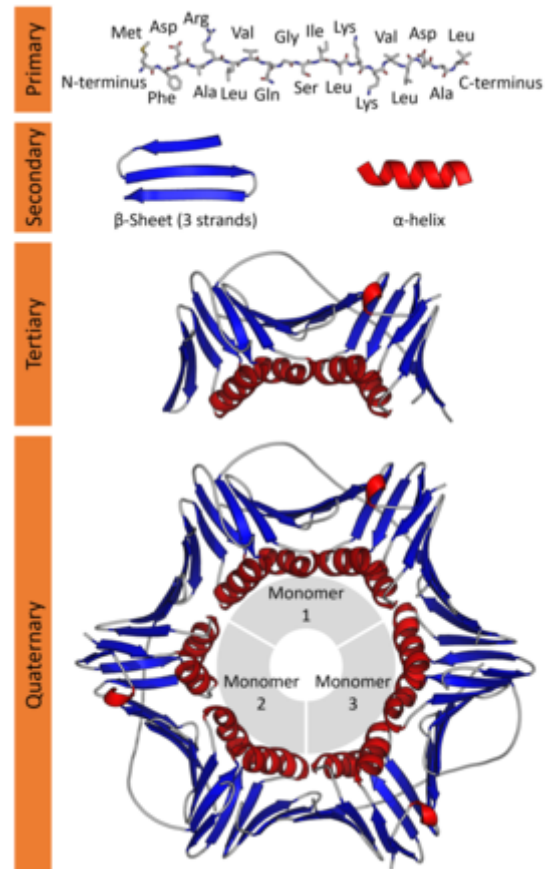# Protein structure

Protein structure is the three-dimensional arrangement of atoms in an amino acid-chain molecule. Proteins are polymers – specifically polypeptides – formed from sequences of amino acids, the monomers of the polymer. A single amino acid monomer may also be called a *residue* indicating a repeating unit of a polymer. Proteins form by amino acids undergoing condensation reactions, in which the amino acids lose one water molecule per reaction in order to attach to one another with a peptide bond. By convention, a chain under 30 amino acids is often identified as a peptide, rather than a protein.[1] To be able to perform their biological function, proteins fold into one or more specific spatial conformations driven by a number of non-covalent interactions such as hydrogen bonding, ionic interactions, Van der Waals forces, and hydrophobic packing. To understand the functions of proteins at a molecular level, it is often necessary to determine their three-dimensional structure. This is the topic of the scientific field of structural biology, which employs techniques such as X-ray crystallography, NMR spectroscopy, cryo electron microscopy (cryo-EM) and dual polarisation interferometry to determine the structure of proteins.

Protein structures range in size from tens to several thousand amino acids.[2] By physical size, proteins are classified as nanoparticles, between 1–100 nm. Very large protein complexes can be formed from protein subunits. For example, many thousands of actin molecules assemble into a microfilament.

Interactive diagram of protein structure, using PCNA as an example. (PDB: 1AXC (https://www.rcsb.org/structure/1AXC))

A protein usually undergoes reversible structural changes in performing its biological function. The alternative structures of the same protein are referred to as different conformations, and transitions between them are called conformational changes.

Contents

Levels of protein structure

**Domains, motifs, and folds in protein structure**

Levels of protein structure

There are four distinct levels of protein structure.

Primary structure

The primary structure of a protein refers to the sequence of amino acids in the polypeptide chain. The primary structure is held together by peptide bonds that are made during the process of protein biosynthesis. The two ends of the polypeptide chain are referred to as the carboxyl terminus (C-terminus) and the amino terminus (N-terminus) based on the nature of the free group on each extremity. Counting of residues always starts at the N-terminal end (NH_2-group), which is the end where the amino group is not involved in a peptide bond. The primary structure of a protein is determined by the gene corresponding to the protein. A specific sequence of

