

Outline and computational approaches of protein misfolding

Xin LIU

The State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2010

Abstract Protein misfolding is a general causation of classical conformational diseases and many pathogenic changes that are the result of structural conversion. Here I review recent progress in clinical and computational approaches for each stage of the misfolding process, aiming to present readers an outline for swift comprehension of this field.

Keywords computational approaches, protein misfolding, conformational diseases

investigated in the context of structural change, such as the H5N1 avian influenza virus (Bornholdt and Prasad, 2008) and A(H1N1) global pandemic outbreak in 2009 (Garten et al, 2009; Liu and Zhao, 2010c). Therefore, CDs are not rare, but are responsible for the development of a wide range of diseases.

Several factors are related to the pathogenic structural change of protein, including multiple stable states, lifespan, molecular environment, evolution, and so on. As each of them is extremely significant, it is hard to make a rank of their importance.

1 Introduction

Proteins are the most abundant bio-molecules in biology. There are about 100 000 different types of proteins that take part in nearly every chemical process of our life (Brändén and Tooze, 1999). The majority of proteins must fold into compact structures to perform their bio-functions. As the bio-chemical reactions among bio-molecules require the atoms that react to each other to be close less than an angstrom scale, the protein structures are extremely specific as if the conformations are the elaborate scaffolds of the corresponding reaction. Therefore, a misfolding of protein can induce an alteration of protein's bio-properties, and result in disease consequently.

As a general causation of illness, protein misfolding is responsible for various types of diseases. Some proteins that have distinct structural change were reported as pathogenic molecules in classical cases of conformational diseases (CDs), such as prion for transmissible spongiform encephalopathy (TSE), and β_2 -microglobulin for dialysis-related amyloidosis. More than 30 different human diseases are related to conformational conversion (Thomas et al., 1995; Kelly, 1996; Carrell and Lomas, 1997; Carrell and Gooptu, 1998; Soto, 2001). Moreover, besides these classical cases, there are lots of diseases that can be

1.1 Multiple stable states

From a physical point of view, a sequence folds into its native structure that has a free energy minimum. In the folding process, a polypeptide chain makes a stochastic search of many conformations that are accessible (Wolynes et al., 1995; Dill and Chan, 1997; Karplus, 1997; Dobson et al., 1998; Dobson, 2004). Consequently, the free energy of the polypeptide chain can be described as a function of its conformational properties. As shown by the energy landscape in Fig. 1, there are usually several metastable states for a protein related to pathogenic structural change. It can be deemed that a protein can be in any of these states, but with different probabilities. Once the protein folds into a conformation other than the 'healthy' state, disease can be developed potentially.

1.2 Lifespan

Some wild type proteins, such as the α -synuclein that is associated with Parkinson's disease (Hardy et al., 2006), can induce conformational disease in old age. On the other hand, some fatal disease, such as hereditary cerebral angiopathy (Abrahamson, 1996; Ólafsson and Grubb, 2000) and TSE, are reported with sporadic events. There are evidences that conformational disease is correlated tightly with human lifespan (Liu and Zhao, 2010b). All proteins can fold into pathogenic state, and have an

