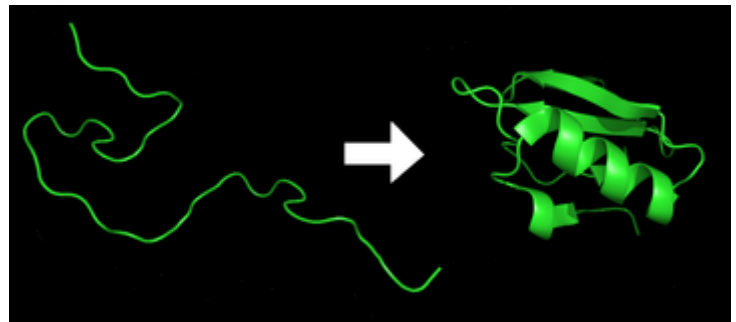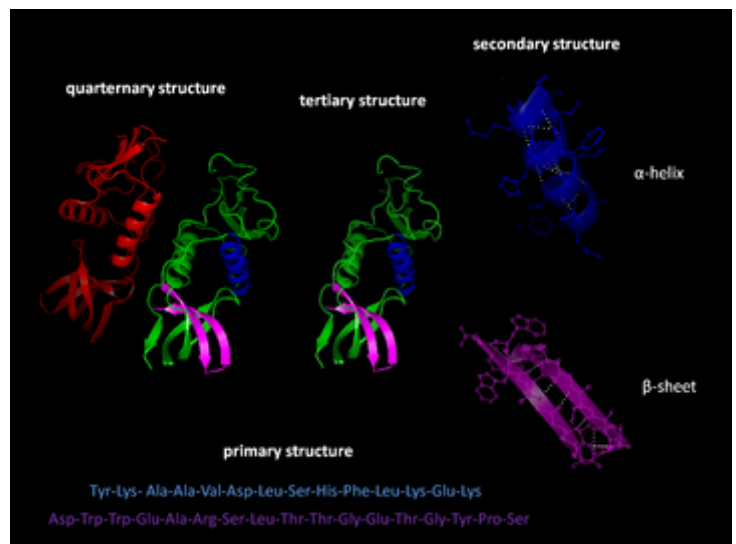# Protein folding

Protein folding is the physical process by which a protein chain is translated to its native three-dimensional structure, typically a "folded" conformation by which the protein becomes biologically functional. Via an expeditious and reproducible process, a polypeptide folds into its characteristic three-dimensional structure from a random coil.[1] Each protein exists first as an unfolded polypeptide or random coil after being translated from a sequence of mRNA to a linear chain of amino acids. At this stage the polypeptide lacks any stable (long-lasting) three-dimensional structure (the left hand side of the first figure). As the polypeptide chain is being synthesized by a ribosome, the linear chain begins to fold into its three-dimensional structure.

Protein before and after folding

Folding of many proteins begins even during translation of the polypeptide chain. Amino acids interact with each other to produce a well-defined three-dimensional structure, the folded protein (the right hand side of the figure), known as the native state. The resulting three-dimensional structure is determined by the amino acid sequence or primary structure (Anfinsen's dogma).[2]

Results of protein folding

The correct three-dimensional structure is essential to function, although some parts of functional proteins may remain unfolded,[3] so that protein dynamics is important. Failure to fold into native structure generally produces inactive proteins, but in some instances misfolded proteins have modified or toxic functionality. Several neurodegenerative and other diseases are believed to result from the accumulation of amyloid fibrils formed by misfolded proteins, infectious varieties of which are known as prions.[4] Many allergies are caused by incorrect folding of some proteins, because the immune system does not produce antibodies for certain protein structures.[5]

Denaturation of proteins is a process of transition from the folded to the unfolded state. It happens in cooking, in burns, in proteinopathies, and in other contexts.

The duration of the folding process varies dramatically depending on the protein of interest. When studied outside the cell, the slowest folding proteins require many minutes or hours to fold primarily due to proline isomerization, and must pass through a number of intermediate states, like checkpoints, before the process is complete.[6] On the other hand, very small single-domain proteins with lengths of up to a hundred amino

acids typically fold in a single step.[7] Time scales of milliseconds are the norm and the very fastest known protein folding reactions are complete within a few microseconds.[8] The folding time scale of a protein depends on its size, contact order and circuit topology.[9]

Understanding and simulating the protein folding process has been an important challenge for computational biology since the late 1960s.

## Contents

Process of protein folding

Primary structure

The primary structure of a protein, its linear amino-acid sequence, determines its native conformation.[10] The specific amino acid residues and their position in the polypeptide chain are the determining factors for which portions of the protein fold closely together and form its three-dimensional conformation. The amino acid composition is not as important as the sequence.[11] The essential fact of folding, however, remains that the amino acid sequence of each protein contains the information that specifies both the native structure and the pathway to attain that state. This is not to say that nearly identical amino acid sequences always fold similarly.[12] Conformations differ based on environmental factors as well; similar proteins fold differently based on where they are found.

Secondary structure
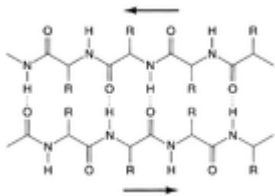
An anti-parallel beta pleated sheet displaying hydrogen bonding within the backbone

Formation of a secondary structure is the first step in the folding process that a protein takes to assume its native structure. Characteristic of secondary structure are the structures known as alpha helices and beta sheets that fold rapidly because they are stabilized by intramolecular hydrogen bonds, as was first characterized by Linus Pauling. Formation of intramolecular hydrogen bonds provides another important contribution to protein stability.[13] α-helices are formed by hydrogen bonding of the backbone to form a spiral shape (refer to figure on the right).[11] The β pleated sheet is a structure that forms with the backbone bending over itself to form the hydrogen bonds (as displayed in the figure to the left). The hydrogen bonds are between the amide hydrogen and carbonyl oxygen of the peptide bond. There exists anti-parallel β pleated sheets and parallel β pleated sheets where the stability of the hydrogen bonds is stronger in the anti-parallel β sheet as it hydrogen bonds with the ideal 180 degree angle compared to the slanted hydrogen bonds formed by parallel sheets.[11]
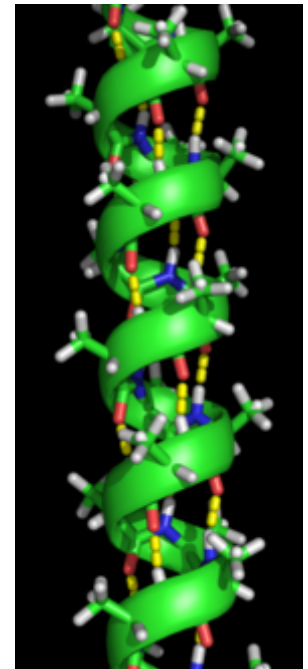
The alpha helix spiral formation

Tertiary structure

The α-Helices and β-Sheets are commonly amphipathic, meaning they have a hydrophilic and a hydrophobic portion. This ability helps in forming tertiary structure of a protein in which folding occurs so that the hydrophilic sides are facing the aqueous environment surrounding the protein and the hydrophobic sides are facing the hydrophobic core of the protein.[14] Secondary structure hierarchically gives way to tertiary structure formation. Once the protein's tertiary structure is formed and stabilized by the hydrophobic interactions, there may also be covalent bonding in the form of disulfide bridges formed between two cysteine residues. These non-covalent and covalent contacts take a specific topological arrangement in a native structure of a protein. Tertiary structure of a protein involves a single polypeptide chain; however, additional interactions of folded polypeptide chains give rise to quaternary structure formation.[15]

Quaternary structure

Tertiary structure may give way to the formation of quaternary structure in some proteins, which usually involves the "assembly" or "coassembly" of subunits that have already folded; in other words, multiple polypeptide chains could interact to form a fully functional quaternary protein.[11]

Driving forces of protein folding

Folding is a spontaneous process that is mainly guided by hydrophobic interactions, formation of intramolecular hydrogen bonds, van der Waals forces, and it is opposed by conformational entropy.[16] The process of folding often begins co-translationally, so that the N-terminus of the protein begins to fold while the C-terminal portion of the protein is still being synthesized by the ribosome; however, a protein molecule may fold spontaneously during or after biosynthesis.[17] While these macromolecules may be regarded as "folding themselves", the process also depends on the solvent (water or lipid bilayer),[18] the concentration of salts, the pH, the temperature, the possible presence of cofactors and of molecular chaperones.
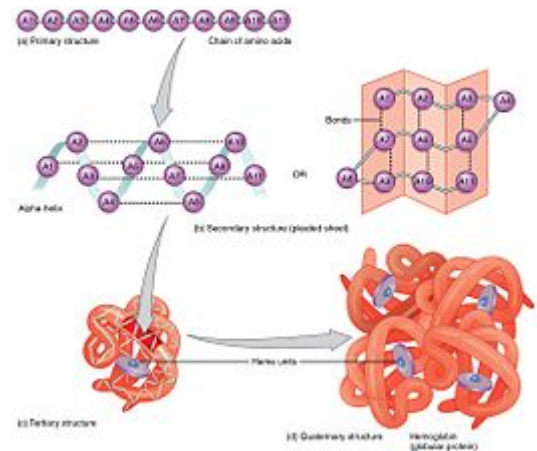
All forms of protein structure summarized

Proteins will have limitations on their folding abilities by the restricted bending angles or conformations that are possible. These allowable angles of protein folding are described with a two-dimensional plot known as the Ramachandran plot, depicted with psi and phi angles of allowable rotation.[19]

Hydrophobic effect
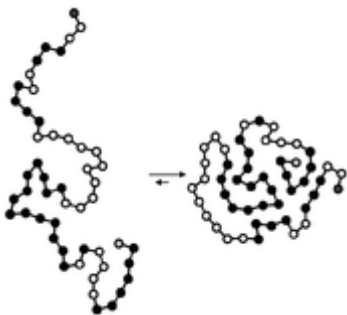
Hydrophobic collapse. In the compact fold (to the right), the hydrophobic amino acids (shown as black spheres) collapse toward the center to become shielded from aqueous environment.