Four levels of protein structure

nucleotides in DNA is transcribed into mRNA, which is read by the ribosome in a process called translation. The sequence of amino acids in insulin was discovered by Frederick Sanger, establishing that proteins have defining amino acid sequences.[3][4] The sequence of a protein is unique to that protein, and defines the structure and function of the protein. The sequence of a protein can be determined by methods such as Edman degradation or tandem mass spectrometry. Often, however, it is read directly from the sequence of the gene using the genetic code. It is strictly recommended to use the words "amino acid residues" when discussing proteins because when a peptide bond is formed, a water molecule is lost, and therefore proteins are made up
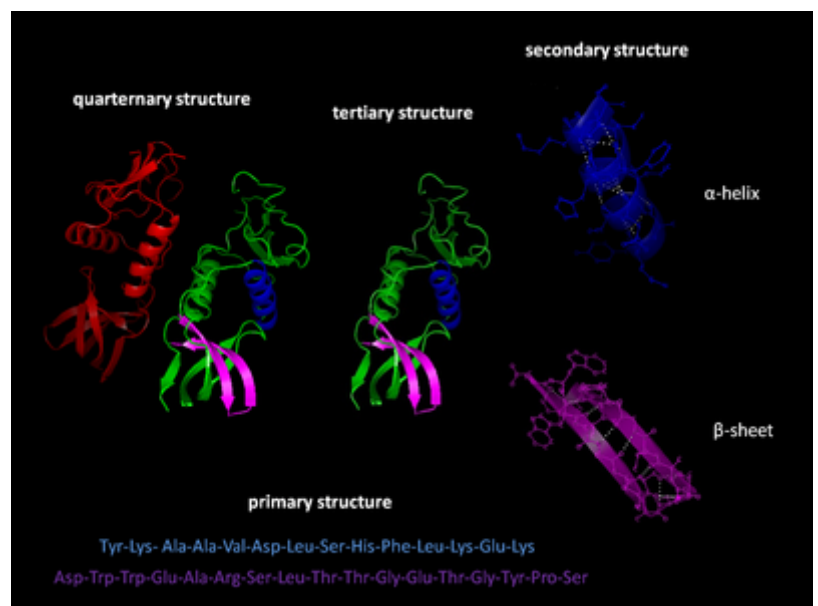
of amino acid residues. Post-translational modifications such as phosphorylations and glycosylations are usually also considered a part of the primary structure, and cannot be read from the gene. For example, insulin is composed of 51 amino acids in 2 chains. One chain has 31 amino acids, and the other has 20 amino acids.

Secondary structure

Secondary structure refers to highly regular local sub-structures on the actual polypeptide backbone chain. Two main types of secondary structure, the α-helix and the β-strand or β-sheets, were suggested in 1951 by Linus Pauling et al.[5] These secondary structures are defined by patterns of hydrogen bonds between the main-chain peptide groups. They have a regular geometry, being constrained to specific values of the dihedral angles ψ and φ on the Ramachandran plot. Both the α-helix and the β-sheet represent a way of saturating all the hydrogen bond donors and acceptors in the peptide backbone. Some parts of the protein are ordered but do not form any regular structures. They should not be confused with random coil, an unfolded polypeptide chain lacking any fixed three-dimensional structure. Several sequential secondary structures may form a "supersecondary unit".[6]
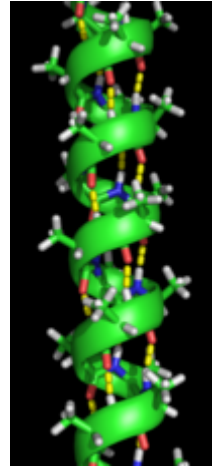
An α-helix with hydrogen bonds (yellow dots)

Tertiary structure

Tertiary structure refers to the three-dimensional structure created by a single protein molecule (a single polypeptide chain). It may include one or several domains. The α-helixes and β-pleated-sheets are folded into a compact globular structure. The folding is driven by the *non-specific* hydrophobic interactions, the burial of hydrophobic residues from water, but the structure is stable only when the parts of a protein domain are locked into place by *specific* tertiary interactions, such as salt bridges, hydrogen bonds, and the tight packing of side chains and disulfide bonds. The disulfide bonds are extremely rare in cytosolic proteins, since the cytosol (intracellular fluid) is generally a reducing environment.

Quaternary structure

Quaternary structure is the three-dimensional structure consisting of the aggregation of two or more individual polypeptide chains (subunits) that operate as a single functional unit (multimer). The resulting multimer is stabilized by the same non-covalent interactions and disulfide bonds as in tertiary structure. There are many possible quaternary structure organisations.[7] Complexes of two or more polypeptides (i.e. multiple subunits) are called multimers. Specifically it would be called a dimer if it contains two subunits, a trimer if it contains three subunits, a tetramer if it contains four subunits, and a pentamer if it contains five subunits. The subunits are frequently related to one another by symmetry operations, such as a 2-fold axis in a dimer. Multimers made up of identical subunits are referred to with a prefix of "homo-" and those made up of different subunits are referred to with a prefix of "hetero-", for example, a heterotetramer, such as the two alpha and two beta chains of hemoglobin.

Domains, motifs, and folds in protein structure

Proteins are frequently described as consisting of several structural units. These units include domains, motifs, and folds. Despite the fact that there are about 100,000 different proteins expressed in eukaryotic systems, there are many fewer different domains, structural motifs and folds.