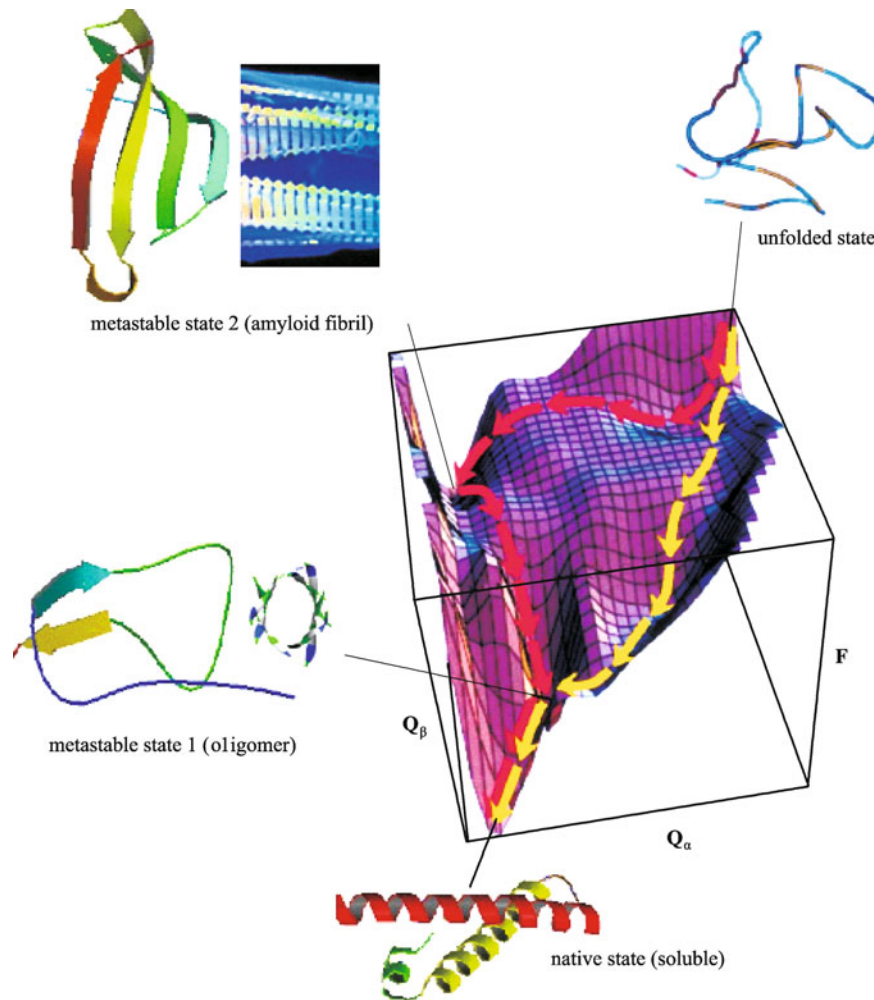


Fig. 1 Sketch map of the energy landscape for disease-related proteins. The free energy (F) is shown as a function of the number of native contacts (Q_α and Q_β) in domain α and β . The metastable state 1 denotes the oligomer in early step of protein aggregation that is highly toxic in amyloidosis. Proteins at metastable state 2 aggregate into amyloid fibrils by ‘Steric zipper’ β -sheets architecture. The red and yellow trajectories denote the fast and slow tracks in folding pathway of the corresponding protein (Dinner et al, 2000).

inherent tendency to aggregate (Dobson, 1999). Selection pressure of evolution has made a filtration, and resulted in proteins that can resist aggregation and other pathogenic state during normal lifespan. The incidence of fatal diseases has been depressed by low probabilities of changing structures from native to morbid states (Liu and Zhao, 2010b). While evolution can do no better than the necessity of enabling us to transmit our genes to offsprings (Dobson, 2002); thus, to prolong life is to cope with the proliferation of these diseases, and to challenge the nature of evolution.

1.3 Molecular environment

The folding process depends on the environment in which the folding takes place. Protein folding beginning as a nascent chain is still attached to ribosome (Hardesty and Kramer, 2001), and complete the major part step by step

after release from the ribosome. As only partially fold, some regions that are buried in the native state can be exposed. Such structures are prone to contact with other molecules inappropriately (Hore et al., 1997; Capaldi et al., 2002). But the living systems have developed a range of elaborate strategies to make the completion of correct folding prior to mistake interactions (Gething and Sambrook, 1992; Hartl and Hayer-Hartl, 2002; Dobson, 2003), such as using molecular chaperones and folding-accelerate catalyst. A typical example is aiding the folding process by molecular chaperonin GroEL that contains a cavity in which incompletely folded polypeptide chains can be held and protected from the outside world. There are large numbers of molecular chaperones that are present in all types of cells and cellular compartments (Dobson, 2004), and nurse the proteins by interacting with nascent chains as they emerge from the ribosome, or by binding non-specifically to protect aggregation-prone regions, and

so on. Error in such quality control mechanism can cause disease.

1.4 Evolution

Conformational disease is a phenomenon of evolution. Therefore, evolution is a major factor of protein misfolding either for the past, the present, or the future. In modern life, there are lots of practices that are not experienced during previous evolution, such as new agricultural practices (Prusiner, 1997), a changing diet associated with type-II diabetes (Höppener et al., 2002), new medical procedures associated with iatrogenic Creutzfeldt-Jakob diseases (Prusiner, 1997). As the practices are introduced much more rapid than evolution, we fatally have not enough time to set up effective protective mechanisms (Dobson, 2002).

