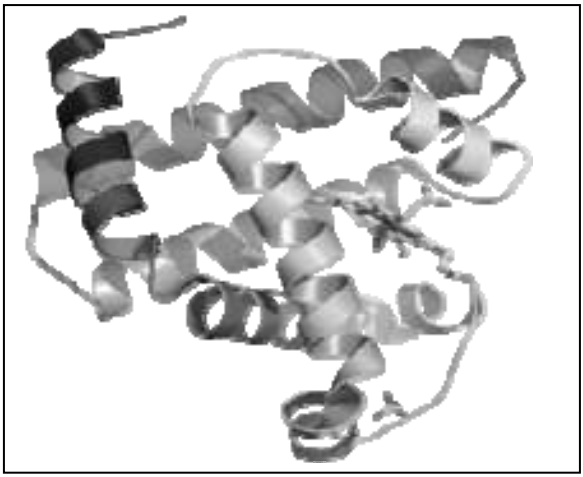

Figure 3: 3-D folding structure of myoglobin

III. Protein folding

1. Folding funnel theory of protein folding

How is a protein able to fold reliably into a predictable conformation? How can the mechanism be described in which the protein is carried from its unfolded random coil to a uniquely folded metastable state? Biochemical studies found that denatured proteins have all of their native three-dimensional structure disrupted. Yet, many of them refold efficiently and completely recover their biological activity when placed under conditions in which the folded form of the protein is stable.[43] Therefore, it is assumed that a native protein exists in some kind of thermodynamic configurational equilibrium. The biologically active state is the one with the lowest configurational energy.[44] The sequence of events guiding the protein folding is called the "protein folding pathway".[45] A random search among the entire conformation space for conformers would require an enormously long time.[46] Proteins, however, are able to

43 Berg, Jeremy M./Tymoczko, John L./Stryer Lubert, Biochemistry, New York, NY, 2005, 72-73.

44 Levinthal, Cyrus, Are there Pathways for Protein Folding? 65 Journal de Chimie Physique 1968, 44, 44.

45 Levinthal, Cyrus, Are there Pathways for Protein Folding?, 65 Journal de Chimie Physique 1968, 44, 44.

46 Although the protein is able to sample new configurations very fast, it will take at least 1027 years to try them all, see Zwanzig, R./Szabo, A./Bagchi, B., Levinthal's paradox, 89 Proceed-

fold within milliseconds to seconds. This implies that only a small amount of conformation space is sampled during the folding process. The problem of how proteins fold rapidly into their three-dimensional conformation despite the infinite number of possible configurations is described as the *Levinthal-Paradox*.[47] Each bond connecting amino acids can occur in several possible states.[48]

Several models attempt to explain the phenomenon of protein folding. A more recent model approaches the issue through a so-called "folding funnel" theory.[49] The conformational energy surface of a protein folding pathway is graphically displayed as a funnel. Convergent kinetic pathways guide the folding to a unique, stable, native conformation.[50] It is assumed that the random polypeptide chain first collapses in a dense structure. Native bonds emerge when fluctuation of the peptide chain randomly associates distant polypeptide sequences. Each native conformation stabilizes the chain and simultaneously narrows the conformational space. Thus, the random search for all further native conformations occurs more rapidly. Through the gradual native configuration, the peptide chain is efficiently transformed into its three-dimensional structure.

As already mentioned, the collapse of the primary amino acid structure into the tertiary folding state can be illustrated through an energy landscape that has the image of a folding funnel. The wide rim demonstrates the multitude of accessible conformation pathways that exist initially. The narrow bottom shows the minimum of configuration flexibility, which mirrors the final state.[51] Each protein follows a different folding path as it approaches its specific native structure. The exact nature of

ings of the National Academy of Science of the United States of America 1992, 20-22; Nienhaus, Ulrich, Physik der Proteine, 3 Physik Journal 2004, 37, 39.

47 Nienhaus, Ulrich, Physik der Proteine, 3 Physik Journal 2004, 37, 39; Zwanzig, R./Szabo, A./Bagchi, B., Levinthal's paradox, 89 Proceedings of the National Academy of Science of the United States of America 1992, 20-22; Levinthal, Cyrus, Are there Pathways for Protein Folding?, 65 Journal de Chimie Physique 1968, 44-45.

48 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 403-404.

49 Onuchic, J. N./Luthey-Schulten, Z./Wolynes, P. G., Theory of Protein Folding: The Energy Landscape Perspective, 48 Annual Revue of Physical Chemistry 1997, 545; Leopold, E. Peter/Montal, Mauricio/Nelson ONuchic, José Nelson, Protein Folding Funnels: A Kinetic Approach to the Sequence-structure Relationship, 89 Proceedings of the National Acadamy of Science of the United States of America 1992, 8721; Nienhaus, Ulrich, Physik der Proteine, 3 Physik Journal 2004, 37, 39.