Structural domain

A structural domain is an element of the protein's overall structure that is self-stabilizing and often folds independently of the rest of the protein chain. Many domains are not unique to the protein products of one gene or one gene family but instead appear in a variety of proteins. Domains often are named and singled out because they figure prominently in the biological function of the protein they belong to; for example, the "calcium-binding domain of calmodulin". Because they are independently stable, domains can be "swapped" by genetic engineering between one protein and another to make chimera proteins. A conservative combination of several domains that occur in different proteins, such as protein tyrosine phosphatase domain and C2 domain pair, was called "a superdomain" that may evolve as a single unit.[8]
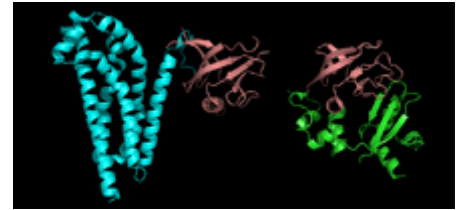
Protein domains. The two shown protein structures share a common domain (maroon), the PH domain, which is involved in phosphatidylinositol (3,4,5)-trisphosphate binding

Structural and sequence motifs

The structural and sequence motifs refer to short segments of protein three-dimensional structure or amino acid sequence that were found in a large number of different proteins

Supersecondary structure

The supersecondary structure refers to a specific combination of secondary structure elements, such as β-α-β units or a helix-turn-helix motif. Some of them may be also referred to as structural motifs.

Protein fold

A protein fold refers to the general protein architecture, like a helix bundle, β-barrel, Rossmann fold or different "folds" provided in the Structural Classification of Proteins database.[9] A related concept is protein topology.

Protein dynamics and conformational ensembles

Proteins are not static objects, but rather populate ensembles of conformational states. Transitions between these states typically occur on nanoscales, and have been linked to functionally relevant phenomena such as allosteric signaling[10] and enzyme catalysis.[11] Protein dynamics and conformational changes allow proteins to function as nanoscale biological machines within cells, often in the form of multi-protein complexes.[12] Examples include motor proteins, such as myosin, which is responsible for muscle contraction, kinesin, which moves cargo inside cells away from the nucleus along microtubules, and dynein, which moves cargo inside cells towards the nucleus and produces the axonemal beating of motile cilia and flagella. "[I]n effect, the [motile cilium] is a nanomachine composed of perhaps over 600 proteins in molecular complexes, many of which also function independently as nanomachines...Flexible linkers allow the mobile protein domains connected by them to recruit their binding partners and induce long-range allostery via protein domain dynamics. "[13]

Proteins are often thought of as relatively stable tertiary structures that experience conformational changes after being affected by interactions with other proteins or as a part of enzymatic activity. However, proteins may have varying degrees of stability, and some of the less stable variants are intrinsically disordered proteins. These proteins exist and function in a relatively 'disordered' state lacking a stable tertiary structure. As a result, they are difficult to describe by a single fixed tertiary structure. Conformational ensembles have been devised

as a way to provide a more accurate and 'dynamic' representation of the conformational state of intrinsically disordered proteins.[15][14]

Protein ensemble files are a representation of a protein that can be considered to have a flexible structure. Creating these files requires determining which of the various theoretically possible protein conformations actually exist. One

Schematic view of the two main ensemble modeling approaches.[14]

approach is to apply computational algorithms to the protein data in order to try to determine the most likely set of conformations for an ensemble file. There are multiple methods for preparing data for the Protein Ensemble Database (https://web.archive.org/web/20180310010556/http://pedb.vib.be/) that fall into two general methodologies – pool and molecular dynamics (MD) approaches (diagrammed in the figure). The pool based approach uses the protein's amino acid sequence to create a massive pool of random conformations. This pool is then subjected to more computational processing that creates a set of theoretical parameters for each conformation based on the structure. Conformational subsets from this pool whose average theoretical parameters closely match known experimental data for this protein are selected. The alternative molecular dynamics approach takes multiple random conformations at a time and subjects all of them to experimental data. Here the experimental data is serving as limitations to be placed on the conformations (e.g. known distances between atoms). Only conformations that manage to remain within the limits set by the experimental data are accepted. This approach often applies large amounts of experimental data to the conformations which is a very computationally demanding task.[14]
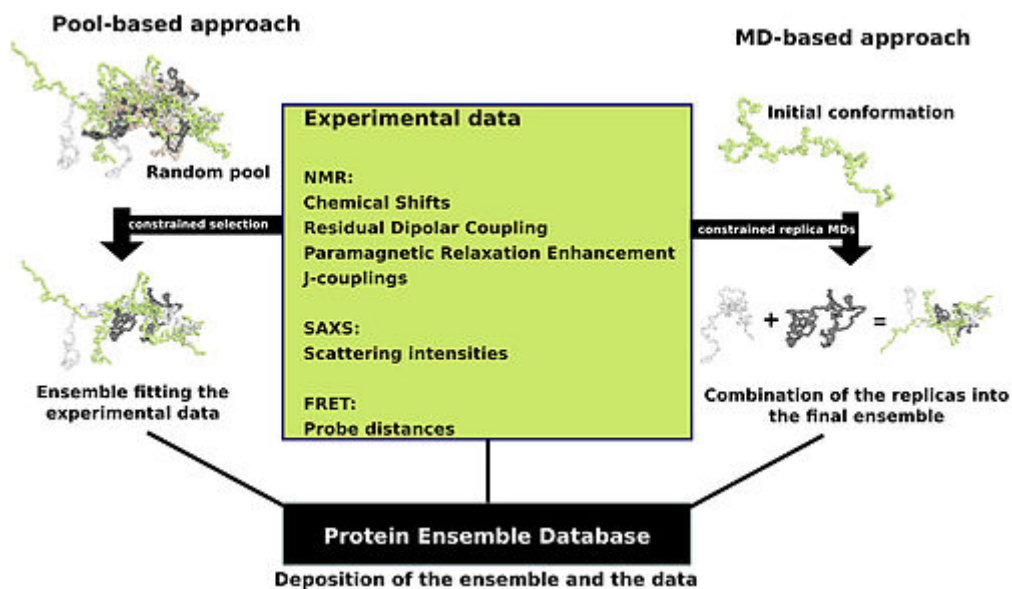
The conformational ensembles were generated for a number of highly dynamic and partially unfolded proteins, such as Sic1/Cdc4,[16] p15 PAF,[17] MKK7,[18] Beta-synuclein[19] and P27[20]

