Protein folding

"Protein thermodynamics" redirects here. For the thermodynamics of reactions catalyzed by proteins, see Enzyme.

Protein folding is the process by which a protein structure assumes its functional shape or conformation. It is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil.¹ Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. This polypeptide lacks any developed three-dimensional structure (the left hand side of the neighboring figure). 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 (Anfinsen's dogma).²

The correct three-dimensional structure is essential to function, although some parts of functional proteins may remain unfolded.³ Failure to fold into native structure produces inactive proteins that are usually toxic. Several neurodegenerative and other diseases are believed to result from the accumulation of amyloid fibrils formed by misfolded proteins.⁴ Many allergies are caused by the folding of the proteins, for the immune system does not produce antibodies for certain protein structures.⁵
Known facts

Relationship between folding and amino acid sequence

The amino-acid sequence of a protein determines
The amino acid sequence of a protein determines its native conformation. A protein molecule folds spontaneously during or after biosynthesis. While these macromolecules may be regarded as "folding themselves", the process also depends on the solvent (water or lipid bilayer), the concentration of salts, the temperature, and the presence of molecular chaperones.

Folded proteins usually have a side chain. Packing stabilizes the folded state, and charged or side chains occupy the solvent-exposed surface where they interact with surrounding water. Minimizing the number of hydrophobic side-chains exposed to water is an important driving force behind the folding process. Formation of intramolecular hydrogen bonds provides another important contribution to protein stability. 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.

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. Specialized proteins called chaperones assist in the folding of other proteins. A well studied example is the bacterial GroEL system, which assists in the folding of globular proteins. In eukaryotic organisms chaperones are known as heat shock proteins. Although most globular proteins are able to assume their native state unassisted, chaperone-assisted folding is often necessary in the crowded intracellular environment to prevent aggregation; chaperones are also used to prevent misfolding and aggregation that may occur as a consequence of exposure to heat or other changes in the cellular environment.

There are two models of protein folding that are currently being confirmed: The first: The diffusion collision
model, in which a nucleus is formed, then the secondary structure is formed, and finally these secondary structures are collided together and pack tightly together. **The second:** The nucleation-condensation model, in which the secondary and tertiary structures of the protein are made at the same time. Recent studies have shown that some proteins show characteristics of both of these folding models.

For the most part, scientists have been able to study many identical molecules folding together *en masse.* At the coarsest level, it appears that in transitioning to the native state, a given amino acid sequence takes on roughly the same route and proceeds through roughly the same intermediates and transition states. Often folding involves first the establishment of regular secondary and supersecondary structures, in particular **alpha helices** and **beta sheets,** and afterward **tertiary structure.** Formation of **quaternary structure** usually involves the "assembly" or "coassembly" of subunits that have already folded. The regular **alpha helix** and **beta sheet** structures fold rapidly because they are stabilized by intramolecular **hydrogen bonds,** as was first characterized by **Linus Pauling.** Protein folding may involve **covalent bonding** in the form of **disulfide bridges** formed between two **cysteine** residues or the formation of metal clusters. Shortly before settling into their more **energetically favourable** native conformation, molecules may pass through an intermediate "molten globule" state.

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. Folding is a **spontaneous process** independent of energy inputs from **nucleoside triphosphates.** The passage of the folded state is mainly guided by hydrophobic interactions, formation of intramolecular **hydrogen bonds,** and **van der Waals forces,** and it is opposed by **conformational entropy.**

### Disruption of the native state

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). High concentrations of **solute**, extremes of **pH**, mechanical forces, and the presence of chemical denaturants can do the same. Protein thermal stability is far from constant, however. For example, **hyperthermophilic bacteria** have been found that grow at...
temperatures as high as 122 °C,[13] which of course requires that their full complement of vital proteins and protein assemblies be stable at that temperature or above.

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.[14] Cells sometimes protect their proteins against the denaturing influence of heat with enzymes known as chaperones or heat shock proteins, which assist other proteins both in folding and in remaining folded. Some proteins never fold in cells at all except with the assistance of chaperone molecules, which either isolate individual proteins so that their folding is not interrupted by interactions with other proteins or help to unfold misfolded proteins, giving them a second chance to refold properly. This function is crucial to prevent the risk of precipitation into insoluble amorphous aggregates.

**Incorrect protein folding and neurodegenerative disease**

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,[15] as well as intracytoplasmic aggregation diseases such as Huntington's and Parkinson's disease.[4][16] These age onset degenerative diseases are associated with the aggregation of misfolded proteins into insoluble, extracellular aggregates and/or intracellular inclusions including cross-beta sheet amyloid fibrils. While 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, the recent European Medicines Agency approval of Tafamidis or Vyndaqel (a kinetic stabilizer of tetrameric transthyretin) for the treatment of the transthyretin amyloid diseases suggests that it is the process of amyloid fibril formation and not the fibrils themselves that causes the degeneration of post-mitotic tissue in human amyloid diseases.[17] 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.
Effect of external factors on the folding of proteins

Several external factors such as temperature, external fields (electric, magnetic),[18] molecular crowding,[19] limitation of space could have a big influence on the folding of proteins.[20] Modification of the local minima by external factors can also induce modifications of the folding trajectory.