50 Leopold, E. Peter/Montal, Mauricio/Nelson ONuchic, José Nelson, Protein Folding Funnels: A Kinetic Approach to the Sequence-structure Relationship, Proceedings of the National Acadamy of Science of the United States of America 1992, 8721, 8721.

51 Nienhaus, Ulrich, Physik der Proteine, 3 Physik Journal 2004, 37, 39-40.

these differences depends on the protein's size, stability, and structure.[52] Figure 4 illustrates this process:[53]

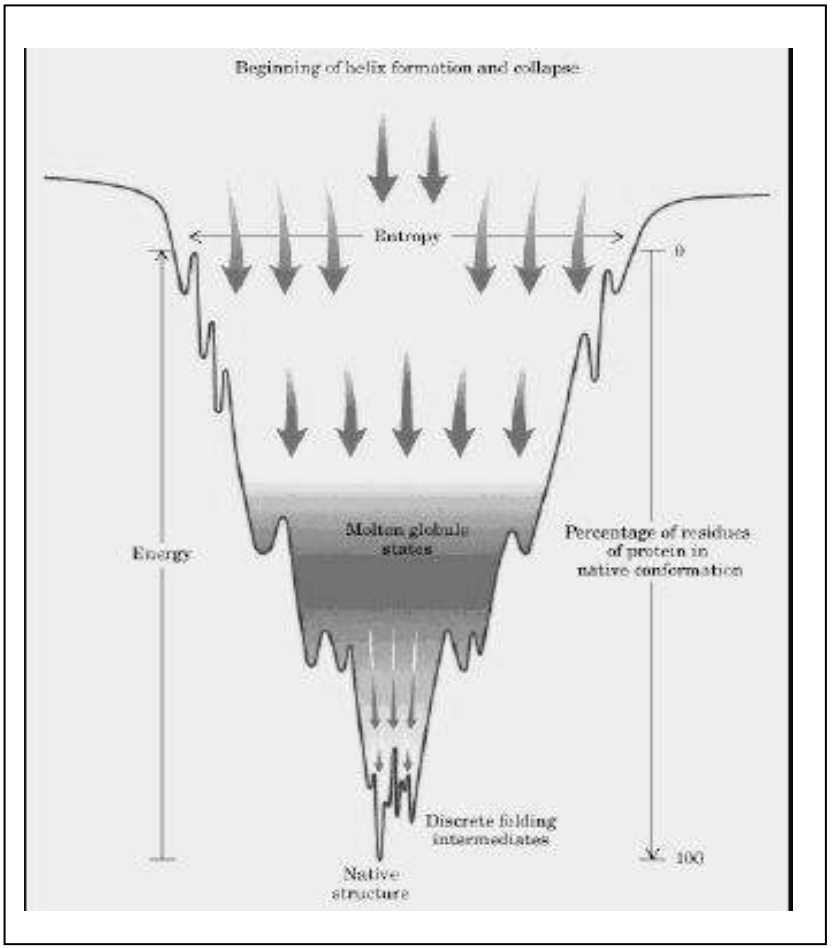

Figure 4: Energy landscape of a protein

52 For a statistical energy landscape approach that explains when and why certain processes such as specific folding pathways occur in some proteins, see Bryngelson, J. D./Onuchic, J. N./Socci, N. D./Wolynes, P. G., Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis, 21 Proteins 1995, 167-195.

53 Based on a figure from Kavraki, Lydia E., Protein folding, available at http://cnx.org/content/m11467/latest/.

2. Protein misfolding and diseases arising from 'folding' defects

With a better understanding of folding, scientists realized that diseases arise because of misfolding.[54] Already small structural defects can give rise to a wide range of folding diseases. Genetic diseases, such as cystic fibrosis and sickle cell anemia, are typically caused by mutations within coding regions.[55] Protein misfolding also plays a crucial role in the pathogenesis of prion diseases.[56] A prion is a protein part that lacks nucleic acid.[57] Normally, it occurs in a harmless form, but its misfiled variation has been identified as the cause of various neurodegenerative disorders such as scrapie, bovine spongiform encephalopathy (BSE, its human equivalent, the Creutzfeld-Jakob disease (CJD), and Kuru. The progression of these diseases is accompanied by the appearance of insoluble protein plaques in the brain (amyloid plaques).[58] The specific protein that has been isolated from the protein plaque is the PrP protein, which has the ability to exist in two stable forms, PrP-C and PrP-Sc. The sequences of PrP-C and PrP-Sc are found to be identical. Biochemical researchers discovered that the brain plaques contain PrP-Sc. All experiments showed that the Prp-Sc was not from an external source, but expressed by the host cell itself. This form of the protein was the only cause of infection in the prion diseases. The disease-specific feature is consequently not the expression of the prion protein, but rather its biophysical and biochemical characteristics. A purely structural change is assumed to cause its aggregation in the brain.[59] The ability of PrP to exist in two stable forms and the fact that the disease-specific feature does not depend on the genetic coding

54 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 426.