Protein folding

As it is translated, polypeptides exit the ribosome mostly as a random coil and folds into its native state.[21][22] The final structure of the protein chain is generally assumed to be determined by its amino acid sequence (Anfinsen's dogma).[23]

Protein stability

Thermodynamic stability of proteins represents the free energy difference between the folded and unfolded protein states. This free energy difference is very sensitive to temperature, hence a change in temperature may result in unfolding or denaturation. Protein denaturation may result in loss of function, and loss of native state. The free energy of stabilization of soluble globular proteins typically does not exceed 50 kJ/mol. Taking into consideration the large number of hydrogen bonds that take place for the stabilization of secondary structures, and the stabilization of the inner core through hydrophobic interactions, the free energy of stabilization emerges as small difference between large numbers.[24]
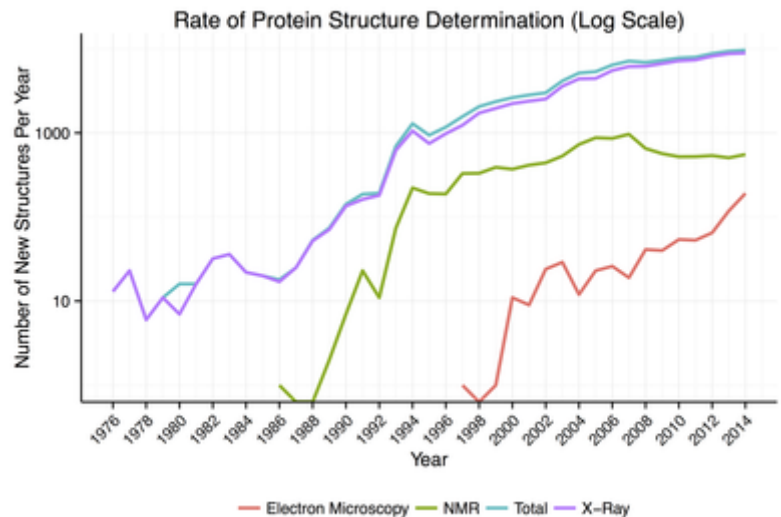
Protein structure determination
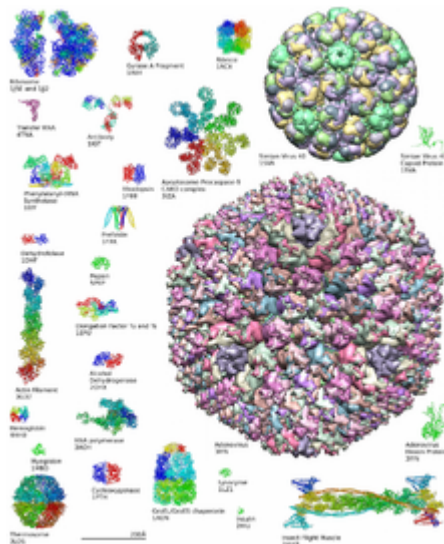
Examples of protein structures from the PDB

Rate of Protein Structure Determination by Method and Year

Around 90% of the protein structures available in the Protein Data Bank have been determined by X-ray crystallography.[25] This method allows one to measure the three-dimensional (3-D) density distribution of electrons in the protein, in the crystallized state, and thereby infer the 3-D coordinates of all the atoms to be determined to a certain resolution. Roughly 9% of the known protein structures have been obtained by nuclear magnetic resonance (NMR) techniques. For larger protein complexes, cryo-electron microscopy can determine protein structures. The resolution is typically lower than that of X-ray crystallography, or NMR, but the maximum resolution is steadily increasing. This technique is still a particularly valuable for very large protein complexes such as virus coat proteins and amyloid fibers.