Many efforts have been made to investigate protein misfolding in computational approaches. Due to the increased power and accuracy, such approaches have attracted many concerns. In order to present readers an outline of this field, I review the recent developments and suggest several points for further studies.

2 The pathological structure of disease-related protein

2.1 Clinical reports

There are several points that have been reported clinically as the features of pathological structure of disease-related protein.

2.1.1 Aggregation

One of the typical features of conformational disease is the formation of insoluble protein, amyloid, that deposits in tissues. For example, the normal prion protein is found on the membranes of cells throughout the body, even in healthy people and animals. In TSE, the soluble cellular isoform of prion (PrP^C) folds inappropriately into the scrapie isoform (PrP^{Sc}), which accumulates and forms fibrils in brain tissue, causes tissue damage and cell death, and leads to degeneration of nerve system. In this process, the normal alpha helix rich PrP^C changes its structure into the pathological beta sheet rich PrP^{Sc} (Kuwata et al., 2007). Similarly, there are obviously structural changes in many classical disease-related proteins also, such as insulin and serpins. Whereas, there are also amyloidogenic proteins that form fibrils in their native state in globular form, in which the local structural change is significant, e.g. β_2 microglobulin, transthyretin, lysozyme (Thomas et al., 1995; Kelly, 1996; Carrell and Lomas, 1997; Carrell and Gooptu, 1998; Soto, 2001).

2.1.2 ‘Steric zipper’ β -sheets amyloid architecture

It was found that the fibrillar structures of various proteins have very similar morphologies, whereby pairs of parallel/antiparallel β -sheets form a dry interface running perpendicular to the fibril axis (Nelson et al., 2005; Sawaya et al., 2007). It is clear that the core structure of the fibrils is stabilized primarily by interactions, particularly hydrogen bonds, involving the polypeptide main chain. As the main chain is common to all polypeptides, this observation explains why fibrils formed by polypeptides of very different amino acid sequences are similar in appearance (Dobson, 2004).

2.1.3 High toxicity of the early pre-fibrillar aggregate

The insoluble protein mass can disrupt the functioning of specific organs (Pepys, 1995), or result in the loss of functional protein that leads to the failure of some crucial cellular process (Thomas et al., 1995). It has been suggested that the early pre-fibrillar aggregates of proteins are highly damaging to cell (Koo et al., 1999; Caughey and Lansbury, 2003). Moreover, it has become clear that the pre-fibrillar aggregates are toxic through a less specific mechanism, such as the exposure of non-native hydrophobic surface (Polverino et al., 2003; Stefani and Dobson, 2003). For example, the pre-fibrillar aggregates of several non-disease-related proteins can be as cytotoxic as those of amyloid β -protein (Bucciantini et al., 2002). By contrast, the mature fibrils have much lower toxicity than that of their precursors (Walsh et al., 2002; Caughey and Lansbury, 2003).

2.1.4 Folding nucleus is distinct from aggregation nucleus

The roles of individual residues in the folding process have been investigated by site-directed mutagenesis. A wide range of studies suggest that there are a small number of key residues, which form the folding nucleus of a protein (Matouschek et al., 1989; Fersht, 1999, 2000); the collapse of the polypeptide chain to stable compact structure can occur only after the majority of the folding nucleus have been formed, i.e., the native structure is a consequence of the formation of folding nucleus. If these key interactions are not formed, the protein cannot usually fold directly to a stable globular structure. It prevents protein misfolding by prolonging the unfolded state. As a result of such ‘quality control’ process, native structure is formed prior to the incorrect one (Vendruscolo et al., 2001; Davis et al., 2002; Makarov and Plaxco, 2003). On the other hand, investigations of the mechanism of amyloid formation suggest that there are aggregation-prone regions, ‘hot spots’ of fibril formation, that are considered to be responsible for aggregation (Ivanova et al., 2004; Ventura et al., 2004).

An important observation is that the residues of the folding nucleus are distinct from those of aggregation (Chiti et al., 2002). It means that the evolutionary pressure may select sequences that favor the assembly process of folding other than aggregation.