Protein folding is a very finely tuned process. Hydrogen bonding between different atoms provides the force required. Hydrophobic interactions between hydrophobic amino acids pack the hydrophobic residues

The Levinthal paradox and kinetics

Levinthal's paradox is a thought experiment, also constituting a self-reference in the theory of protein folding. 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.

The Levinthal paradox[21] observes that if a protein were folded by sequentially 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). 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.

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.[22] On the other hand, very small single-domain proteins with lengths of up to a hundred amino acids typically fold in a single step.[23] Time scales of milliseconds are the norm and the very fastest known protein folding reactions are complete within a few microseconds.[24]

Experimental techniques for studying protein folding

While inferences about protein folding can be made through mutation studies[citation needed]; typically, experimental techniques for studying protein folding rely on the gradual unfolding or folding of a solution of proteins and observing conformational changes using standard non-crystallographic techniques for observing
Protein nuclear magnetic resonance spectroscopy

Main article: Protein NMR

Protein folding is routinely studied using NMR spectroscopy, for example by monitoring hydrogen-deuterium exchange of partially folded intermediates. [citation needed]

Circular dichroism

Main article: 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. [citation needed] 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 melting temperature ($T_m$) of the protein. This type of spectroscopy can also be combined with fast-mixing devices, such as stopped flow, to measure protein folding kinetics and to generate chevron plots.

Dual polarisation interferometry

Main article: Dual polarisation interferometry

Dual polarisation interferometry is a surface based technique for measuring the optical properties of molecular layers. When used to characterise 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 [citation needed]. 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.
Vibrational circular dichroism of proteins

The more recent developments of vibrational circular dichroism (VCD) techniques for proteins, currently involving Fourier transform (FFT) instruments, provide powerful means for determining protein conformations in solution even for very large protein molecules. Such VCD studies of proteins are often combined with X-ray diffraction of protein crystals, FT-IR data for protein solutions in heavy water (D\textsubscript{2}O), or \textit{ab initio} quantum computations to provide unambiguous structural assignments that are unobtainable from CD.\footnote{citation needed}

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. These are experimental methods for rapidly triggering the folding of a sample of unfolded protein, and then observing the resulting dynamics. Fast techniques in widespread use include neutron scattering,\footnote{25} 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.

Computational methods for studying protein folding

\textit{Main article: Protein structure prediction}

Energy landscape of protein folding

The protein folding phenomenon was largely an experimental endeavor until the formulation of an energy landscape theory of proteins by Joseph Bryngelson and Peter Wolynes in the late 1980s and early 1990s. This approach introduced the principle of minimal frustration.\footnote{26} This principle says that nature has chosen amino acid sequences so that the folded state of the protein is very stable. In addition, the undesired interactions between amino acids along the folding pathway are reduced making the acquisition of the folded state a very 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" (coined by José Onuchic)[27] 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,[26] and it has been used to improve methods for protein structure prediction and design.[26] The description of protein folding by the leveling free-energy landscape is also consistent with the 2nd law of thermodynamics.[28] 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 highly dimensional phase space in which manifolds might take a variety of more complicated topological forms.[29]

**Modeling of protein folding**

*De novo* or *ab initio* techniques for computational protein structure prediction are related to, but strictly distinct from experimental studies of protein folding. *Molecular Dynamics* (MD) is an important tool for studying protein folding and dynamics in silico. First equilibrium folding simulations were done using implicit solvent model and *Umbrella Sampling*. Because of computational cost, ab initio MD folding simulations with explicit water are limited to peptides and very small proteins.[30] MD simulations of larger proteins remain restricted to dynamics of the experimental structure or its high-temperature unfolding. In order to simulate long-time folding processes (beyond about 1 microsecond), like folding of small-size proteins (about 50 residues) or larger, some approximations or simplifications in protein models need to be introduced. An approach using reduced protein representation (pseudo-atoms representing groups of atoms are defined) and statistical potential are useful in protein structure prediction and modeling of the folding pathways.[33]

There are distributed computing projects which use *idle CPU* or *GPU* time of personal computers to solve problems such as protein folding or prediction of protein structure, one prominent example being the *Folding@home* project. People can run these programs on their computer or *PlayStation 3* to support them.

**See also**

- Chevron plot
Denaturation (biochemistry)
Denaturation midpoint
Downhill folding
Folding (chemistry)
Folding@Home
Foldit computer game
Levinthal paradox
Protein design
Protein dynamics
Protein Misfolding Cyclic Amplification
Protein structure prediction
Protein structure prediction software
Proteopathy
Rosetta@home
Statistical potential
Software for molecular mechanics modeling

References


18. ^Ojeda, P., Garcia, M. (2010). "Electric Field-Driven Disruption of a Native β-Sheet Protein Conformation


31. ^ "Fragment-based Protein Folding Simulations".

32. ^ "Protein folding" (by Molecular Dynamics).

**External links**

- FoldIt - Folding Protein Game
- Folding@Home
- Rosetta@Home
- Human Proteome Folding Project
- BHAGEERATH-H: Protein tertiary structure prediction server

http://www.englandlab.com/protein-folding.html

**Rate this page**

*What's this?*

- Trustworthy
- Objective
- Complete
- Well-written

I am highly knowledgeable about this topic (optional)

Submit ratings