Protein folding must be thermodynamically favorable within a cell in order for it to be a spontaneous reaction. Since it is known that protein folding is a spontaneous reaction, then it must assume a negative Gibbs free energy value. Gibbs free energy in protein folding is directly related to enthalpy and entropy.[11] For a negative delta G to arise and for protein folding to become thermodynamically favorable, then either enthalpy, entropy, or both terms must be favorable.

Minimizing the number of hydrophobic side-chains exposed to water is an important driving force behind the folding process.[20] The hydrophobic effect is the phenomenon in which the hydrophobic chains of a protein collapse into the core of the protein (away from the hydrophilic environment).[11] In an aqueous environment, the water molecules tend to aggregate around the hydrophobic regions or side chains of the protein, creating water shells of ordered water molecules.[21] An ordering of water molecules around a hydrophobic region increases order in a system and therefore contributes a negative change in entropy (less entropy in the system). The water molecules are fixed in these water cages which drives the hydrophobic collapse, or the inward folding of the hydrophobic groups. The hydrophobic collapse introduces entropy back to the system via the breaking of the water cages which frees the ordered water molecules.[11] The multitude of hydrophobic groups interacting within the core of the globular folded protein contributes a significant amount to protein stability after folding, because of the vastly accumulated van der Waals forces (specifically London Dispersion forces).[11] The hydrophobic effect exists as a driving force in thermodynamics only if there is the presence of an aqueous medium with an amphiphilic

molecule containing a large hydrophobic region.[22] The strength of hydrogen bonds depends on their environment; thus, H-bonds enveloped in a hydrophobic core contribute more than H-bonds exposed to the aqueous environment to the stability of the native state.[23]

In proteins with globular folds, hydrophobic amino acids tend to be interspersed along the primary sequence, rather than randomly distributed or clustered together.[24][25] However, proteins that have recently been born de novo, which tend to be intrinsically disordered,[26][27] show the opposite pattern of hydrophobic amino acid clustering along the primary sequence.[28]

Play media
Entropy is decreased as the water molecules become more orderly near the hydrophobic solute.

Chaperones

Molecular chaperones are a class of proteins that aid in the correct folding of other proteins in vivo. Chaperones exist in all cellular compartments and interact with the polypeptide chain in order to allow the native three-dimensional conformation of the protein to form; however, chaperones themselves are not included in the final structure of the protein they are assisting in.[29] Chaperones may assist in folding even when the nascent polypeptide is being synthesized by the ribosome.[30] Molecular chaperones operate by binding to stabilize an otherwise unstable structure of a protein in its folding pathway, but chaperones do not contain the necessary information to know the correct native structure of the protein they are aiding; rather, chaperones work by preventing incorrect folding conformations.[30] In this way, chaperones do not actually increase the rate of individual steps involved in the folding pathway toward the native structure; instead, they work by reducing possible unwanted aggregations of

Example of a small eukaryotic heat shock protein

the polypeptide chain that might otherwise slow down the search for the proper intermediate and they provide a more efficient pathway for the polypeptide chain to assume the correct conformations.[29] Chaperones are not to be confused with folding catalyst proteins, which catalyze chemical reactions responsible for slow steps in folding pathways. Examples of folding catalysts are protein disulfide isomerases and peptidyl-prolyl isomerases that may be involved in formation of disulfide bonds or interconversion between cis and trans stereoisomers of peptide group.[30] Chaperones are shown to be critical in the process of protein folding in vivo because they provide the protein with the aid needed to assume its proper alignments and conformations efficiently enough to become "biologically relevant".[31] This means that the polypeptide chain could theoretically fold into its native structure without the aid of chaperones, as demonstrated by protein folding experiments conducted in vitro;[31] however, this process proves to be too inefficient or too slow to exist in biological systems; therefore, chaperones are necessary for protein folding in vivo. Along with its role in aiding native structure formation, chaperones are shown to be involved in various roles such as protein transport, degradation, and even allow denatured proteins exposed to certain external denaturant factors an opportunity to refold into their correct native structures.[32]

A fully denatured protein lacks both tertiary and secondary structure, and exists as a so-called random coil. Under certain conditions some proteins can refold; however, in many cases, denaturation is irreversible.[33] Cells sometimes protect their proteins against the denaturing influence of heat with enzymes known as heat shock proteins (a type of chaperone), which assist other proteins both in folding and in remaining folded.
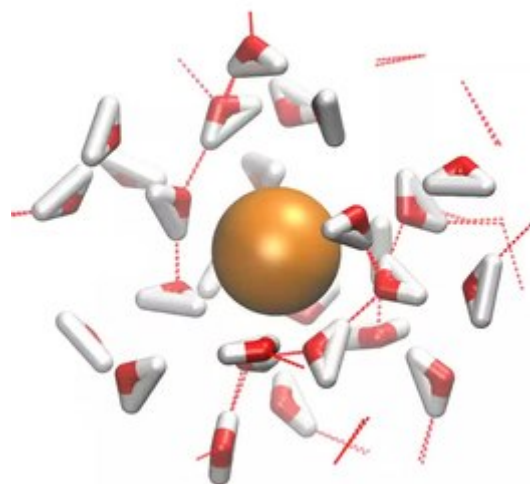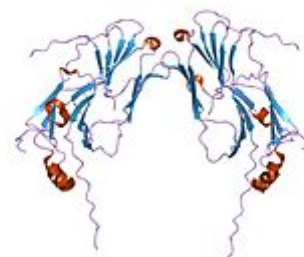
Heat shock proteins have been found in all species examined, from bacteria to humans, suggesting that they evolved very early and have an important function. Some proteins never fold in cells at all except with the assistance of chaperones which either isolate individual proteins so that their folding is not interrupted by interactions with other proteins or help to unfold misfolded proteins, allowing them to refold into the correct native structure.[34] This function is crucial to prevent the risk of precipitation into insoluble amorphous aggregates. The external factors involved in protein denaturation or disruption of the native state include temperature, external fields (electric, magnetic),[35] molecular crowding,[36] and even the limitation of space (i.e. confinement), which can have a big influence on the folding of proteins.[37] High concentrations of solutes, extremes of pH, mechanical forces, and the presence of chemical denaturants can contribute to protein denaturation, as well. These individual factors are categorized together as stresses. Chaperones are shown to exist in increasing concentrations during times of cellular stress and help the proper folding of emerging proteins as well as denatured or misfolded ones.[29]

Under some conditions proteins will not fold into their biochemically functional forms. Temperatures above or below the range that cells tend to live in will cause thermally unstable proteins to unfold or denature (this is why boiling makes an egg white turn opaque). Protein thermal stability is far from constant, however; for example, hyperthermophilic bacteria have been found that grow at temperatures as high as 122 °C,[38] which of course requires that their full complement of vital proteins and protein assemblies be stable at that temperature or above.

The bacterium *E. coli* is the host for bacteriophage T4, and the phage encoded gp31 protein (P17313 (http s://www.uniprot.org/uniprot/P17313)) appears to be structurally and functionally homologous to *E. coli* chaperone protein GroES and able to substitute for it in the assembly of bacteriophage T4 virus particles during infection.[39] Like GroES, gp31 forms a stable complex with GroEL chaperonin that is absolutely necessary for the folding and assembly in vivo of the bacteriophage T4 major capsid protein gp23.[39]

Fold switching

Some proteins have multiple native structures, and change their fold based on some external factors. For example, the KaiB protein switches fold throughout the day, acting as a clock for cyanobacteria. It has been estimated that around 0.5–4% of PDB (Protein Data Bank) proteins switch folds.[40]