Besides the aforementioned points, there are also other pathogenic features in classical disease-related proteins, such as the protein unstable in haemoglobin and serpins, structural topological change in apolipoprotein AI. As numerous efforts have been focused on the classical cases of conformational diseases, their features of pathological structure have been revealed more than those of non-classical one. In the latter, it has been suggested that the pathological structure of disease-related protein can introduce new targets to human immune system, and be associated with the highly pathogenic H5N1 avian influenza and A(H1N1) 2009 global pandemic (Liu and Zhao, 2010c).

2.2 Computational approaches

Since aggregation appears in the majority of classical conformational diseases, amyloidogenic mechanism is the main focus of computational approaches.

2.2.1 Prediction of amyloid core

The discovery of aggregation-prone region has promoted the development of a number of algorithms and models for predicting the aggregation propensity of proteins (Fernandez-Escamilla et al., 2004; López and Serrano, 2004; Yoon and Welsh, 2004; Pawar et al., 2005; Sánchez et al., 2005; Bemporad et al., 2006; Caffisch, 2006; Galzitskaya et al., 2006; Saiki et al., 2006; Zhang et al., 2007). Some of them possess quite good capability in identifying residues that are buried in the amyloid core, and have become facility tools in the analysis of pathological structure. As the amyloid core observed in experimentation often contains a large number of residues, it is difficult to rank the significance of individual residues in the amyloidogenic mechanism. Complementarily, the pre-residue score in these algorithms may provide such information, and aid in the comprehension of amyloid in the coming efforts.

2.2.2 Prediction of pathological structure

Because there is usually a block of proteins in the amyloid, the purification of pathological structures is quite difficult in experiment. Thus, there is a shortage of high quality coordinates. Computational approach has been an efficient way in uncovering such information. For example, in contrast to the high-resolution data for the PrP^C, the structure of PrP^{Sc} is largely unknown. Huang, Prusiner, and Cohen have developed a three-dimensional model of PrP^{Sc}

with a combination of computational techniques and experimental data (Huang et al., 1996). However, there are lots of unsolved problems, and it remains a hot spot in recent studies (Smirnovas et al., 2009).

2.2.3 Investigation of interactions in fibril

Computational approaches have gained great achievement in the study of molecular architecture of protein amyloid, particularly for cases in lack of pathological structure. For example, the molecular architecture of PrP^{Sc} amyloid has been investigated with theoretical models by Govaerts et al. and DeMarco et al. (DeMarco and Daggett, 2004; Govaerts et al., 2004). Features of the interaction surface in amyloidogenic regions can also be investigated by a combination of several different algorithms (Castillo and Ventura, 2009).

2.2.4 Investigation of the early steps of aggregation

Since the early oligomers during the aggregation process are the primary toxic species in amyloidosis, the investigation focusing on the initial assemblies of oligomers is one of the hot spots at present. However, the transient/premature atomic details are difficult to be characterized using biophysical methods (Hardy and Selkoe, 2002; Bitan et al., 2003; Mastrangelo et al., 2006). Therefore, computational approaches contribute great to this field. Although only very short polypeptide (usually less than ten residues) can be simulated, such efforts have achieved some patterns of self-assembly, which could be useful for the studies of true amyloidosis (Wei et al., 2007).

2.3 Computational assistant treatment

Scientists have attempted to prove their theory of amyloidosis with experimental efforts. In 2002, López et al. designed amyloid hexapeptide sequences using a computer-designed algorithm. Sequences with a high propensity to form homopolymeric β -sheets were validated experimentally. It has been shown that the *de novo* designed peptide self-associates efficiently into β -sheets. Whereas, some point mutations that were predicted to be unfavorable for fibrils inhibited the polymerization. The delicate balance of interactions involved in fibrils formation to those in more disorder aggregates was uncovered (López et al., 2002). Other evidences suggested that the increased aggregation propensity and the decreased stability of amyloid protein are significantly correlated with the decreased patient survival (Meiering, 2008). Therefore, the computational assistant analysis and treatment may promise new clinical therapeutics in amyloidosis.

