

Total extract → Affinity → Tag removal → Affinity → Gel filtration

Cloning and expression

- Protein production in prokaryotic cells
  - E. Coli cells
- Protein production in eukaryotic cells
  - Yeast cells
  - Insect cells
  - Mammalian cells
- Cell free systems

Stabilisation, solubilisation



Stabilisation

- Find the right buffer composition to maintain the integrity of the purified protein
- The release of intra cellular components by grinding releases proteases
- Use of protease inhibitors
- Work at 4°C

Stabilization: REDOX potential

- Potential REDOX (use of reducing agent or oxydo/reducing mixtures)
  - Preventing the formation of inter molecules disulfides bridges
  - Preventing the formation of incorrect intra protein disulfide bridges
  - Maintaining correct disulfide bridge

Solubility

- Ionic force (solubility depending on the concentration and nature of salt)
- Detergents
- Additives (glycerol, fluoro-alcohol, Arg/Glu,...) for protein solubilization
- pH (solubility vs. pH, using buffers)

Solubility: ionic force

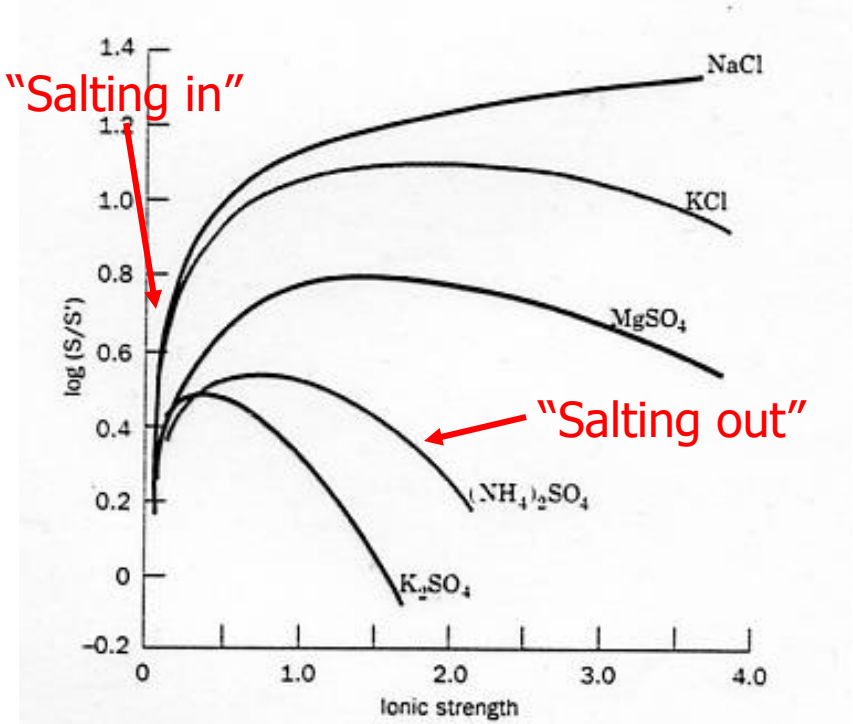


Figure 5-3
Solubility of carboxy-hemoglobin at its isoelectric point as a function of ionic strength and ion type. Here S and S′ are, respectively, the solubilities of the protein in the salt solution and in pure water. The logarithm of their ratios is plotted so that the solubility curves can be placed on a common scale. [After Green, A. A., *J. Biol. Chem.* **95**, 47 (1932).]