55 Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 426.

56 Tatzelt, Jorg/Winklhofer, Konstanze F., Folding and Misfolding of the Prion Protein in the Secretory Pathway, 11 Amyloid 2004, 162-172.

57 The term 'prion' is derived from 'proteinaceous infectious particle, see Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 431. For a definition, see Medline Plus, Medical Dictionary, available at: http://www.nlm.nih.gov/medlineplus/mplusdictionary.html, last checked on January 21, 2008.

58 Scarpie occurs in sheep and leads to a progressive loss of motoric coordination, finally ending in an inability to stand unsupported. It was first identified in the 17th century in the United Kingdom, with similar forms discovered more recently in other animals, such as mink, deer and elk. CJD typically occurs in humans above 50 years of age. Infected persons exhibit dementia and loss of motoric coordination. CDJ was first described in the 1920s. Kuru or 'the laughing dead' occurred in the 1960 in the Highlands of Papua New Guinea, where the conduct of cannibalism was held to be responsible for the progression of the disease. A decline in cannibalism resulted in a decline in Kuru, although the precise agent has yet not been identified, as described in Whitford, David, Proteins: Structure and Function, Chichester, West Sussex, U.K., 2005, 433. For the potential causes of scarpie, see also Prusiner, Stanley B., Novel Proteinaceous Infectious particles cause scrapie, 216 Science 1982, 136.

59 Tatzelt, Jorg/Winklhofer, Konstanze F., Folding and Misfolding of the Prion Protein in the Secretory Pathway, 11 Amyloid 2004, 162, 162.

seems contrary to the long-held hypothesis that an amino acid codes for a single unique 3-D structure.[60] Amyloid plaques or protein aggregations in the brain are also associated with Alzheimer's disease and Parkinson's disease, conditions not considered prion-based but also dependent on aberrant protein folding.[61] The primary structure does not absolutely determine the tertiary folding structure. It is now widely believed that gene expression alone largely, but not exclusively, controls the protein's 3-D properties.

III. Structurally similar, sequence dissimilar proteins

With the discovery of increasingly more protein structures, it has further become evident that many proteins that possess similar structures share only a very small number of identical residues in structurally associated positions.[62] Various structurally similar protein pairs have only a minimal amount of sequence identity. This suggests that many sequence positions do not play a significant role in structure determination, and folding determinants are restricted to a limited number of sequence residues.[63] Structurally similar proteins do therefore not necessarily reflect sequence-similar proteins.[64] Some proteins bearing diverse sequences with essentially no sequence homology, do fold into the same structure. With the protein's effect depending on the structure, large numbers of different proteins are able to perform the same functions.[65]

IV. Posttranslational modifications (PTM)

An important component of protein regulation and function is the modification of protein structures, which occur either co- or posttranslationally. Translation refers to

60 Tatzelt, Jorg/Winklhofer, Konstanze F., Folding and Misfolding of the Prion Protein in the Secretory Pathway, 11 Amyloid 2004, 162, 166.

61 Tatzelt, Jorg/Winklhofer, Konstanze F., Folding and Misfolding of the Prion Protein in the Secretory Pathway, 11 Amyloid 2004, 162, 162.

62 Jaenichen, Hans-Rainer/Mcdonell, Leslie A./Haley, James F., Jr., From Clones to Claims, Cologne, Berlin, Bonn, Munich 2002, 167; molecular biologists thus attempted to identify the common hidden information within these sequences that directs them to assume similar folds.

63 Kleist, Peter, Biomarker und Surrogat-Endpunkte: Garanten für eine schnellere Zulassung von neuen Arzneimitteln?, 83 Schweizerische Ärztezeitung 2022, 2347, 2350.

64 Wachenfeld, Joachim, The Patenting of Protein Structures, http://www.vossiusandpartner.com/eng/publication/mip-yearbook.html 2002.

65 Structural protein families are also called 'protein superfamilies'; see: Hultquist, Steven J./Robert Harrison, and Yongzhi Yang, Patenting Bioinformatic Inventions: Emerging Trends in the United States, 20 Nature Biotechnology 2002, 743; 771. A list of protein superfamilies with structure-based-sequence-alignment is available at: http://www-cryst.bioc.cam.ac.uk/~campass/superfamily.html, last checked on May 06, 2005.