3 The switch of pathogenic structural change

3.1 Clinical reports

Either the preliminary or the later stage of the process of pathological conformational conversion, e.g. the protein aggregation, is the result of switching on the misfolding pathway of the native state. It is ideal for a treatment to cure a disease by prohibiting proteins falling into the misfolding pathway; that is, by preventing the disease-related misfolding other than coping with the mass of subsequent pathological changes of misfolding. As sites significant for switching on the misfolding pathway can be perfect binding targets for drugs in clinical treatment, many efforts have been made to note them experimentally, using information such as disease-related point mutations observed in clinical practice, sites involved in the inhibition of the pathogenic change process, and the abnormal cleavage site responsible for amyloidosis (Liu and Zhao, 2010c).

A typical study dealing with switch sites was reported by Kuwata et al. (Kuwata et al., 2007). Based on a series of relevant works, the authors selected 14 amino acid residues for an in-depth investigation. The switch region that is responsible for the pathogenic structural conversion of prion protein is identified as a pocket formed by five residues. The intercalation of an anti-prion compound GN8 to this pocket can inhibit the pathogenic change process of the prion protein, and prolong the survival of TSE-infected mice.

There is evidence that the amyloid-related mutations are not necessarily involved in aggregation-prone regions. So it is still not clear whether switch sites occur in hot spots of aggregation or not (Ivanova et al., 2004; Ventura et al., 2004; Liu and Zhao, 2010b). It is also an open question whether such a difference is related to the distinction between folding nucleus and aggregation nucleus.

3.2 Prediction of switch region responsible for pathogenic structural change

Usually, the identification of switch sites needs large amount of experiments. The literatures may even conflict with each other. Moreover, the long duration is another important restriction. A prediction algorithm of switch region can help the comprehension of clinical reports, speed up the clinical investigation, decrease the hardness and knowledge threshold especially for the inexperienced and people in interdisciplinary sciences. However, such prediction is difficult and in absence for a long time because multiple factors must be jointly considered.

Based on a joint consideration of protein stability and the selection pressure of protein evolution, we have developed the first algorithm in the prediction of switch

region responsible for pathogenic structural change (Liu and Zhao, 2010c). The remote homologous relationships among polypeptides can be identified with a high accuracy based on the discovery of the significant role of molecular mechanics properties in protein evolution (Liu and Zhao, 2009a, 2010a). Using this highly accurate algorithm, it was revealed that there are only two major clusters in the phase space of polypeptide: the helix-donut zone that consists of the helix segments and the N/C-terminal helix caps, and the strand-arc part that is mainly comprised of β -sheet segments and the N/C terminal strand caps. A query protein is treated as successive residue segments (Liu and Zhao, 2009b). In the native fold of a protein, each segment belongs to one of the clusters in the phase space. Its probability of being in the other cluster determines the capability of each segment in arousing pathogenic structural change.

The algorithm can identify the residue segments that are responsible for the start of pathogenic structural changes with an accuracy of 94%, and find the residues that are tightly associated with conformational diseases about eight times the capability of random dicing (Liu and Zhao, 2010b). It would be a useful tool in identifying the riskiest region of a protein, and form a foundation for further investigation.

4 Summary and outlook

Here I review the investigations of protein misfolding in a full process of the formation of pathogenic change, including studies about switching on the misfolding pathway, the early step of aggregation, the final pathological structure per molecule, hot spots of amyloid, the interactions in morbid polymer, and architecture of amyloid. This review provides typical references of clinical and computational analysis in each step therein, and singles out some important conceptions in the comprehension of protein misfolding, such as multiple stable states, energy landscape, contribution of protein evolution, high toxicity of the early pre-fibrillar aggregates, amyloid architecture, the difference of folding nucleus and aggregation nucleus, and the difference of switch region and amyloid core. I hope it is helpful for the readers to comprehend the outline of pathogenic structural change and conformational disease swiftly.

As shown in the aforementioned text, there are computational efforts for each step of the misfolding process. Although the present letter mainly focuses on the investigation of the disease-related proteins, there are notable progress in the algorithm of protein-protein interaction (Bonvin, 2006; Gray, 2006), medicine design (Available Chemicals Directory, MDL Information System, San Leandro, CA), toxicity evaluation (Toxicity, www.symyx.com), and so on. One of the bottle-neck of

computational approach is that, as an all-around study cannot be accomplished with the present computational technique, the significant sites must be selected according to experimental reports. With the development of the algorithm of switch region prediction, some significant sites can be identified computationally. As conformational change is responsible for many diseases, this would benefit large amount of research. Investigating diseases in an aspect of structural change can be a promising methodology in pathology (Liu and Zhao, 2010c).