General secondary structure composition can be determined via circular dichroism. Vibrational spectroscopy can also be used to characterize the conformation of peptides, polypeptides, and proteins.[26] Two-dimensional infrared spectroscopy has become a valuable method to investigate the structures of flexible peptides and proteins that cannot be studied with other methods.[27][28] A more qualitative picture of protein structure is often obtained by proteolysis, which is also useful to screen for more crystallizable protein samples. Novel implementations of this approach, including fast parallel proteolysis (FASTpp), can probe the structured fraction and its stability without the need for purification.[29] Once a protein's structure has been experimentally determined, further detailed studies can be done computationally, using molecular dynamic simulations of that structure.[30]

Protein structure databases

A protein structure database is a database that is modeled around the various experimentally determined protein structures. The aim of most protein structure databases is to organize and annotate the protein structures, providing the biological community access to the experimental data in a useful way. Data included in protein structure databases often includes 3D coordinates as well as experimental information, such as unit cell dimensions and angles for x-ray crystallography determined structures. Though most instances, in this case either proteins or a specific structure determinations of a protein, also contain sequence information and some databases even provide means for performing sequence based queries, the primary attribute of a structure database is structural information, whereas sequence databases focus on sequence information, and contain no structural information for the majority of entries. Protein structure databases are critical for many efforts in

computational biology such as structure based drug design, both in developing the computational methods used and in providing a large experimental dataset used by some methods to provide insights about the function of a protein.[31]

Structural classifications of proteins

Protein structures can be grouped based on their structural similarity, topological class or a common evolutionary origin. The Structural Classification of Proteins database[32] and CATH database[33] provide two different structural classifications of proteins. When the structural similarity is large the two proteins have possibly diverged from a common ancestor,[34] and shared structure between proteins is considered evidence of homology. Structure similarity can then be used to group proteins together into protein superfamilies.[35] If shared structure is significant but the fraction shared is small, the fragment shared may be the consequence of a more dramatic evolutionary event such as horizontal gene transfer, and joining proteins sharing these fragments into protein superfamilies is no longer justified.[34] Topology of a protein can be used to classify proteins as well. Knot theory and circuit topology are two topology frameworks developed for classification of protein folds based on chain crossing and intrachain contacts respectively.

Computational prediction of protein structure

The generation of a protein sequence is much easier than the determination of a protein structure. However, the structure of a protein gives much more insight in the function of the protein than its sequence. Therefore, a number of methods for the computational prediction of protein structure from its sequence have been developed.[36] *Ab initio* prediction methods use just the sequence of the protein. Threading and homology modeling methods can build a 3-D model for a protein of unknown structure from experimental structures of evolutionarily-related proteins, called a protein family.

See also

- Biomolecular structure
- Gene structure
- Nucleic acid structure
- Ribbon diagram 3D schematic representation of proteins

References

1. H. Stephen Stoker (1 January 2015). *Organic and Biological Chemistry* (https://books.google.com/books?id=HRCdBQAAQBAJ&pg=PA371). Cengage Learning. p. 371. ISBN 978-1-305-68645-8.
2. Brocchieri L, Karlin S (10 June 2005). "Protein length in eukaryotic and prokaryotic proteomes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1150220). *Nucleic Acids Research*. **33** (10): 3390–3400. doi:10.1093/nar/gki615 (https://doi.org/10.1093%2Fnar%2Fgki615). PMC 1150220 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1150220). PMID 15951512 (https://pubmed.ncbi.nlm.nih.gov/15951512).
3. Sanger, F.; Tuppy, H. (1 September 1951). "The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1197535). *The Biochemical Journal*. **49** (4): 463–481. doi:10.1042/bj0490463 (https://doi.org/10.1042%2Fbj0490463). ISSN 0264-6021 (https://www.worldcat.org/issn/0264-6021). PMC 1197535 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1197535). PMID 14886310 (https://pubmed.ncbi.nlm.nih.gov/14886310).

4. Sanger, F. (15 May 1959). "Chemistry of Insulin". *Science*. **129** (3359): 1340–1344. Bibcode:1959Sci...129.1340G (https://ui.adsabs.harvard.edu/abs/1959Sci...129.1340G). doi:10.1126/science.129.3359.1340 (https://doi.org/10.1126%2Fscience.129.3359.1340). ISSN 0036-8075 (https://www.worldcat.org/issn/0036-8075). PMID 13658959 (https://pubmed.ncbi.nlm.nih.gov/13658959).

5. Pauling L, Corey RB, Branson HR (1951). "The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063337). *Proc Natl Acad Sci USA*. **37** (4): 205–211. Bibcode:1951PNAS...37..205P (https://ui.adsabs.harvard.edu/abs/1951PNAS...37..205P). doi:10.1073/pnas.37.4.205 (https://doi.org/10.1073%2Fpnas.37.4.205). PMC 1063337 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063337). PMID 14816373 (https://pubmed.ncbi.nlm.nih.gov/14816373).

6. Chiang YS, Gelfand TI, Kister AE, Gelfand IM (2007). "New classification of supersecondary structures of sandwich-like proteins uncovers strict patterns of strand assemblage". *Proteins*. **68** (4): 915–921. doi:10.1002/prot.21473 (https://doi.org/10.1002%2Fprot.21473). PMID 17557333 (https://pubmed.ncbi.nlm.nih.gov/17557333).