Protein misfolding and neurodegenerative disease

A protein is considered to be misfolded if it cannot achieve its normal native state. This can be due to mutations in the amino acid sequence or a disruption of the normal folding process by external factors.[41] The misfolded protein typically contains β-sheets that are organized in a supramolecular arrangement known as a cross-β structure. These β-sheet-rich assemblies are very stable, very insoluble, and generally resistant to proteolysis.[42] The structural stability of these fibrillar assemblies is caused by extensive interactions between the protein monomers, formed by backbone hydrogen bonds between their β-strands.[42] The misfolding of proteins can trigger the further misfolding and accumulation of other proteins into aggregates or oligomers. The increased levels of aggregated proteins in the cell leads to formation of amyloid-like structures which can cause degenerative disorders and cell death.[41] The amyloids are fibrillary structures that contain intermolecular hydrogen bonds which are highly insoluble and made from converted protein aggregates.[41] Therefore, the proteasome pathway may not be efficient enough to degrade the misfolded proteins prior to aggregation. Misfolded proteins can interact with one another and form structured aggregates and gain toxicity through intermolecular interactions.[41]
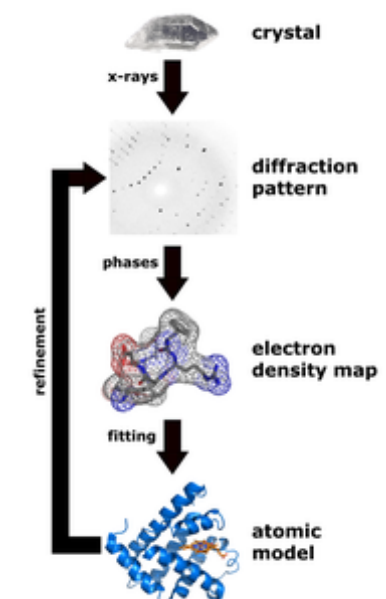
Aggregated proteins are associated with prion-related illnesses such as Creutzfeldt–Jakob disease, bovine spongiform encephalopathy (mad cow disease), amyloid-related illnesses such as Alzheimer's disease and familial amyloid cardiomyopathy or polyneuropathy,[43] as well as intracellular aggregation diseases such as Huntington's and Parkinson's disease.[4][44] These age onset degenerative diseases are associated with the aggregation of misfolded proteins into insoluble, extracellular aggregates and/or intracellular inclusions including cross-β amyloid fibrils. It is not completely clear whether the aggregates are the cause or merely a reflection of the loss of protein homeostasis, the balance between synthesis, folding, aggregation and protein turnover. Recently the European Medicines Agency approved the use of Tafamidis or Vyndaqel (a kinetic stabilizer of tetrameric transthyretin) for the treatment of transthyretin amyloid diseases. This suggests that the process of amyloid fibril formation (and not the fibrils themselves) causes the degeneration of post-mitotic tissue in human amyloid diseases.[45] Misfolding and excessive degradation instead of folding and function leads to a number of proteopathy diseases such as antitrypsin-associated emphysema, cystic fibrosis and the lysosomal storage diseases, where loss of function is the origin of the disorder. While protein replacement therapy has historically been used to correct the latter disorders, an emerging approach is to use pharmaceutical chaperones to fold mutated proteins to render them functional.

Experimental techniques for studying protein folding

While inferences about protein folding can be made through mutation studies, typically, experimental techniques for studying protein folding rely on the gradual unfolding or folding of proteins and observing conformational changes using standard non-crystallographic techniques.

X-ray crystallography

X-ray crystallography is one of the more efficient and important methods for attempting to decipher the three dimensional configuration of a folded protein.[46] To be able to conduct X-ray crystallography, the protein under investigation must be located inside a crystal lattice. To place a protein inside a crystal lattice, one must have a suitable solvent for crystallization, obtain a pure protein at supersaturated levels in solution, and precipitate the crystals in solution.[47] Once a protein is crystallized, X-ray beams can be concentrated through the crystal lattice which would diffract the beams or shoot them outwards in various directions. These exiting beams are correlated to the specific three-dimensional configuration of the protein enclosed within. The X-rays specifically interact with the electron clouds surrounding the individual atoms within the protein crystal lattice and produce a discernible diffraction pattern.[14] Only by relating the electron density clouds with the amplitude of the X-rays can this pattern be read and lead to assumptions of the phases or phase angles involved that complicate this method.[48] Without the relation established through a mathematical basis known as Fourier transform, the "phase problem" would render predicting the diffraction patterns very difficult.[14]

Steps of X-ray crystallography

Emerging methods like multiple isomorphous replacement use the presence of a heavy metal ion to diffract the X-rays into a more predictable manner, reducing the number of variables involved and resolving the phase problem.[46]

Fluorescence spectroscopy

Fluorescence spectroscopy is a highly sensitive method for studying the folding state of proteins. Three amino acids, phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp), have intrinsic fluorescence properties, but only Tyr and Trp are used experimentally because their quantum yields are high enough to give good fluorescence signals. Both Trp and Tyr are excited by a wavelength of 280 nm, whereas only Trp is excited by a wavelength of 295 nm. Because of their aromatic character, Trp and Tyr residues are often found fully or partially buried in the hydrophobic core of proteins, at the interface between two protein domains, or at the interface between subunits of oligomeric proteins. In this apolar environment, they have high quantum yields and therefore high fluorescence intensities. Upon disruption of the protein's tertiary or quaternary structure, these side chains become more exposed to the hydrophilic environment of the solvent, and their quantum yields decrease, leading to low fluorescence intensities. For Trp residues, the wavelength of their maximal fluorescence emission also depend on their environment.

Fluorescence spectroscopy can be used to characterize the equilibrium unfolding of proteins by measuring the variation in the intensity of fluorescence emission or in the wavelength of maximal emission as functions of a denaturant value.[49][50] The denaturant can be a chemical molecule (urea, guanidinium hydrochloride), temperature, pH, pressure, etc. The equilibrium between the different but discrete protein states, i.e. native state, intermediate states, unfolded state, depends on the denaturant value; therefore, the global fluorescence signal of their equilibrium mixture also depends on this value. One thus obtains a profile relating the global protein signal to the denaturant value. The profile of equilibrium unfolding may enable one to detect and identify intermediates of unfolding.[51][52] General equations have been developed by Hugues Bedouelle to obtain the thermodynamic parameters that characterize the unfolding equilibria for homomeric or heteromeric proteins, up to trimers and potentially tetramers, from such profiles.[49] Fluorescence spectroscopy can be combined with fast-mixing devices such as stopped flow, to measure protein folding kinetics,[53] generate a chevron plot and derive a Phi value analysis.

Circular dichroism

Circular dichroism is one of the most general and basic tools to study protein folding. Circular dichroism spectroscopy measures the absorption of circularly polarized light. In proteins, structures such as alpha helices and beta sheets are chiral, and thus absorb such light. The absorption of this light acts as a marker of the degree of foldedness of the protein ensemble. This technique has been used to measure equilibrium unfolding of the protein by measuring the change in this absorption as a function of denaturant concentration or temperature. A denaturant melt measures the free energy of unfolding as well as the protein's m value, or denaturant dependence. A temperature melt measures the denaturation temperature (Tm) of the protein.[49] As for fluorescence spectroscopy, circular-dichroism spectroscopy can be combined with fast-mixing devices such as stopped flow to measure protein folding kinetics and to generate chevron plots.

Vibrational circular dichroism of proteins

The more recent developments of vibrational circular dichroism (VCD) techniques for proteins, currently involving Fourier transform (FT) instruments, provide powerful means for determining protein conformations in solution even for very large protein molecules. Such VCD studies of proteins can be combined with X-ray diffraction data for protein crystals, FT-IR data for protein solutions in heavy water (D$_2$O), or quantum computations.

Protein nuclear magnetic resonance spectroscopy

Protein nuclear magnetic resonance (NMR) is able to collect protein structural data by inducing a magnet field through samples of concentrated protein. In NMR, depending on the chemical environment, certain nuclei will absorb specific radio-frequencies.[54][55] Because protein structural changes operate on a time scale from ns to ms, NMR is especially equipped to study intermediate structures in timescales of ps to s.[56] Some of the main techniques for studying proteins structure and non-folding protein structural changes include COSY, TOCSY, HSQC, time relaxation (T1 & T2), and NOE.[54] NOE is especially useful because magnetization transfers can be observed between spatially proximal hydrogens are observed.[54] Different NMR experiments have varying degrees of timescale sensitivity that are appropriate for different protein structural changes. NOE can pick up bond vibrations or side chain rotations, however, NOE is too sensitive to pick up protein folding because it occurs at larger timescale.[56]

Because protein folding takes place in about 50 to 3000 s^{-1} CPMG Relaxation dispersion and chemical exchange saturation transfer have become some of the primary techniques for NMR analysis of folding.[55] In addition, both techniques are used to uncover excited intermediate states in the protein folding landscape.[57] To do this, CPMG Relaxation dispersion takes advantage of the spin echo phenomenon. This technique exposes the target nuclei to a 90 pulse followed by one or more 180 pulses.[58] As the nuclei refocus, a broad distribution indicates the target nuclei is

Timescale of protein structural changes matched with NMR experiments. For protein folding, CPMG Relaxation Dispersion (CPMG RD) and chemical exchange saturation transfer (CEST) collect data in the appropriate timescale.

involved in an intermediate excited state. By looking at Relaxation dispersion plots the data collect information on the thermodynamics and kinetics between the excited and ground.[58][57] Saturation Transfer measures changes in signal from the ground state as excited states become perturbed. It uses weak radio frequency irradiation to saturate the excited state of a particular nuclei which transfers its saturation to the ground state.[55] This signal is amplified by decreasing the magnetization (and the signal) of the ground state.[55][57]

The main limitations in NMR is that its resolution decreases with proteins that are larger than 25 kDa and is not as detailed as X-ray crystallography.[55] Additionally, protein NMR analysis is quite difficult and can propose multiple solutions from the same NMR spectrum.[54]

In a study focused on the folding of an amyotrophic lateral sclerosis involved protein SOD1, excited intermediates were studied with relaxation dispersion and Saturation transfer.[59] SOD1 had been previously tied to many disease causing mutants which were assumed to be involved in protein aggregation, however the mechanism was still unknown. By using Relaxation Dispersion and Saturation Transfer experiments many excited intermediate states were uncovered misfolding in the SOD1 mutants.[59]

Dual-polarization interferometry

Dual polarisation interferometry is a surface-based technique for measuring the optical properties of molecular layers. When used to characterize protein folding, it measures the conformation by determining the overall size of a monolayer of the protein and its density in real time at sub-Angstrom resolution,[60] although real-time measurement of the kinetics of protein folding are limited to processes that occur slower than ~10 Hz. Similar to circular dichroism, the stimulus for folding can be a denaturant or temperature.