References

- Abrahamson M (1996). Molecular basis for amyloidosis related to hereditary brain hemorrhage. *Scand J Clin Lab Invest*, 226: 47–56
- Bemporad F, Calloni G, Campioni S, Plakoutsi G, Taddei N, Chiti F (2006). Sequence and structural determinants of amyloid fibril formation. *Acc Chem Res*, 39: 620–627
- Bitan G, Vollers S S, Teplow D B (2003). Elucidation of primary structure elements controlling early amyloid-beta protein oligomerisation. *J Biol Chem*, 278: 34882–34889
- Bonvin A M (2006). Flexible protein-protein docking. *Curr Opin Struct Biol*, 16: 194–200
- Bornholdt Z A, Prasad B V V (2008). X-ray structure of NS1 from a highly pathogenic H5N1 influenza virus. *Nature*, 456: 985–988
- Brändén C I, Tooze J (1999). Introduction to protein structure. 2nd ed. New York: Garland Publishing
- Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, Stefani M (2002). Inherent cytotoxicity of aggregates implies a common origin for protein misfolding diseases. *Nature*, 416: 507–511
- Cafilisch A (2006). Computational models for the prediction of polypeptide aggregation propensity. *Curr Opin Chem Biol*, 10: 437–444
- Capaldi A P, Kleanthous C, Radford S E (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol*, 9: 209–216
- Carrell R W, Gooptu B (1998). Conformational changes and diseases—serpins, prions, and Alzheimer's. *Curr Opin Struct Biol*, 8: 799–809
- Carrell R W, Lomas D A (1997). Conformational disease. *Lancet*, 350: 134–138
- Castillo V, Ventura S (2009). Amyloidogenic Regions and Interaction Surfaces Overlap in Globular Proteins Related to Conformational Diseases. *PLoS Comput Biol*, 5: e1000476
- Caughey B, Lansbury P T (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci*, 26: 267–298
- Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, Dobson C M (2002). Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol*, 9:137–143
- Davis R, Dobson C M, Vendruscolo M (2002). Determination of the structures of distinct transition state ensembles for a β -sheet peptide with parallel folding pathways. *J Chem Phys*, 117: 9510–9517
- DeMarco M L, Daggett V (2004). From conversion to aggregation: Protofibril formation of the prion protein. *Proc Natl Acad Sci USA*, 101:2293–2298
- Dill K A, Chan H S (1997). From Levinthal to pathways to funnels. *Nat Struct Biol*, 4: 10–19
- Dinner A R, Sali A, Smith L J, Dobson C M, Karplus M (2000). Understanding protein folding via free energy surfaces from theory and experiment. *Trends Biochem Sci*, 25: 331–339
- Dobson C M (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci*, 24: 329–332
- Dobson C M (2002). Getting out of shape—protein misfolding diseases. *Nature*, 418: 729–730
- Dobson C M (2003). Protein folding and disease: a view from the first Horizon symposium. *Nat Rev Drug Discov*, 2: 154–160
- Dobson C M (2004). Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Bio*, 15: 3–16
- Dobson C M, Sali A, Karplus M (1998). Protein folding: a perspective from theory and experiment. *Angew Chem Int Ed Eng*, 37: 868–893
- Fernandez-Escamilla A M, Rousseau F, Schymkowitz J, Serrano L (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*, 22: 1302–1306
- Fersht A (1999). Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: WH Freeman
- Fersht A R (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci USA*, 97:1525–1529
- Galzitskaya O V, Garbuzynskiy S O, Lobanov M Y (2006). Is it possible to predict amyloidogenic regions from sequence alone? *J Bioinform Comput Biol*, 4: 373–388
- Garten R J, Davis C T, Russell C A, Shu B, Lindstrom S, Balish A, Sessions W M, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith C B, Emery S L, Hillman M J, Rivaller P, Smagala J, de Graaf M, Burke D F, Fouchier R A, Pappas C, Alpuche-Aranda C M, López-Gatell H, Olivera H, López I, Myers C A, Faix D, Blair P J, Yu C, Keene K M, Dotson P D Jr, Boxrud D, Sambol A R, Abid S H, St George K, Bannerman T, Moore A L, Stringer D J, Blevins P, Demmler-Harrison G J, Ginsberg M, Kriner P, Waterman S, Smole S, Guevara H F, Belongia E A, Clark P A, Beatrice S T, Donis R, Katz J, Finelli L, Bridges C B, Shaw M, Jernigan D B, Uyeki T M, Smith D J, Klimov A I, Cox N J (2009). Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science*, 325: 197–201
- Gething M J, Sambrook J (1992). Protein folding in the cell. *Nature*, 355: 33–45
- Govaerts C, Wille H, Prusiner S B, Cohen F E (2004). Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci USA*, 101: 8342–8347
- Gray J J (2006). High-resolution protein-protein docking. *Curr Opin Struct Biol*, 16: 183–193
- Hardesty B, Kramer G (2001). Folding of a nascent peptide on the ribosome. *Prog Nucleic Acid Res Mol Biol*, 66:41–66
- Hardy J, Cai H, Cookson M R, Gwinn-Hardy K, Singleton A (2006). Genetics of Parkinson's disease and parkinsonism. *Ann Neurol*, 60: 389–398
- Hardy J, Selkoe D J (2002). The amyloid hypothesis of Alzheimer's