7. Moutevelis E, Woolfson DN (January 2009). "A periodic table of coiled-coil protein structures". *J. Mol. Biol*. **385** (3): 726–32. doi:10.1016/j.jmb.2008.11.028 (https://doi.org/10.1016%2Fj.jmb.2008.11.028). ISSN 0022-2836 (https://www.worldcat.org/issn/0022-2836). PMID 19059267 (https://pubmed.ncbi.nlm.nih.gov/19059267).

8. Haynie DT, Xue B (2015). "Superdomain in the protein structure hierarchy: the case of PTP-C2" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4420535). *Protein Science*. **24** (5): 874–82. doi:10.1002/pro.2664 (https://doi.org/10.1002%2Fpro.2664). PMC 4420535 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4420535). PMID 25694109 (https://pubmed.ncbi.nlm.nih.gov/25694109).

9. Govindarajan S, Recabarren R, Goldstein RA (17 September 1999). "Estimating the total number of protein folds" (https://archive.today/20130105075413/http://www3.interscience.wiley.com/journal/65000323/abstract). *Proteins*. **35** (4): 408–414. doi:10.1002/(SICI)1097-0134(19990601)35:4<408::AID-PROT4>3.0.CO;2-A (https://doi.org/10.1002%2F%28SICI%291097-0134%2819990601%2935%3A4%3C408%3A%3AAID-PROT4%3E3.0.CO%3B2-A). hdl:2027.42/34969 (https://hdl.handle.net/2027.42%2F34969). PMID 10382668 (https://pubmed.ncbi.nlm.nih.gov/10382668). Archived from the original (http://www3.interscience.wiley.com/journal/65000323/abstract) on 5 January 2013.

10. Bu Z, Callaway DJ (2011). "Proteins MOVE! Protein dynamics and long-range allostery in cell signaling" (http://linkinghub.elsevier.com/retrieve/pii/B978-0-12-381262-9.00005-7). *Protein Structure and Diseases*. Advances in Protein Chemistry and Structural Biology. **83**. pp. 163–221. doi:10.1016/B978-0-12-381262-9.00005-7 (https://doi.org/10.1016%2FB978-0-12-381262-9.00005-7). ISBN 9780123812629. PMID 21570668 (https://pubmed.ncbi.nlm.nih.gov/21570668).

11. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (December 2009). "Hidden alternative structures of proline isomerase essential for catalysis" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805857). *Nature*. **462** (7273): 669–673. Bibcode:2009Natur.462..669F (https://ui.adsabs.harvard.edu/abs/2009Natur.462..669F). doi:10.1038/nature08615 (https://doi.org/10.1038%2Fnature08615). PMC 2805857 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805857). PMID 19956261 (https://pubmed.ncbi.nlm.nih.gov/19956261).

12. Donald, Voet (2011). *Biochemistry*. Voet, Judith G. (4th ed.). Hoboken, NJ: John Wiley & Sons. ISBN 9780470570951. OCLC 690489261 (https://www.worldcat.org/oclc/690489261).

13. Satir, Peter; Søren T. Christensen (26 March 2008). "Structure and function of mammalian cilia" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2386530). *Histochemistry and Cell Biology*. **129** (6): 687–93. doi:10.1007/s00418-008-0416-9 (https://doi.org/10.1007%2Fs00418-008-0416-9). PMC 2386530 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2386530). PMID 18365235 (https://pubmed.ncbi.nlm.nih.gov/18365235). 1432-119X.

14. Varadi, Mihaly; Vranken, Wim; Guharoy, Mainak; Tompa, Peter (1 January 2015). "Computational approaches for inferring the functions of intrinsically disordered proteins" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4525029). *Frontiers in Molecular Biosciences*. **2**: 45. doi:10.3389/fmolb.2015.00045 (https://doi.org/10.3389%2Ffmolb.2015.00045). PMC 4525029 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4525029). PMID 26301226 (https://pubmed.ncbi.nlm.nih.gov/26301226).

15. Protein Ensemble Database (https://web.archive.org/web/20180310010556/http://pedb.vib.be/)

16. Mittag, Tanja; Marsh, Joseph; Grishaev, Alexander; Orlicky, Stephen; Lin, Hong; Sicheri, Frank; Tyers, Mike; Forman-Kay, Julie D. (14 March 2010). "Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2924144). *Structure*. **18** (4): 494–506. doi:10.1016/j.str.2010.01.020 (https://doi.org/10.1016%2Fj.str.2010.01.020). ISSN 1878-4186 (https://www.worldcat.org/issn/1878-4186). PMC 2924144 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2924144). PMID 20399186 (https://pubmed.ncbi.nlm.nih.gov/20399186).