Studies of folding with high time resolution

The study of protein folding has been greatly advanced in recent years by the development of fast, time-resolved techniques. Experimenters rapidly trigger the folding of a sample of unfolded protein and observe the resulting dynamics. Fast techniques in use include neutron scattering,[61] ultrafast mixing of solutions, photochemical methods, and laser temperature jump spectroscopy. Among the many scientists who have contributed to the development of these techniques are Jeremy Cook, Heinrich Roder, Harry Gray, Martin Gruebele, Brian Dyer, William Eaton, Sheena Radford, Chris Dobson, Alan Fersht, Bengt Nölting and Lars Konermann.

Proteolysis

Proteolysis is routinely used to probe the fraction unfolded under a wide range of solution conditions (e.g. fast parallel proteolysis (FASTpp).[62][63]

Single-molecule force spectroscopy

Single molecule techniques such as optical tweezers and AFM have been used to understand protein folding mechanisms of isolated proteins as well as proteins with chaperones.[64] Optical tweezers have been used to stretch single protein molecules from their C- and N-termini and unfold them to allow study of the subsequent refolding.[65] The technique allows one to measure folding rates at single-molecule level; for example, optical tweezers have been recently applied to study folding and unfolding of proteins involved in blood coagulation. von Willebrand factor (vWF) is a protein with an essential role in blood clot formation process. It discovered – using single molecule optical tweezers measurement – that calcium-bound vWF acts as a shear force sensor in the blood. Shear force leads to unfolding of the A2 domain of vWF, whose refolding rate is dramatically enhanced in the presence of calcium.[66] Recently, it was also shown that the simple src SH3 domain accesses multiple unfolding pathways under force.[67]

Biotin painting

Biotin painting enables condition-specific cellular snapshots of (un)folded proteins. Biotin 'painting' shows a bias towards predicted Intrinsically disordered proteins.[68]

Computational studies of protein folding

Computational studies of protein folding includes three main aspects related to the prediction of protein stability, kinetics, and structure. A 2013 review summarizes the available computational methods for protein folding. [69]

Levinthal's paradox

In 1969, Cyrus Levinthal noted that, because of the very large number of degrees of freedom in an unfolded polypeptide chain, the molecule has an astronomical number of possible conformations. An estimate of 3^{300} or 10^{143} was made in one of his papers.[70] Levinthal's paradox is a thought experiment based on the observation that if a protein were folded by sequential sampling of all possible conformations, it would take an astronomical amount of time to do so, even if the conformations were sampled at a rapid
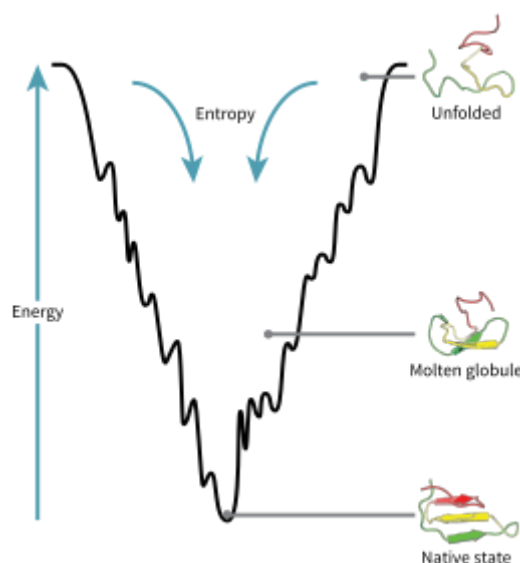
rate (on the nanosecond or picosecond scale).[71] Based upon the observation that proteins fold much faster than this, Levinthal then proposed that a random conformational search does not occur, and the protein must, therefore, fold through a series of meta-stable intermediate states.

Energy landscape of protein folding

The configuration space of a protein during folding can be visualized as an energy landscape. According to Joseph Bryngelson and Peter Wolynes, proteins follow the *principle of minimal frustration*, meaning that naturally evolved proteins have optimized their folding energy landscapes,[72] and that nature has chosen amino acid sequences so that the folded state of the protein is sufficiently stable. In addition, the acquisition of the folded state had to become a sufficiently fast process. Even though nature has reduced the level of *frustration* in proteins, some degree of it remains up to now as can be observed in the presence of local minima in the energy landscape of proteins.

A consequence of these evolutionarily selected sequences is that proteins are generally thought to have globally "funneled energy landscapes" (a term coined by José Onuchic)[73] that are largely directed toward the native state. This "folding funnel" landscape allows the protein to fold to the native state through any of a large number of pathways and intermediates, rather than being restricted to a single mechanism. The theory is supported by both computational simulations of model proteins and experimental studies,[72] and it has been used to improve methods for protein structure prediction and design.[72] The description of protein folding by the leveling free-energy landscape is also consistent with the 2nd law of thermodynamics.[74] Physically, thinking of landscapes in terms of visualizable potential or total energy surfaces simply with maxima, saddle points, minima, and funnels, rather like geographic landscapes, is perhaps a little misleading. The relevant description is really a high-dimensional phase space in which manifolds might take a variety of more complicated topological forms.[75]

The energy funnel by which an unfolded polypeptide chain assumes its native structure

The unfolded polypeptide chain begins at the top of the funnel where it may assume the largest number of unfolded variations and is in its highest energy state. Energy landscapes such as these indicate that there are a large number of initial possibilities, but only a single native state is possible; however, it does not reveal the numerous folding pathways that are possible. A different molecule of the same exact protein may be able to follow marginally different folding pathways, seeking different lower energy intermediates, as long as the same native structure is reached.[76] Different pathways may have different frequencies of utilization depending on the thermodynamic favorability of each pathway. This means that if one pathway is found to be more thermodynamically favorable than another, it is likely to be used more frequently in the pursuit of the native structure.[76] As the protein begins to fold and assume its various conformations, it always seeks a more thermodynamically favorable structure than before and thus continues through the energy funnel. Formation of secondary structures is a strong indication of increased stability within the protein, and only one combination of secondary structures assumed by the polypeptide backbone will have the lowest energy and therefore be present in the native state of the protein.[76] Among the first structures to form once the polypeptide begins to fold are alpha helices and beta turns, where alpha helices can form in as little as 100 nanoseconds and beta turns in 1 microsecond.[29]

There exists a saddle point in the energy funnel landscape where the transition state for a particular protein is found.[29] The transition state in the energy funnel diagram is the conformation that must be assumed by every molecule of that protein if the protein wishes to finally assume the native structure. No protein may assume the native structure without first passing through the transition state.[29] The transition state can be referred to as a variant or premature form of the native state rather than just another intermediary step.[77] The folding of the transition state is shown to be rate-determining, and even though it exists in a higher energy state than the native fold, it greatly resembles the native structure. Within the transition state, there exists a nucleus around which the protein is able to fold, formed by a process referred to as "nucleation condensation" where the structure begins to collapse onto the nucleus.[77]

Modeling of protein folding

De novo or *ab initio* techniques for computational protein structure prediction can be used for simulating various aspects of protein folding. Molecular Dynamics (MD) was used in simulations of protein folding and dynamics in silico.[78] First equilibrium folding simulations were done using implicit solvent model and umbrella sampling.[79] Because of computational cost, ab initio MD folding simulations with explicit water are limited to peptides and very small proteins.[80][81] MD simulations of larger proteins remain restricted to dynamics of the experimental structure or its high-temperature unfolding. Long-time folding processes (beyond about 1 millisecond), like folding of small-size proteins (about 50 residues) or larger, can be accessed using coarse-grained models.[82][83][84]
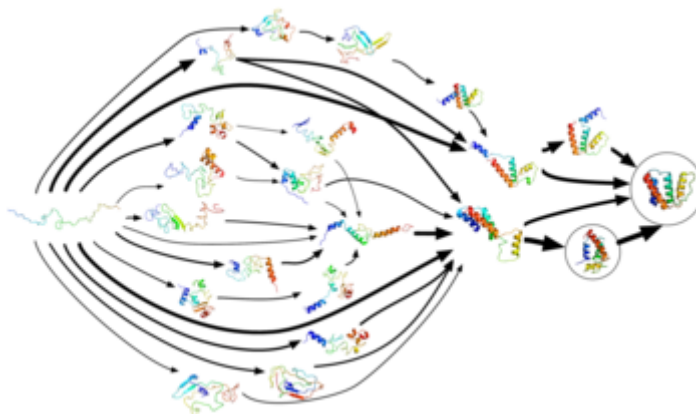
Folding@home uses Markov state models, like the one diagrammed here, to model the possible shapes and folding pathways a protein can take as it condenses from its initial randomly coiled state (left) into its native 3D structure (right).