- disease: progress and problems on the road to therapeutics. *Science*, 297: 353–356
- Hartl F U, Hayer-Hartl M (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, 295: 1852–1858
- Hore P J, Winder S L, Roberts C H, Dobson C M (1997). Stopped-flow photo-CIDNP observation of protein folding. *J Am Chem Soc*, 119: 5049–5050
- Höppener J W, Nieuwenhuis M G, Vroom T M, Ahrén B, Lips C J (2002). Role of islet amyloid in type 2 diabetes mellitus: consequence or cause? *Mol Cell Endocrinol*, 197: 205–212
- Huang Z W, Prusiner S B, Cohen F E (1996). Scrapie prions: a three-dimensional model of an infectious fragment. *Fold Des*, 1: 13–19
- Ivanova M I, Sawaya M R, Gingery M, Attinger A, Eisenberg D (2004). An amyloid-forming segment of beta2-microglobulin suggests a molecular model for the fibril. *Proc Natl Acad Sci USA*, 101: 10584–10589
- Karplus M (1997). The Levinthal paradox, yesterday and today. *Fold Des*, 2: S69–S76
- Kelly J W (1996). Alternative conformations of amyloidogenic proteins govern their behavior. *Curr Opin Struct Biol*, 6: 11–17
- Koo E H, Lansbury P T Jr, Kelly J W (1999). Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc Natl Acad Sci USA*, 96: 9989–9990
- Kuwata K, Nishida N, Matsumoto T, Kamatari Y O, Hosokawa-Muto J, Kodama K, Nakamura H K, Kimura K, Kawasaki M, Takakura Y, Shirabe S, Takata J, Kataoka Y, Katamine S (2007). Hot spots in prion protein for pathogenic conversion. *Proc Natl Acad Sci USA*, 104: 11921–11926
- Liu X, Zhao Y P (2009a). A scheme for multiple sequence alignment optimization— an improvement based on family representative mechanics features. *J Theor Biol*, 261: 593–597
- Liu X, Zhao Y P (2009b). Donut-shaped fingerprint in homologous polypeptide relationships—a topological feature related to pathogenic structural conversion of conformational disease. *J Theor Biol*, 258: 294–301
- Liu X, Zhao Y P (2010a). Generating artificial homologous proteins according to the representative family characteristic in molecular mechanics properties. *FEBS Lett*, 584: 1059–1065
- Liu X, Zhao Y P (2010b). Simulated pathogenic conformational switch regions matched well with the biochemical findings. *J Biomed Inform* (in press). doi: 10.1016/j.jbi.2009.12.003
- Liu X, Zhao Y P (2010c). Switch region for pathogenic structural change in conformational disease and its prediction. *Plos One*, 5(1): e8441
- López de la Paz M, Goldie K, Zurdo J, Lacroix E, Dobson C M, Hoenger A, Serrano L (2002). *De novo* designed peptide-based amyloid fibrils. *Proc Natl Acad Sci USA*, 99: 16052–16057
- López de la Paz M, Serrano L (2004). Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci USA*, 101: 87–92
- Makarov D E, Plaxco K W (2003). The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. *Protein Sci*, 12: 17–26
- Mastrangelo I, Ahmed M, Sato T, Liu W, Wang C, Hough P, Smith S O (2006). High-resolution atomic force microscopy of soluble Aβ42 oligomers. *J Mol Biol*, 358: 106–109
- Matouschek A, Kellis J T, Serrano L, Fersht A R (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, 342: 122–126
- Meiering E M (2008). The threat of instability: neurodegeneration predicted by protein destabilization and aggregation propensity. *Plos Biol*, 6: e193