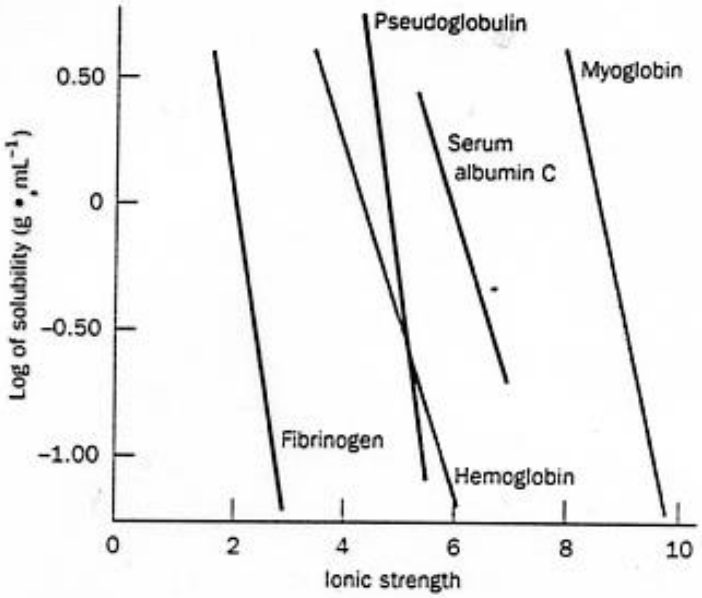


Figure 5-2
Solubilities of several proteins in ammonium sulfate solutions. [After Cohn, E. J. and Edsall, J. T., *Proteins, Amino Acids and Peptides, p.* 602, Academic Press (1943).]

Stabilisation, solubilisation

- Proteins denatured by contact with air - water interface (foam)
- If low amount, a significant fraction may be lost by adsorption to surfaces
- Minimize freezing
- Oxidation of cysteine: addition of reducers (βmercaptoethanol, DTT)
- pH stabilization by buffers
- Stabilization of ionic strength (NaCl, KCl)
- Addition of detergents in the case of aggregation due to hydrophobic interactions
- Proteins denature slowly by chemical or proteolytic degradation: addition of EDTA to remove heavy metals, inhibitors of proteases, purification at 4° C.
- Avoid bacterial contamination: NaN_3

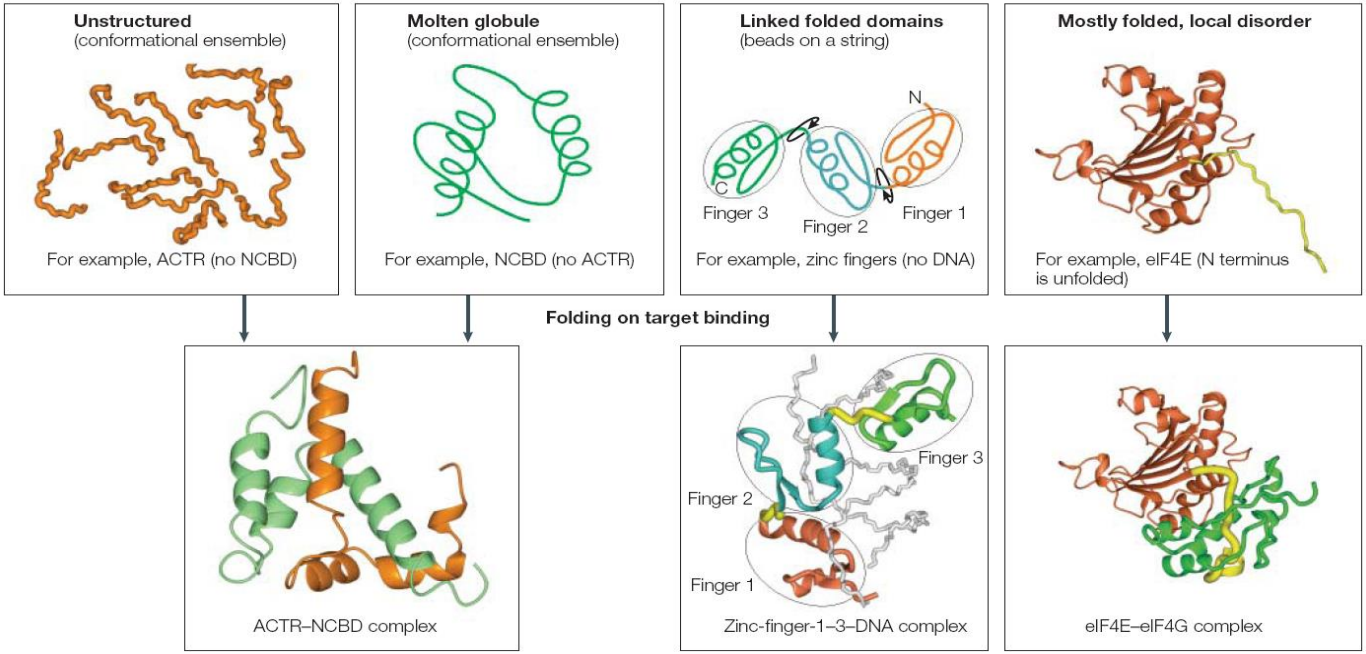
Typical composition of a solution used for cells breaking:

NaCl 100-200 mM, EDTA 0.1 mM, TRIS/HCl pH=7.5, DTT 2 mM, PMSF 0.1 mM

Intrinsically Disordered Proteins (IDP)

Definitions: (Keith Dunker - Indiana University)

Many proteins contain regions that lack specific 3-D structure; indeed some proteins lack specific 3-D structure in their entireties under physiological conditions and yet carry out biological function. Such proteins and regions have been called natively unfolded, intrinsically unstructured, naturally disordered, …



H. Jane Dyson and Peter E.Wright, Nature Reviews In Molecular Cell Biology, 2005, 6, 197-208

Intrinsic disorder and protein interaction networks

Dunker et al, Flexible nets: The roles of intrinsic disorder in protein interaction networks, FEBS J. 2005, 272, 5129-5148)

- plasticity of the disorder regions allows the binding of multiple partners
- Few proteins bind to many partners (Hub proteins) but most proteins bind to few partners
- Many disordered regions are modified post-transcriptionaly (Phosphorylation, Acetylation, ubiquitination, proteolytic cleavage)

Intrinsically disordered proteins: function, folding, and flexibility

IDPs participate in important regulatory functions in the cell, including transcription, translation, the cell cycle, and numerous signal transduction events. Disordered proteins often undergo coupled folding and binding transitions upon interaction with their cellular targets.

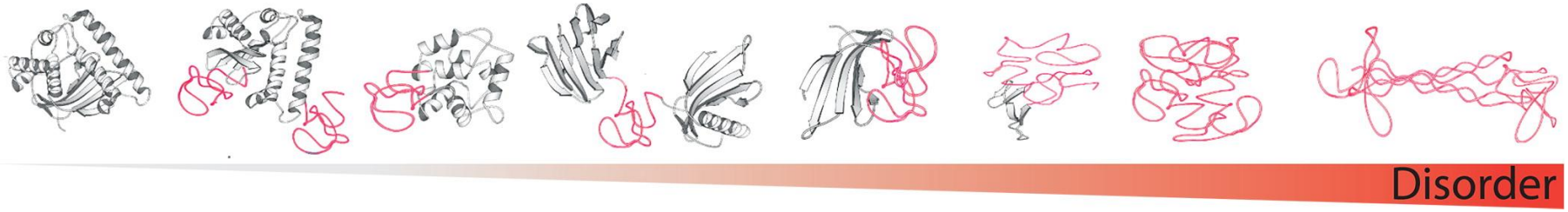
The lack of stable globular structure confers numerous functional advantages, but not without cost; many disordered proteins are associated with amyloid disease and with chromosomal translocations in cancer.

Intrinsically disordered proteins occupy a continuum of "conformational space", ranging from highly unstructured, through molten globule, to local disorder within an otherwise folded domain.

Intrinsically disordered proteins: function, folding, and flexibility

While many proteins need to adopt a well-defined structure to carry out their function, a large fraction of the proteome of any organism consists of polypeptide segments that are not likely to form a defined three-dimensional structure, but are nevertheless functional. These protein segments are referred to as intrinsically disordered regions (IDRs).