Several large-scale computational projects, such as Rosetta@home,[85] Folding@home[86] and Foldit,[87] target protein folding.

Long continuous-trajectory simulations have been performed on Anton, a massively parallel supercomputer designed and built around custom ASICs and interconnects by D. E. Shaw Research. The longest published result of a simulation performed using Anton is a 2.936 millisecond simulation of NTL9 at 355 K.[88]

See also

- Chevron plot
- Denaturation midpoint
- Downhill folding
- Folding (chemistry)
- Phi value analysis
- Potential energy of protein
- Protein dynamics
- Protein misfolding cyclic amplification
- Protein structure prediction software
- Proteopathy
- Time-resolved mass spectrometry

References

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walters P (2002). "The Shape and Structure of Proteins" (https://www.ncbi.nlm.nih.gov/books/NBK26830/). *Molecular Biology of the Cell; Fourth Edition*. New York and London: Garland Science. ISBN 978-0-8153-3218-3.

2. Anfinsen CB (July 1972). "The formation and stabilization of protein structure" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1173893). *The Biochemical Journal*. **128** (4): 737–49. doi:10.1042/bj1280737 (https://doi.org/10.1042%2Fbj1280737). PMC 1173893 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1173893). PMID 4565129 (https://pubmed.ncbi.nlm.nih.gov/4565129).

3. Berg JM, Tymoczko JL, Stryer L (2002). "3. Protein Structure and Function" (https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=stryer%5Bbook%5D+AND+215168%5Buid%5D&rid=stryer.chapter.280). *Biochemistry*. San Francisco: W. H. Freeman. ISBN 978-0-7167-4684-3.

4. Selkoe DJ (December 2003). "Folding proteins in fatal ways". *Nature*. **426** (6968): 900–4. Bibcode:2003Natur.426..900S (https://ui.adsabs.harvard.edu/abs/2003Natur.426..900S). doi:10.1038/nature02264 (https://doi.org/10.1038%2Fnature02264). PMID 14685251 (https://pubmed.ncbi.nlm.nih.gov/14685251). S2CID 6451881 (https://api.semanticscholar.org/CorpusID:6451881).

5. Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2010). "Protein Structure and Function". *Essential cell biology* (Third ed.). New York, NY: Garland Science. pp. 120–70. ISBN 978-0-8153-4454-4.

6. Kim PS, Baldwin RL (1990). "Intermediates in the folding reactions of small proteins". *Annual Review of Biochemistry*. **59**: 631–60. doi:10.1146/annurev.bi.59.070190.003215 (https://doi.org/10.1146%2Fannurev.bi.59.070190.003215). PMID 2197986 (https://pubmed.ncbi.nlm.nih.gov/2197986).

7. Jackson SE (1998). "How do small single-domain proteins fold?" (https://doi.org/10.1016%2FS1359-0278%2898%2900033-9). *Folding & Design*. **3** (4): R81-91. doi:10.1016/S1359-0278(98)00033-9 (https://doi.org/10.1016%2FS1359-0278%2898%2900033-9). PMID 9710577 (https://pubmed.ncbi.nlm.nih.gov/9710577).

8. Kubelka J, Hofrichter J, Eaton WA (February 2004). "The protein folding 'speed limit' " (https://zenodo.org/record/1259347). *Current Opinion in Structural Biology*. **14** (1): 76–88. doi:10.1016/j.sbi.2004.01.013 (https://doi.org/10.1016%2Fj.sbi.2004.01.013). PMID 15102453 (https://pubmed.ncbi.nlm.nih.gov/15102453).

9. https://pubs.rsc.org/en/content/articlelanding/2021/cp/d1cp03390e

10. Anfinsen CB (July 1973). "Principles that govern the folding of protein chains". *Science*. **181** (4096): 223–30. Bibcode:1973Sci...181..223A (https://ui.adsabs.harvard.edu/abs/1973Sci...181..223A). doi:10.1126/science.181.4096.223 (https://doi.org/10.1126%2Fscience.181.4096.223). PMID 4124164 (https://pubmed.ncbi.nlm.nih.gov/4124164).

11. Voet D, Voet JG, Pratt CW (2016). *Principles of Biochemistry* (Fifth ed.). Wiley. ISBN 978-1-118-91840-1.

12. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (July 2007). "The design and characterization of two proteins with 88% sequence identity but different structure and function" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1906725). *Proceedings of the National Academy of Sciences of the United States of America*. **104** (29): 11963–8. Bibcode:2007PNAS..10411963A (https://ui.adsabs.harvard.edu/abs/2007PNAS..10411963A). doi:10.1073/pnas.0700922104 (https://doi.org/10.1073%2Fpnas.0700922104). PMC 1906725 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1906725). PMID 17609385 (https://pubmed.ncbi.nlm.nih.gov/17609385).

13. Rose GD, Fleming PJ, Banavar JR, Maritan A (November 2006). "A backbone-based theory of protein folding" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636505). *Proceedings of the National Academy of Sciences of the United States of America*. **103** (45): 16623–33. Bibcode:2006PNAS..10316623R (https://ui.adsabs.harvard.edu/abs/2006PNAS..10316623R). CiteSeerX 10.1.1.630.5487 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.630.5487). doi:10.1073/pnas.0606843103 (https://doi.org/10.1073%2Fpnas.0606843103). PMC 1636505 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636505). PMID 17075053 (https://pubmed.ncbi.nlm.nih.gov/17075053).

14. Fersht A (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (https://books.google.com/books?id=QdpZz_ahA5UC&pg=PR20). Macmillan. ISBN 978-0-7167-3268-6.

15. "Protein Structure" (http://www.nature.com/scitable/topicpage/protein-structure-14122136). *Scitable*. Nature Education. Retrieved 2016-11-26.

16. Pratt C, Cornely K (2004). "Thermodynamics" (http://www.wiley.com/college/pratt/0471393878/instructor/review/thermodynamics/7_relationship.html). *Essential Biochemistry* (https://archive.org/details/essentialbiochem00char). Wiley. ISBN 978-0-471-39387-0. Retrieved 2016-11-26.

17. Zhang G, Ignatova Z (February 2011). "Folding at the birth of the nascent chain: coordinating translation with co-translational folding". *Current Opinion in Structural Biology*. **21** (1): 25–31. doi:10.1016/j.sbi.2010.10.008 (https://doi.org/10.1016%2Fj.sbi.2010.10.008). PMID 21111607 (https://pubmed.ncbi.nlm.nih.gov/21111607).

18. van den Berg B, Wain R, Dobson CM, Ellis RJ (August 2000). "Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC306593). *The EMBO Journal*. **19** (15): 3870–5. doi:10.1093/emboj/19.15.3870 (https://doi.org/10.1093%2Femboj%2F19.15.3870). PMC 306593 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC306593). PMID 10921869 (https://pubmed.ncbi.nlm.nih.gov/10921869).

19. Al-Karadaghi S. "Torsion Angles and the Ramachnadran Plot in Protein Structures" (http://www.proteinstructures.com/Structure/Structure/Ramachandran-plot.html). *www.proteinstructures.com*. Retrieved 2016-11-26.

20. Pace CN, Shirley BA, McNutt M, Gajiwala K (January 1996). "Forces contributing to the conformational stability of proteins". *FASEB Journal*. **10** (1): 75–83. doi:10.1096/fasebj.10.1.8566551 (https://doi.org/10.1096%2Ffasebj.10.1.8566551). PMID 8566551 (https://pubmed.ncbi.nlm.nih.gov/8566551). S2CID 20021399 (https://api.semanticscholar.org/CorpusID:20021399).

21. Cui D, Ou S, Patel S (December 2014). "Protein-spanning water networks and implications for prediction of protein-protein interactions mediated through hydrophobic effects". *Proteins*. **82** (12): 3312–26. doi:10.1002/prot.24683 (https://doi.org/10.1002%2Fprot.24683). PMID 25204743 (https://pubmed.ncbi.nlm.nih.gov/25204743). S2CID 27113763 (https://api.semanticscholar.org/CorpusID:27113763).

22. Tanford C (June 1978). "The hydrophobic effect and the organization of living matter". *Science*. **200** (4345): 1012–8. Bibcode:1978Sci...200.1012T (https://ui.adsabs.harvard.edu/abs/1978Sci...200.1012T). doi:10.1126/science.653353 (https://doi.org/10.1126%2Fscience.653353). PMID 653353 (https://pubmed.ncbi.nlm.nih.gov/653353).

23. Deechongkit S, Nguyen H, Powers ET, Dawson PE, Gruebele M, Kelly JW (July 2004). "Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics". *Nature*. **430** (6995): 101–5. Bibcode:2004Natur.430..101D (https://ui.adsabs.harvard.edu/abs/2004Natur.430..101D). doi:10.1038/nature02611 (https://doi.org/10.1038%2Fnature02611). PMID 15229605 (https://pubmed.ncbi.nlm.nih.gov/15229605). S2CID 4315026 (https://api.semanticscholar.org/CorpusID:4315026).