- Nelson R, Sawaya M R, Balbirnie M, Madsen A Ø, Riek C, Grothe R, Eisenberg D (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435: 773–778
- Ólafsson Í, Grubb A (2000). Hereditary cystatin C amyloid angiopathy. *Amyloid*, 7: 70–79
- Pawar A P, Dubay K F, Zurdo J, Chiti F, Vendruscolo M, Dobson C M (2005). Prediction of “aggregation-prone” and “aggregation-susceptible” regions in proteins associated with neurodegenerative diseases. *J Mol Biol*, 350: 379–392
- Pepys M B (1995). *The Oxford textbook of medicine*. 3rd ed. Oxford: Oxford University Press, 1512–1524
- Polverino de Laureto P, Taddei N, Frare E, Capanni C, Costantini S, Zurdo J, Chiti F, Dobson C M, Fontana A (2003). Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J Mol Biol*, 334: 129–141
- Prusiner S B (1997). Prion diseases and the BSE crisis. *Science*, 278: 245–251
- Saiki M, Konakahara T, Morii H (2006). Interaction-based evaluation of the propensity for amyloid formation with cross-beta structure. *Biochem Biophys Res Commun*, 343: 1262–1271
- Sánchez de Groot N, Pallarés I, Avilés FX, Vendrell J, Ventura S (2005). Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct Biol*, 5: 18
- Sawaya M R, Sambashivan S, Nelson R, Ivanova M I, Sievers S A, Apostol M I, Thompson MJ, Balbirnie M, Wiltzius J J, McFarlane H T, Madsen A Ø, Riek C, Eisenberg D (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, 447 (7143): 453–457
- Smirnovas V, Kim J I, Lu X, Atarashi R, Caughey B, Surewicz W K (2009). Distinct structures of Scrapie Prion Protein (PrP^{Sc})-seeded *versus* spontaneous recombinant prion protein fibrils revealed by hydrogen/deuterium exchange. *J Biol Chem*, 284(36): 24233–24241
- Soto C (2001). Protein misfolding and disease; protein refolding and therapy. *FEBS Lett*, 498: 204–207
- Stefani M, Dobson C M (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med*, 81: 678–699
- Thomas P J, Qu B H, Pedersen P L (1995). Defective protein folding as a basis of human disease. *Trends Biochem Sci*, 20: 456–459
- Walsh D M, Klyubin I, Fadeeva J V, Cullen W K, Anwyl R, Wolfe M S, Rowan M J, Selkoe D J (2002). Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature*, 416: 535–539
- Wei G, Mousseau N, Derreumaux P (2007). Computational simulations of the early steps of pProtein aggregation. *Prion*, 1: 3–8
- Wolynes P G, Onuchic J N, Thirumalai D (1995). Navigating the folding routes. *Science*, 267: 1619–1623
- Vendruscolo M, Paci E, Dobson C M, Karplus M (2001). Three key residues form a critical contact network in a transition state for protein folding. *Nature*, 409: 641–645
- Ventura S, Zurdo J, Narayanan S, Parreño M, Mangues R, Reif B, Chiti F, Giannoni E, Dobson C M, Aviles F X, Serrano L (2004). Short

- amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc Natl Acad Sci USA*, 101: 7258–7263
- Yoon S, Welsh W J (2004). Detecting hidden sequence propensity for amyloid fibril formation. *Protein Sci*, 13: 2149–2160
- Zhang Z, Chen H, Lai L (2007). Identification of amyloid fibril-forming segments based on structure and residue-based statistical potential. *Bioinformatics*, 23: 2218–2225