Proteins sequences in a genome can be viewed as modular because they are made up of combinations of structured and disordered regions. Proteins without IDRs are called structured proteins, and proteins with entirely disordered sequences that do not adopt any tertiary structure are referred to as intrinsically disordered proteins (IDPs). The majority of eukaryotic proteins are made up of both structured and disordered regions, and both are important for the repertoire of functions that a protein can have in a variety of cellular contexts.

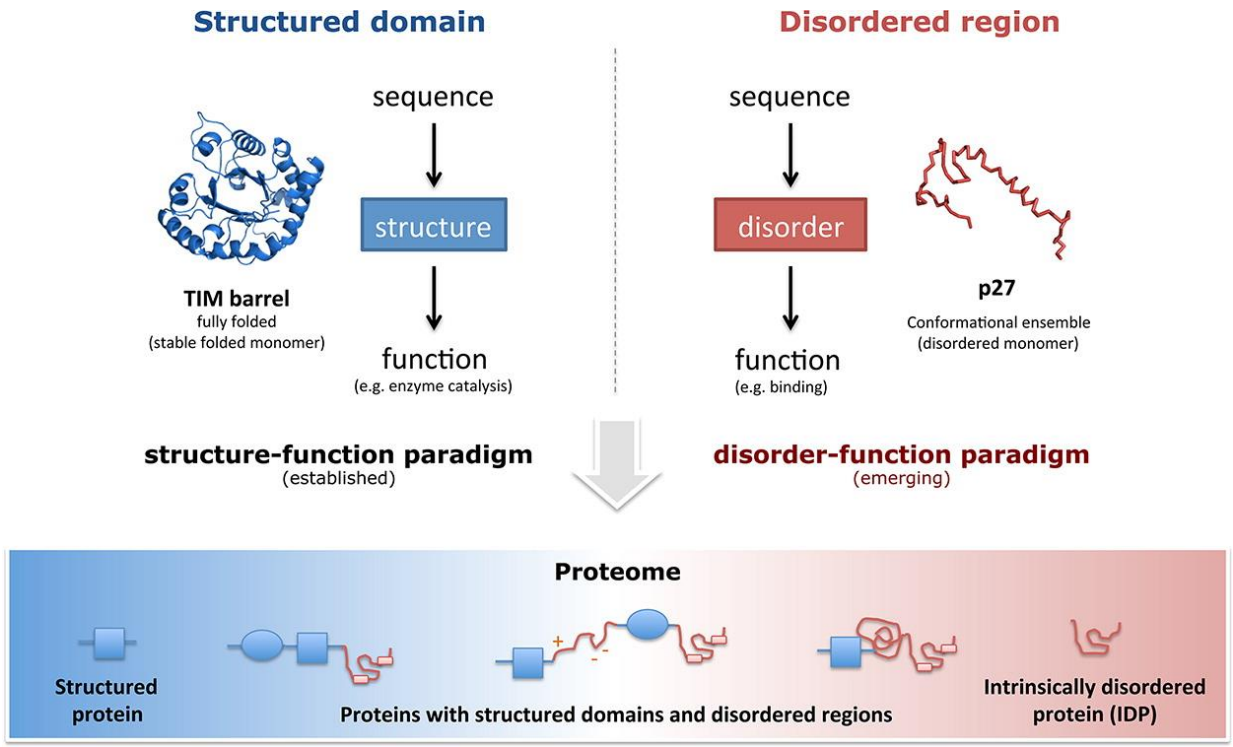


IDRs and IDPs are prevalent in eukaryotic genomes. For instance, 44% of human protein-coding genes contain disordered segments of >30 amino acids in length.

Johnny Habchi; Peter Tompa; Sonia Longhi; Vladimir N. Uversky; *Chem. Rev.* 2014, 114, 6561-6588.

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* 2014, 114, 6589-6631. DOI: 10.1021/cr400525m.