24. Irbäck A, Sandelin E (November 2000). "On hydrophobicity correlations in protein chains" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1301114). *Biophysical Journal*. **79** (5): 2252–8. arXiv:cond-mat/0010390 (https://arxiv.org/abs/cond-mat/0010390). Bibcode:2000BpJ....79.2252I (https://ui.adsabs.harvard.edu/abs/2000BpJ....79.2252I). doi:10.1016/S0006-3495(00)76472-1 (https://doi.org/10.1016%2FS0006-3495%2800%2976472-1). PMC 1301114 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1301114). PMID 11053106 (https://pubmed.ncbi.nlm.nih.gov/11053106).

25. Irbäck A, Peterson C, Potthast F (September 1996). "Evidence for nonrandom hydrophobicity structures in protein chains" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC38463). *Proceedings of the National Academy of Sciences of the United States of America*. **93** (18): 9533–8. arXiv:chem-ph/9512004 (https://arxiv.org/abs/chem-ph/9512004). Bibcode:1996PNAS...93.9533I (https://ui.adsabs.harvard.edu/abs/1996PNAS...93.9533I). doi:10.1073/pnas.93.18.9533 (https://doi.org/10.1073%2Fpnas.93.18.9533). PMC 38463 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC38463). PMID 8790365 (https://pubmed.ncbi.nlm.nih.gov/8790365).

26. Wilson BA, Foy SG, Neme R, Masel J (June 2017). "De Novo Gene Birth" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5476217). *Nature Ecology & Evolution*. **1** (6): 0146–146. doi:10.1038/s41559-017-0146 (https://doi.org/10.1038%2Fs41559-017-0146). PMC 5476217 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5476217). PMID 28642936 (https://pubmed.ncbi.nlm.nih.gov/28642936).

27. Willis S, Masel J (September 2018). "Gene Birth Contributes to Structural Disorder Encoded by Overlapping Genes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6116962). *Genetics*. **210** (1): 303–313. doi:10.1534/genetics.118.301249 (https://doi.org/10.1534%2Fgenetics.118.301249). PMC 6116962 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6116962). PMID 30026186 (https://pubmed.ncbi.nlm.nih.gov/30026186).

28. Foy SG, Wilson BA, Bertram J, Cordes MH, Masel J (April 2019). "A Shift in Aggregation Avoidance Strategy Marks a Long-Term Direction to Protein Evolution" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6456324). *Genetics*. **211** (4): 1345–1355. doi:10.1534/genetics.118.301719 (https://doi.org/10.1534%2Fgenetics.118.301719). PMC 6456324 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6456324). PMID 30692195 (https://pubmed.ncbi.nlm.nih.gov/30692195).

29. Dobson CM (December 2003). "Protein folding and misfolding". *Nature*. **426** (6968): 884–90. Bibcode:2003Natur.426..884D (https://ui.adsabs.harvard.edu/abs/2003Natur.426..884D). doi:10.1038/nature02261 (https://doi.org/10.1038%2Fnature02261). PMID 14685248 (https://pubmed.ncbi.nlm.nih.gov/14685248). S2CID 1036192 (https://api.semanticscholar.org/CorpusID:1036192).

30. Hartl FU (June 1996). "Molecular chaperones in cellular protein folding". *Nature*. **381** (6583): 571–9. Bibcode:1996Natur.381..571H (https://ui.adsabs.harvard.edu/abs/1996Natur.381..571H). doi:10.1038/381571a0 (https://doi.org/10.1038%2F381571a0). PMID 8637592 (https://pubmed.ncbi.nlm.nih.gov/8637592). S2CID 4347271 (https://api.semanticscholar.org/CorpusID:4347271).

31. Hartl FU, Bracher A, Hayer-Hartl M (July 2011). "Molecular chaperones in protein folding and proteostasis". *Nature*. **475** (7356): 324–32. doi:10.1038/nature10317 (https://doi.org/10.1038%2Fnature10317). PMID 21776078 (https://pubmed.ncbi.nlm.nih.gov/21776078). S2CID 4337671 (https://api.semanticscholar.org/CorpusID:4337671).

32. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU (2013). "Molecular chaperone functions in protein folding and proteostasis". *Annual Review of Biochemistry*. **82**: 323–55. doi:10.1146/annurev-biochem-060208-092442 (https://doi.org/10.1146%2Fannurev-biochem-060208-092442). PMID 23746257 (https://pubmed.ncbi.nlm.nih.gov/23746257).

33. Shortle D (January 1996). "The denatured state (the other half of the folding equation) and its role in protein stability". *FASEB Journal*. **10** (1): 27–34. doi:10.1096/fasebj.10.1.8566543 (https://doi.org/10.1096%2Ffasebj.10.1.8566543). PMID 8566543 (https://pubmed.ncbi.nlm.nih.gov/8566543). S2CID 24066207 (https://api.semanticscholar.org/CorpusID:24066207).

34. Lee S, Tsai FT (2005). "Molecular chaperones in protein quality control" (https://doi.org/10.5483%2FBMBRep.2005.38.3.259). *Journal of Biochemistry and Molecular Biology*. **38** (3): 259–65. doi:10.5483/BMBRep.2005.38.3.259 (https://doi.org/10.5483%2FBMBRep.2005.38.3.259). PMID 15943899 (https://pubmed.ncbi.nlm.nih.gov/15943899).

35. Ojeda-May P, Garcia ME (July 2010). "Electric field-driven disruption of a native beta-sheet protein conformation and generation of a helix-structure" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2905109). *Biophysical Journal*. **99** (2): 595–9. Bibcode:2010BpJ....99..595O (https://ui.adsabs.harvard.edu/abs/2010BpJ....99..595O). doi:10.1016/j.bpj.2010.04.040 (https://doi.org/10.1016%2Fj.bpj.2010.04.040). PMC 2905109 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2905109). PMID 20643079 (https://pubmed.ncbi.nlm.nih.gov/20643079).

36. van den Berg B, Ellis RJ, Dobson CM (December 1999). "Effects of macromolecular crowding on protein folding and aggregation" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1171756). *The EMBO Journal*. **18** (24): 6927–33. doi:10.1093/emboj/18.24.6927 (https://doi.org/10.1093%2Femboj%2F18.24.6927). PMC 1171756 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1171756). PMID 10601015 (https://pubmed.ncbi.nlm.nih.gov/10601015).

37. Ellis RJ (July 2006). "Molecular chaperones: assisting assembly in addition to folding". *Trends in Biochemical Sciences*. **31** (7): 395–401. doi:10.1016/j.tibs.2006.05.001 (https://doi.org/10.1016%2Fj.tibs.2006.05.001). PMID 16716593 (https://pubmed.ncbi.nlm.nih.gov/16716593).

38. Takai K, Nakamura K, Toki T, Tsunogai U, Miyazaki M, Miyazaki J, Hirayama H, Nakagawa S, Nunoura T, Horikoshi K (August 2008). "Cell proliferation at 122 degrees C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2490668). *Proceedings of the National Academy of Sciences of the United States of America*. **105** (31): 10949–54. Bibcode:2008PNAS..10510949T (https://ui.adsabs.harvard.edu/abs/2008PNAS..10510949T). doi:10.1073/pnas.0712334105 (https://doi.org/10.1073%2Fpnas.0712334105). PMC 2490668 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2490668). PMID 18664583 (https://pubmed.ncbi.nlm.nih.gov/18664583).

39. Marusich EI, Kurochkina LP, Mesyanzhinov VV. Chaperones in bacteriophage T4 assembly. Biochemistry (Mosc). 1998;63(4):399-406

40. Porter, Lauren L.; Looger, Loren L. (5 June 2018). "Extant fold-switching proteins are widespread" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6003340). *Proceedings of the National Academy of Sciences*. **115** (23): 5968–5973. doi:10.1073/pnas.1800168115 (https://doi.org/10.1073%2Fpnas.1800168115). PMC 6003340 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6003340). PMID 29784778 (https://pubmed.ncbi.nlm.nih.gov/29784778).

41. Chaudhuri TK, Paul S (April 2006). "Protein-misfolding diseases and chaperone-based therapeutic approaches" (https://doi.org/10.1111%2Fj.1742-4658.2006.05181.x). *The FEBS Journal*. **273** (7): 1331–49. doi:10.1111/j.1742-4658.2006.05181.x (https://doi.org/10.1111%2Fj.1742-4658.2006.05181.x). PMID 16689923 (https://pubmed.ncbi.nlm.nih.gov/16689923). S2CID 23370420 (https://api.semanticscholar.org/CorpusID:23370420).

42. Soto C, Estrada L, Castilla J (March 2006). "Amyloids, prions and the inherent infectious nature of misfolded protein aggregates". *Trends in Biochemical Sciences*. **31** (3): 150–5. doi:10.1016/j.tibs.2006.01.002 (https://doi.org/10.1016%2Fj.tibs.2006.01.002). PMID 16473510 (https://pubmed.ncbi.nlm.nih.gov/16473510).

43. Hammarström P, Wiseman RL, Powers ET, Kelly JW (January 2003). "Prevention of transthyretin amyloid disease by changing protein misfolding energetics". *Science*. **299** (5607): 713–6. Bibcode:2003Sci...299..713H (https://ui.adsabs.harvard.edu/abs/2003Sci...299..713H). doi:10.1126/science.1079589 (https://doi.org/10.1126%2Fscience.1079589). PMID 12560553 (https://pubmed.ncbi.nlm.nih.gov/12560553). S2CID 30829998 (https://api.semanticscholar.org/CorpusID:30829998).