17. De Biasio, Alfredo; Ibáñez de Opakua, Alain; Cordeiro, Tiago N.; Villate, Maider; Merino, Nekane; Sibille, Nathalie; Lelli, Moreno; Diercks, Tammo; Bernadó, Pau (18 February 2014). "p15PAF is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3944474). *Biophysical Journal*. **106** (4): 865–874. Bibcode:2014BpJ...106..865D (https://ui.adsabs.harvard.edu/abs/2014BpJ...106..865D). doi:10.1016/j.bpj.2013.12.046 (https://doi.org/10.1016%2Fj.bpj.2013.12.046). ISSN 1542-0086 (https://www.worldcat.org/issn/1542-0086). PMC 3944474 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3944474). PMID 24559989 (https://pubmed.ncbi.nlm.nih.gov/24559989).

18. Kragelj, Jaka; Palencia, Andrés; Nanao, Max H.; Maurin, Damien; Bouvignies, Guillaume; Blackledge, Martin; Jensen, Malene Ringkjøbing (17 March 2015). "Structure and dynamics of the MKK7-JNK signaling complex" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4371970). *Proceedings of the National Academy of Sciences of the United States of America*. **112** (11): 3409–3414. Bibcode:2015PNAS..112.3409K (https://ui.adsabs.harvard.edu/abs/2015PNAS..112.3409K). doi:10.1073/pnas.1419528112 (https://doi.org/10.1073%2Fpnas.1419528112). ISSN 1091-6490 (https://www.worldcat.org/issn/1091-6490). PMC 4371970 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4371970). PMID 25737554 (https://pubmed.ncbi.nlm.nih.gov/25737554).

19. Allison, Jane R.; Rivers, Robert C.; Christodoulou, John C.; Vendruscolo, Michele; Dobson, Christopher M. (25 November 2014). "A relationship between the transient structure in the monomeric state and the aggregation propensities of α-synuclein and β-synuclein" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4245978). *Biochemistry*. **53** (46): 7170–7183. doi:10.1021/bi5009326 (https://doi.org/10.1021%2Fbi5009326). ISSN 1520-4995 (https://www.worldcat.org/issn/1520-4995). PMC 4245978 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4245978). PMID 25389903 (https://pubmed.ncbi.nlm.nih.gov/25389903).

20. Sivakolundu, Sivashankar G.; Bashford, Donald; Kriwacki, Richard W. (11 November 2005). "Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation". *Journal of Molecular Biology*. **353** (5): 1118–1128. doi:10.1016/j.jmb.2005.08.074 (https://doi.org/10.1016%2Fj.jmb.2005.08.074). ISSN 0022-2836 (https://www.worldcat.org/issn/0022-2836). PMID 16214166 (https://pubmed.ncbi.nlm.nih.gov/16214166).

21. Zhang, Gong; Ignatova, Zoya (1 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). ISSN 0959-440X (https://www.worldcat.org/issn/0959-440X). PMID 21111607 (https://pubmed.ncbi.nlm.nih.gov/21111607).

22. Alberts, Bruce; Alexander Johnson; Julian Lewis; Martin Raff; Keith Roberts; Peter Walters (2002). "The Shape and Structure of Proteins" (https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=mboc4%5Bbook%5D+AND+372270%5Buid%5D&rid=mboc4.section.388). *Molecular Biology of the Cell; Fourth Edition*. New York and London: Garland Science. ISBN 978-0-8153-3218-3.

23. Anfinsen, C. (1972). "The formation and stabilization of protein structure" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1173893). *Biochem. J.* **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).

24. Jaenicke, R.; Heber, U.; Franks, F.; Chapman, D.; Griffin, Mary C. A.; Hvidt, A.; Cowan, D. A. (1990). "Protein Structure and Function at Low Temperatures [and Discussion]" (https://doi.org/10.1098%2Frstb.1990.0030). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. **326** (1237): 535–553. doi:10.1098/rstb.1990.0030 (https://doi.org/10.1098%2Frstb.1990.0030). JSTOR 2398703 (https://www.jstor.org/stable/2398703). PMID 1969647 (https://pubmed.ncbi.nlm.nih.gov/1969647).

25. Kendrew, J.C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D.C. (1958). "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis". *Nature*. **181** (4610): 662–666. Bibcode:1958Natur.181..662K (https://ui.adsabs.harvard.edu/abs/1958Natur.181..662K). doi:10.1038/181662a0 (https://doi.org/10.1038%2F181662a0). PMID 13517261 (https://pubmed.ncbi.nlm.nih.gov/13517261). S2CID 4162786 (https://api.semanticscholar.org/CorpusID:4162786).

26. Krimm S, Bandekar J (1986). "Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins". *Advances in Protein Chemistry Volume 38. Adv. Protein Chem*. Advances in Protein Chemistry. **38**. pp. 181–364. doi:10.1016/S0065-3233(08)60528-8 (https://doi.org/10.1016%2FS0065-3233%2808%2960528-8). ISBN 9780120342389. PMID 3541539 (https://pubmed.ncbi.nlm.nih.gov/3541539).