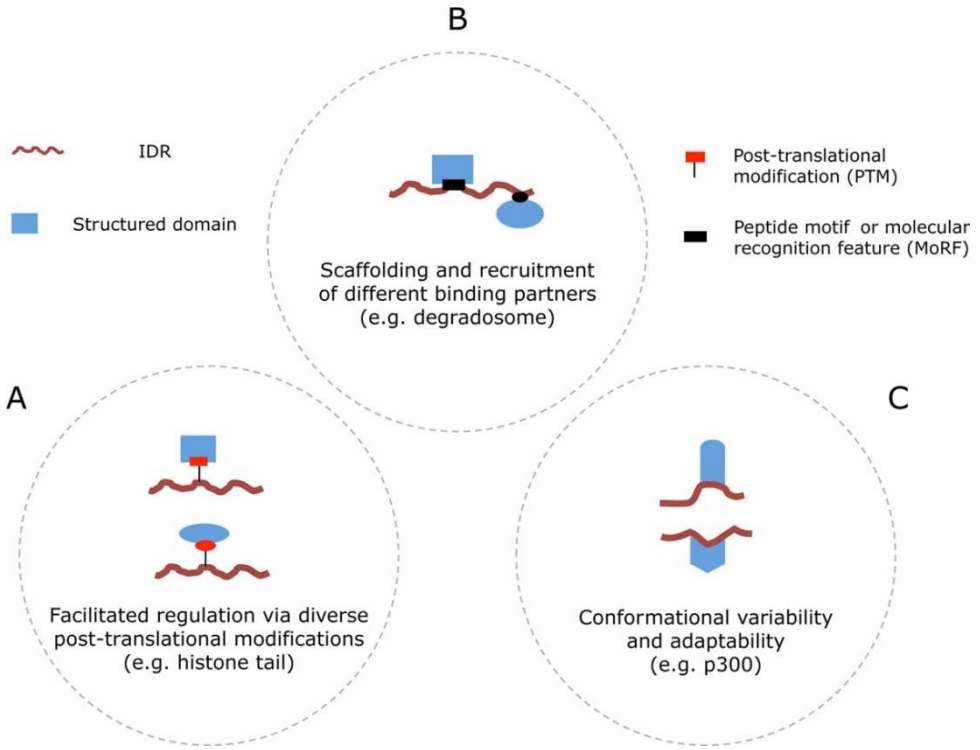
Intrinsically disordered proteins: function, folding, and flexibility



Structured domains and intrinsically disordered regions (IDRs) are two fundamental classes of functional building blocks of proteins. The synergy between disordered regions and structured domains increases the functional versatility of proteins.

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* 2014, 114, 6589-6631. DOI: 10.1021/cr400525m.

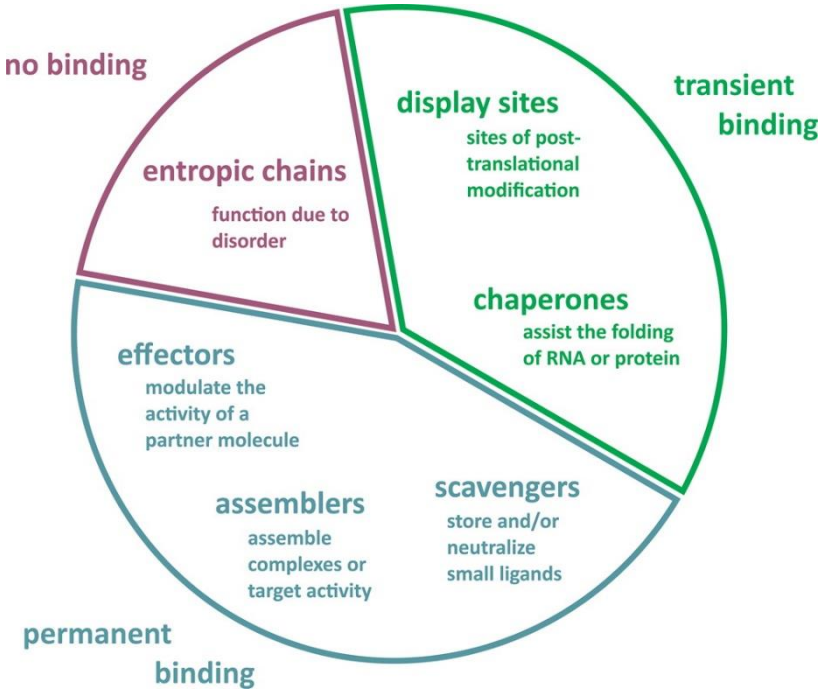
Intrinsically disordered proteins: function, folding, and flexibility



Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* 2014, 114, 6589-6631. DOI: 10.1021/cr400525m.

Functional classification of IDRs according to their interaction features. (A) The flexibility of IDRs facilitates access to enzymes that catalyze post-translational modifications and effectors that bind these PTMs. This permits combinatorial regulation and reuse of the same components in multiple biological processes. (B) The availability of molecular recognition features and linear motifs within the IDRs enables the fishing for ("fly casting") and gathering of different partners. (C) Conformational variability enables a nearly perfect molding to fit the binding interfaces of very diverse interaction partners. Context-dependent folding of an IDR can activate signalling processes in one case or inhibit them in another, resulting in completely different outcomes.

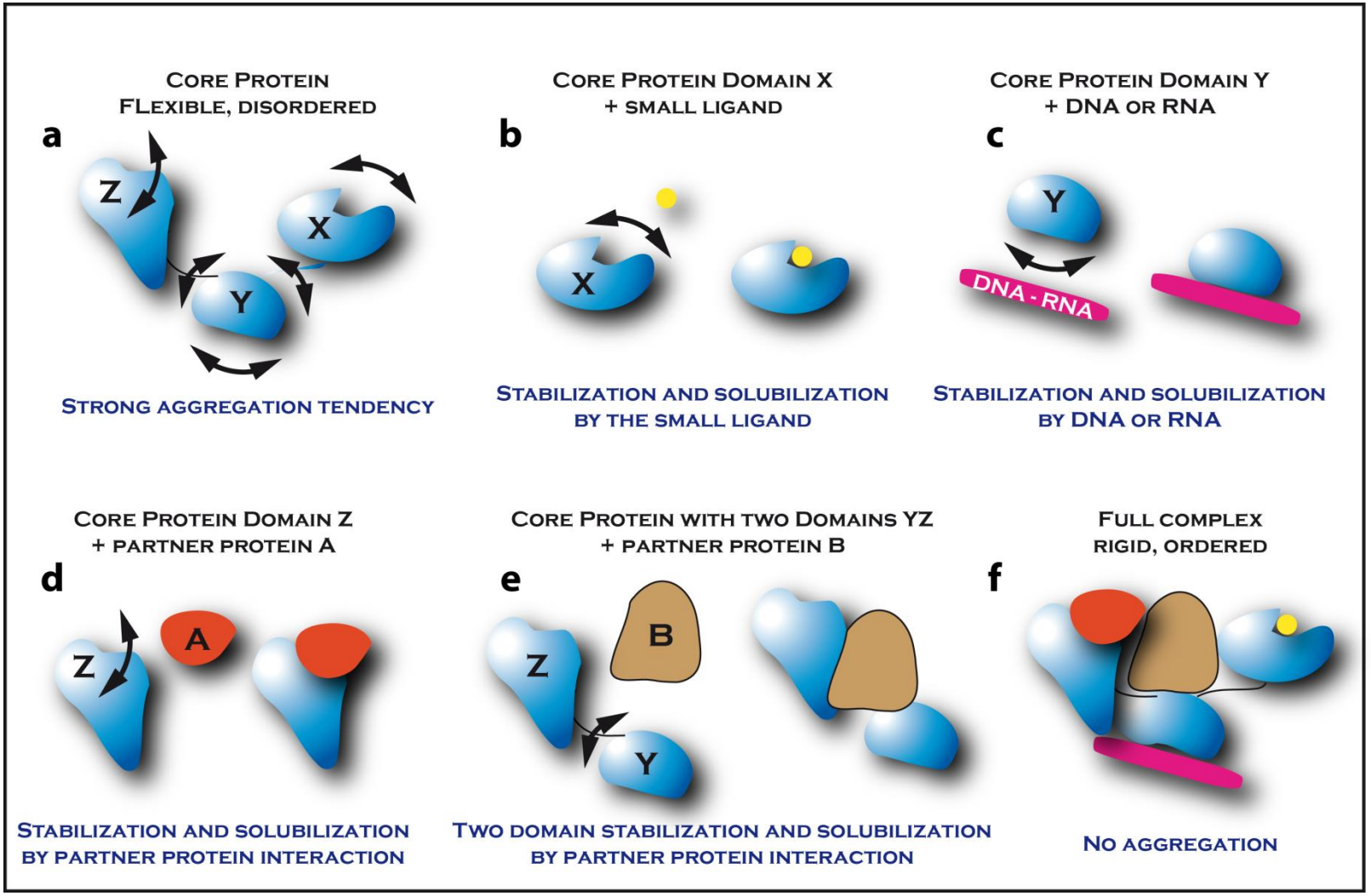
Intrinsically disordered proteins: function, folding, and flexibility