44. Chiti F, Dobson CM (2006). "Protein misfolding, functional amyloid, and human disease". *Annual Review of Biochemistry*. **75**: 333–66. doi:10.1146/annurev.biochem.75.101304.123901 (https://doi.org/10.1146%2Fannurev.biochem.75.101304.123901). PMID 16756495 (https://pubmed.ncbi.nlm.nih.gov/16756495).

45. Johnson SM, Wiseman RL, Sekijima Y, Green NS, Adamski-Werner SL, Kelly JW (December 2005). "Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses". *Accounts of Chemical Research*. **38** (12): 911–21. doi:10.1021/ar020073i (https://doi.org/10.1021%2Far020073i). PMID 16359163 (https://pubmed.ncbi.nlm.nih.gov/16359163).

46. Cowtan K (2001). "Phase Problem in X-ray Crystallography, and Its Solution" (http://people.bu.edu/mfk/restricted566/phaseproblem.pdf) (PDF). *Encyclopedia of Life Sciences*. Macmillan Publishers Ltd, Nature Publishing Group. Retrieved November 3, 2016.

47. Drenth J (2007-04-05). *Principles of Protein X-Ray Crystallography* (https://books.google.com/books?id=Jobr7svN0IIC&pg=PR5). Springer Science & Business Media. ISBN 978-0-387-33746-3.

48. Taylor G (2003). "The phase problem" (https://doi.org/10.1107%2FS0907444903017815). *Acta Crystallographica Section D*. **59** (11): 1881–90. doi:10.1107/S0907444903017815 (https://doi.org/10.1107%2FS0907444903017815). PMID 14573942 (https://pubmed.ncbi.nlm.nih.gov/14573942).

49. Bedouelle H (February 2016). "Principles and equations for measuring and interpreting protein stability: From monomer to tetramer". *Biochimie*. **121**: 29–37. doi:10.1016/j.biochi.2015.11.013 (https://doi.org/10.1016%2Fj.biochi.2015.11.013). PMID 26607240 (https://pubmed.ncbi.nlm.nih.gov/26607240).

50. Monsellier E, Bedouelle H (September 2005). "Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength" (https://doi.org/10.1093%2Fprotein%2Fgzi046). *Protein Engineering, Design & Selection*. **18** (9): 445–56. doi:10.1093/protein/gzi046 (https://doi.org/10.1093%2Fprotein%2Fgzi046). PMID 16087653 (https://pubmed.ncbi.nlm.nih.gov/16087653).

51. Park YC, Bedouelle H (July 1998). "Dimeric tyrosyl-tRNA synthetase from Bacillus stearothermophilus unfolds through a monomeric intermediate. A quantitative analysis under equilibrium conditions" (https://doi.org/10.1074%2Fjbc.273.29.18052). *The Journal of Biological Chemistry*. **273** (29): 18052–9. doi:10.1074/jbc.273.29.18052 (https://doi.org/10.1074%2Fjbc.273.29.18052). PMID 9660761 (https://pubmed.ncbi.nlm.nih.gov/9660761).

52. Ould-Abeih MB, Petit-Topin I, Zidane N, Baron B, Bedouelle H (June 2012). "Multiple folding states and disorder of ribosomal protein SA, a membrane receptor for laminin, anticarcinogens, and pathogens". *Biochemistry*. **51** (24): 4807–21. doi:10.1021/bi300335r (https://doi.org/10.1021%2Fbi300335r). PMID 22640394 (https://pubmed.ncbi.nlm.nih.gov/22640394).

53. Royer CA (May 2006). "Probing protein folding and conformational transitions with fluorescence". *Chemical Reviews*. **106** (5): 1769–84. doi:10.1021/cr0404390 (https://doi.org/10.1021%2Fcr0404390). PMID 16683754 (https://pubmed.ncbi.nlm.nih.gov/16683754).

54. Wüthrich K (December 1990). "Protein structure determination in solution by NMR spectroscopy" (http://www.jbc.org/content/265/36/22059). *The Journal of Biological Chemistry*. **265** (36): 22059–62. doi:10.1016/S0021-9258(18)45665-7 (https://doi.org/10.1016%2FS0021-9258%2818%2945665-7). PMID 2266107 (https://pubmed.ncbi.nlm.nih.gov/2266107).

55. Zhuravleva A, Korzhnev DM (May 2017). "Protein folding by NMR" (http://www.sciencedirect.com/science/article/pii/S0079656516300280). *Progress in Nuclear Magnetic Resonance Spectroscopy*. **100**: 52–77. doi:10.1016/j.pnmrs.2016.10.002 (https://doi.org/10.1016%2Fj.pnmrs.2016.10.002). PMID 28552172 (https://pubmed.ncbi.nlm.nih.gov/28552172).

56. Ortega G, Pons M, Millet O (2013-01-01). Karabencheva-Christova T (ed.). "Protein functional dynamics in multiple timescales as studied by NMR spectroscopy". *Advances in Protein Chemistry and Structural Biology*. Dynamics of Proteins and Nucleic Acids. Academic Press. **92**: 219–51. doi:10.1016/b978-0-12-411636-8.00006-7 (https://doi.org/10.1016%2Fb978-0-12-411636-8.00006-7). ISBN 9780124116368. PMID 23954103 (https://pubmed.ncbi.nlm.nih.gov/23954103).

57. Vallurupalli P, Bouvignies G, Kay LE (May 2012). "Studying "invisible" excited protein states in slow exchange with a major state conformation". *Journal of the American Chemical Society*. **134** (19): 8148–61. doi:10.1021/ja3001419 (https://doi.org/10.1021%2Fja3001419). PMID 22554188 (https://pubmed.ncbi.nlm.nih.gov/22554188).

58. Neudecker P, Lundström P, Kay LE (March 2009). "Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2717354). *Biophysical Journal*. **96** (6): 2045–54. Bibcode:2009BpJ....96.2045N (https://ui.adsabs.harvard.edu/abs/2009BpJ....96.2045N). doi:10.1016/j.bpj.2008.12.3907 (https://doi.org/10.1016%2Fj.bpj.2008.12.3907). PMC 2717354 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2717354). PMID 19289032 (https://pubmed.ncbi.nlm.nih.gov/19289032).

59. Sekhar A, Rumfeldt JA, Broom HR, Doyle CM, Sobering RE, Meiering EM, Kay LE (November 2016). "Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5111666). *Proceedings of the National Academy of Sciences of the United States of America*. **113** (45): E6939–E6945. doi:10.1073/pnas.1611418113 (https://doi.org/10.1073%2Fpnas.1611418113). PMC 5111666 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5111666). PMID 27791136 (https://pubmed.ncbi.nlm.nih.gov/27791136).

60. Cross GH, Freeman NJ, Swann MJ (2008). "Dual Polarization Interferometry: A Real-Time Optical Technique for Measuring (Bio)molecular Orientation, Structure and Function at the Solid/Liquid Interface". *Handbook of Biosensors and Biochips*. doi:10.1002/9780470061565.hbb055 (https://doi.org/10.1002%2F9780470061565.hbb055). ISBN 978-0-470-01905-4.

61. Bu Z, Cook J, Callaway DJ (September 2001). "Dynamic regimes and correlated structural dynamics in native and denatured alpha-lactalbumin". *Journal of Molecular Biology*. **312** (4): 865–73. doi:10.1006/jmbi.2001.5006 (https://doi.org/10.1006%2Fjmbi.2001.5006). PMID 11575938 (https://pubmed.ncbi.nlm.nih.gov/11575938).

62. Minde DP, Maurice MM, Rüdiger SG (2012). "Determining biophysical protein stability in lysates by a fast proteolysis assay, FASTpp" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3463568). *PLOS ONE*. **7** (10): e46147. Bibcode:2012PLoSO...746147M (https://ui.adsabs.harvard.edu/abs/2012PLoSO...746147M). doi:10.1371/journal.pone.0046147 (https://doi.org/10.1371%2Fjournal.pone.0046147). PMC 3463568 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3463568). PMID 23056252 (https://pubmed.ncbi.nlm.nih.gov/23056252).

63. Park C, Marqusee S (March 2005). "Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding". *Nature Methods*. **2** (3): 207–12. doi:10.1038/nmeth740 (https://doi.org/10.1038%2Fnmeth740). PMID 15782190 (https://pubmed.ncbi.nlm.nih.gov/15782190). S2CID 21364478 (https://api.semanticscholar.org/CorpusID:21364478).

64. Mashaghi A, Kramer G, Lamb DC, Mayer MP, Tans SJ (January 2014). "Chaperone action at the single-molecule level". *Chemical Reviews*. **114** (1): 660–76. doi:10.1021/cr400326k (https://doi.org/10.1021%2Fcr400326k). PMID 24001118 (https://pubmed.ncbi.nlm.nih.gov/24001118).

65. Jagannathan B, Marqusee S (November 2013). "Protein folding and unfolding under force" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4065244). *Biopolymers*. **99** (11): 860–9. doi:10.1002/bip.22321 (https://doi.org/10.1002%2Fbip.22321). PMC 4065244 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4065244). PMID 23784721 (https://pubmed.ncbi.nlm.nih.gov/23784721).