27. Lessing, J.; Roy, S.; Reppert, M.; Baer, M.; Marx, D.; Jansen, T.L.C.; Knoester, J.; Tokmakoff, A. (2012). "Identifying Residual Structure in Intrinsically Disordered Systems: A 2D IR Spectroscopic Study of the GVGXPGVG Peptide" (https://www.rug.nl/research/portal/en/publications/identifying-residual-structure-in-intrinsically-disordered-systems(ff19c09b-088a-48f0-afee-2111a9b19252).html). *J. Am. Chem. Soc*. **134** (11): 5032–5035. doi:10.1021/ja2114135 (https://doi.org/10.1021%2Fja2114135). PMID 22356513 (https://pubmed.ncbi.nlm.nih.gov/22356513).

28. Jansen, T.L.C.; Knoester, J. (2008). "Two-dimensional infrared population transfer spectroscopy for enhancing structural markers of proteins" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242754). *Biophys. J.* **94** (5): 1818–1825. Bibcode:2008BpJ....94.1818J (https://ui.adsabs.harvard.edu/abs/2008BpJ....94.1818J). doi:10.1529/biophysj.107.118851 (https://doi.org/10.1529%2Fbiophysj.107.118851). PMC 2242754 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242754). PMID 17981904 (https://pubmed.ncbi.nlm.nih.gov/17981904).

29. 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).

30. Kumari I, Sandhu P, Ahmed M, Akhter Y (August 2017). "Molecular Dynamics Simulations, Challenges and Opportunities: A Biologist's Prospective". *Curr. Protein Pept. Sci*. **18** (11): 1163–1179. doi:10.2174/1389203718666170622074741 (https://doi.org/10.2174%2F1389203718666170622074741). PMID 28637405 (https://pubmed.ncbi.nlm.nih.gov/28637405).

31. Laskowski, RA (2011). "Protein structure databases". *Mol Biotechnol*. **48** (2): 183–98. doi:10.1007/s12033-010-9372-4 (https://doi.org/10.1007%2Fs12033-010-9372-4). PMID 21225378 (https://pubmed.ncbi.nlm.nih.gov/21225378). S2CID 45184564 (https://api.semanticscholar.org/CorpusID:45184564).

32. Murzin, A. G.; Brenner, S.; Hubbard, T.; Chothia, C. (1995). "SCOP: A structural classification of proteins database for the investigation of sequences and structures" (https://web.archive.org/web/20120426170732/http://scop.mrc-lmb.cam.ac.uk/scop/ref/1995-jmb-scop.pdf) (PDF). *Journal of Molecular Biology*. **247** (4): 536–540. doi:10.1016/S0022-2836(05)80134-2 (https://doi.org/10.1016%2FS0022-2836%2805%2980134-2). PMID 7723011 (https://pubmed.ncbi.nlm.nih.gov/7723011). Archived from the original (http://scop.mrc-lmb.cam.ac.uk/scop/ref/1995-jmb-scop.pdf) (PDF) on 26 April 2012.

33. Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. (1997). "CATH--a hierarchic classification of protein domain structures". *Structure*. **5** (8): 1093–1108. doi:10.1016/S0969-2126(97)00260-8 (https://doi.org/10.1016%2FS0969-2126%2897%2900260-8). PMID 9309224 (https://pubmed.ncbi.nlm.nih.gov/9309224).

34. Pascual-García, A.; Abia, D.; Ortiz, A.R.; Bastolla, U. (2009). "Cross-over between discrete and continuous protein structure space: insights into automatic classification and networks of protein structures" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2654728). *PLOS Computational Biology*. **5** (3): e1000331. Bibcode:2009PLSCB...5E0331P (https://ui.adsabs.harvard.edu/abs/2009PLSCB...5E0331P). doi:10.1371/journal.pcbi.1000331 (https://doi.org/10.1371%2Fjournal.pcbi.1000331). PMC 2654728 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2654728). PMID 19325884 (https://pubmed.ncbi.nlm.nih.gov/19325884).

35. Holm, L; Rosenström, P (July 2010). "Dali server: conservation mapping in 3D" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896194). *Nucleic Acids Research*. **38** (Web Server issue): W545–9. doi:10.1093/nar/gkq366 (https://doi.org/10.1093%2Fnar%2Fgkq366). PMC 2896194 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896194). PMID 20457744 (https://pubmed.ncbi.nlm.nih.gov/20457744).

36. Zhang Y (2008). "Progress and challenges in protein structure prediction" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2680823). *Curr Opin Struct Biol*. **18** (3): 342–348. doi:10.1016/j.sbi.2008.02.004 (https://doi.org/10.1016%2Fj.sbi.2008.02.004). PMC 2680823 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2680823). PMID 18436442 (https://pubmed.ncbi.nlm.nih.gov/18436442).

Further reading

- 50 Years of Protein Structure Determination Timeline - HTML Version - National Institute of General Medical Sciences (http://publications.nigms.nih.gov/psi/timeline_text.html) at NIH

External links

- Media related to Protein structures at Wikimedia Commons