Functional classification scheme of IDRs. The function of disordered regions can stem directly from their highly flexible nature, when they fulfil entropic chain functions (such as linkers and spacers, indicated in dark-tone red), or from their ability to bind to partner molecules (proteins, other macromolecules, or small molecules). In the latter case, they bind either transiently as display sites of post-translational modifications or as chaperones (indicated in green), or they bind permanently as effectors, assemblers, or scavengers (indicated in dark-tone blue).

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* 2014, 114, 6589-6631. DOI: 10.1021/cr400525m.

From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

Current Methods for Function Prediction of Intrinsically Disordered Regions and Proteins

basis for method	description	method	Web site
linear motifs	annotation of well-characterized linear motifs, which can be mapped onto other protein sequences	ELM	http://elm.eu.org/
		MiniMotif	http://mnm.engr.uconn.edu/
	identification of putative uncharacterized motifs in protein sequences	SLiMPrints	http://bioware.ucd.ie/slimprints.html
		phylo-HMM	http://www.moseslab.csb.utoronto.ca/phylo_HMM/
		DiliMot	http://dilimot.russelllab.org/
		SLiMFinder	http://bioware.ucd.ie/slimfinder.html
PTM sites	resources of experimentally verified PTM sites, mostly phosphorylation	Phospho.ELM	http://phospho.elm.eu.org/
		PhosphoSite	http://www.phosphosite.org/
		PHOSIDA	http://www.phosida.com/
	identification and collection of peptide motifs that direct post-translational modifications	ScanSite	http://scansite.mit.edu/
		NetPhorest	http://netphorest.info/
		NetworKIN	http://networkin.info/
		PhosphoNET(	http://www.phosphonet.ca/

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* 2014, 114, 6589-6631. DOI: 10.1021/cr400525m.

Current Methods for Function Prediction of Intrinsically Disordered Regions and Proteins

molecular recognition features	collection of verified sequence elements that undergo coupled folding and binding	IDEAL	http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/
	prediction of sequences that undergo disorder-to-order transitions	MoRFpred	http://biomine.ece.ualberta.ca/MoRFpred/
		ANCHOR	http://anchor.enzim.hu/
intrinsically disordered domains	annotation of disordered protein domains, which can be detected by sequence profiles	Pfam	http://pfam.sanger.ac.uk/
other	prediction of gene ontology functions using protein sequence features such as intrinsic disorder	FFPred	http://bioinf.cs.ucl.ac.uk/psipred/
	function annotation of experimentally verified disordered protein regions	DisProt	http://www.disprot.org/
	predictions of disordered regions combined with information on MoRFs, PTM sites, and domains	D^2P^2	http://d2p2.pro/

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* 2014, 114, 6589-6631. DOI: 10.1021/cr400525m.