66. Jakobi AJ, Mashaghi A, Tans SJ, Huizinga EG (July 2011). "Calcium modulates force sensing by the von Willebrand factor A2 domain" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144584). *Nature Communications*. **2**: 385. Bibcode:2011NatCo...2..385J (https://ui.adsabs.harvard.edu/abs/2011NatCo...2..385J). doi:10.1038/ncomms1385 (https://doi.org/10.1038%2Fncomms1385). PMC 3144584 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144584). PMID 21750539 (https://pubmed.ncbi.nlm.nih.gov/21750539).

67. Jagannathan B, Elms PJ, Bustamante C, Marqusee S (October 2012). "Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497811). *Proceedings of the National Academy of Sciences of the United States of America*. **109** (44): 17820–5. Bibcode:2012PNAS..10917820J (https://ui.adsabs.harvard.edu/abs/2012PNAS..10917820J). doi:10.1073/pnas.1201800109 (https://doi.org/10.1073%2Fpnas.1201800109). PMC 3497811 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497811). PMID 22949695 (https://pubmed.ncbi.nlm.nih.gov/22949695).

68. Minde DP, Ramakrishna M, Lilley KS (2018). "Biotinylation by proximity labelling favours unfolded proteins" (https://doi.org/10.1101%2F274761). *bioRxiv*. doi:10.1101/274761 (https://doi.org/10.1101%2F274761).

69. Compiani M, Capriotti E (December 2013). "Computational and theoretical methods for protein folding". *Biochemistry*. **52** (48): 8601–24. doi:10.1021/bi4001529 (https://doi.org/10.1021%2Fbi4001529). PMID 24187909 (https://pubmed.ncbi.nlm.nih.gov/24187909).

70. "Structural Biochemistry/Proteins/Protein Folding - Wikibooks, open books for an open world" (https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins/Protein_Folding#The_Levinthal_Paradox). *en.wikibooks.org*. Retrieved 2016-11-05.

71. Levinthal C (1968). "Are there pathways for protein folding?" (https://web.archive.org/web/20090902211239/http://www.biochem.wisc.edu/courses/biochem704/Reading/Levinthal1968.pdf) (PDF). *Journal de Chimie Physique et de Physico-Chimie Biologique*. **65**: 44–45. Bibcode:1968JCP....65...44L (https://ui.adsabs.harvard.edu/abs/1968JCP....65...44L). doi:10.1051/jcp/1968650044 (https://doi.org/10.1051%2Fjcp%2F1968650044). Archived from the original (http://www.biochem.wisc.edu/courses/biochem704/Reading/Levinthal1968.pdf) (PDF) on 2009-09-02.

72. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (March 1995). "Funnels, pathways, and the energy landscape of protein folding: a synthesis". *Proteins*. **21** (3): 167–95. arXiv:chem-ph/9411008 (https://arxiv.org/abs/chem-ph/9411008). doi:10.1002/prot.340210302 (https://doi.org/10.1002%2Fprot.340210302). PMID 7784423 (https://pubmed.ncbi.nlm.nih.gov/7784423). S2CID 13838095 (https://api.semanticscholar.org/CorpusID:13838095).

73. Leopold PE, Montal M, Onuchic JN (September 1992). "Protein folding funnels: a kinetic approach to the sequence-structure relationship" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC49992). *Proceedings of the National Academy of Sciences of the United States of America*. **89** (18): 8721–5. Bibcode:1992PNAS...89.8721L (https://ui.adsabs.harvard.edu/abs/1992PNAS...89.8721L). doi:10.1073/pnas.89.18.8721 (https://doi.org/10.1073%2Fpnas.89.18.8721). PMC 49992 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC49992). PMID 1528885 (https://pubmed.ncbi.nlm.nih.gov/1528885).

74. Sharma V, Kaila VR, Annila A (2009). "Protein folding as an evolutionary process". *Physica A: Statistical Mechanics and Its Applications*. **388** (6): 851–62. Bibcode:2009PhyA..388..851S (https://ui.adsabs.harvard.edu/abs/2009PhyA..388..851S). doi:10.1016/j.physa.2008.12.004 (https://doi.org/10.1016%2Fj.physa.2008.12.004).

75. Robson B, Vaithilingam A (2008). "Protein Folding Revisited". *Molecular Biology of Protein Folding, Part B*. Progress in Molecular Biology and Translational Science. Vol. 84. pp. 161–202. doi:10.1016/S0079-6603(08)00405-4 (https://doi.org/10.1016%2FS0079-6603%2808%2900405-4). ISBN 978-0-12-374595-8. PMID 19121702 (https://pubmed.ncbi.nlm.nih.gov/19121702).

76. Dill KA, MacCallum JL (November 2012). "The protein-folding problem, 50 years on". *Science*. **338** (6110): 1042–6. Bibcode:2012Sci...338.1042D (https://ui.adsabs.harvard.edu/abs/2012Sci...338.1042D). doi:10.1126/science.1219021 (https://doi.org/10.1126%2Fscience.1219021). PMID 23180855 (https://pubmed.ncbi.nlm.nih.gov/23180855). S2CID 5756068 (https://api.semanticscholar.org/CorpusID:5756068).

77. Fersht AR (February 2000). "Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC26468). *Proceedings of the National Academy of Sciences of the United States of America*. **97** (4): 1525–9. Bibcode:2000PNAS...97.1525F (https://ui.adsabs.harvard.edu/abs/2000PNAS...97.1525F). doi:10.1073/pnas.97.4.1525 (https://doi.org/10.1073%2Fpnas.97.4.1525). PMC 26468 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC26468). PMID 10677494 (https://pubmed.ncbi.nlm.nih.gov/10677494).

78. Rizzuti B, Daggett V (March 2013). "Using simulations to provide the framework for experimental protein folding studies" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4084838). *Archives of Biochemistry and Biophysics*. **531** (1–2): 128–35. doi:10.1016/j.abb.2012.12.015 (https://doi.org/10.1016%2Fj.abb.2012.12.015). PMC 4084838 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4084838). PMID 23266569 (https://pubmed.ncbi.nlm.nih.gov/23266569).

79. Schaefer M, Bartels C, Karplus M (December 1998). "Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model". *Journal of Molecular Biology*. **284** (3): 835–48. doi:10.1006/jmbi.1998.2172 (https://doi.org/10.1006%2Fjmbi.1998.2172). PMID 9826519 (https://pubmed.ncbi.nlm.nih.gov/9826519).

80. Jones D. "Fragment-based Protein Folding Simulations" (http://www.cs.ucl.ac.uk/staff/d.jones/t42morph.html). University College London.

81. "Protein folding" (http://www.biomolecular-modeling.com/Abalone/Protein-folding.html) (by Molecular Dynamics).

82. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A (July 2016). "Coarse-Grained Protein Models and Their Applications" (https://doi.org/10.1021%2Facs.chemrev.6b00163). *Chemical Reviews*. **116** (14): 7898–936. doi:10.1021/acs.chemrev.6b00163 (https://doi.org/10.1021%2Facs.chemrev.6b00163). PMID 27333362 (https://pubmed.ncbi.nlm.nih.gov/27333362).

83. Kmiecik S, Kolinski A (July 2007). "Characterization of protein-folding pathways by reduced-space modeling" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1941469). *Proceedings of the National Academy of Sciences of the United States of America*. **104** (30): 12330–5. Bibcode:2007PNAS..10412330K (https://ui.adsabs.harvard.edu/abs/2007PNAS..10412330 K). doi:10.1073/pnas.0702265104 (https://doi.org/10.1073%2Fpnas.0702265104). PMC 1941469 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1941469). PMID 17636132 (https://pubmed.ncbi.nlm.nih.gov/17636132).

84. Adhikari AN, Freed KF, Sosnick TR (October 2012). "De novo prediction of protein folding pathways and structure using the principle of sequential stabilization" (https://www.ncbi.nlm. nih.gov/pmc/articles/PMC3491489). *Proceedings of the National Academy of Sciences of the United States of America*. **109** (43): 17442–7. Bibcode:2012PNAS..10917442A (https://u i.adsabs.harvard.edu/abs/2012PNAS..10917442A). doi:10.1073/pnas.1209000109 (https://d oi.org/10.1073%2Fpnas.1209000109). PMC 3491489 (https://www.ncbi.nlm.nih.gov/pmc/arti cles/PMC3491489). PMID 23045636 (https://pubmed.ncbi.nlm.nih.gov/23045636).

85. Rosetta@Home (http://boinc.bakerlab.org/rosetta)

86. Folding@Home (https://foldingathome.org/about/the-foldinghome-consortium/)

87. FoldIt - Folding Protein Game (http://fold.it/portal/info/science)

88. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (October 2011). "How fast-folding proteins fold". *Science*. **334** (6055): 517–20. Bibcode:2011Sci...334..517L (https://ui.adsabs.harvard. edu/abs/2011Sci...334..517L). doi:10.1126/science.1208351 (https://doi.org/10.1126%2Fsci ence.1208351). PMID 22034434 (https://pubmed.ncbi.nlm.nih.gov/22034434). S2CID 27988268 (https://api.semanticscholar.org/CorpusID:27988268).

External links

- Human Proteome Folding Project (http://www.worldcommunitygrid.org/research/proteome/ov erview.do)