Role of water in protein folding, oligomerization, amyloidosis and miniprotein

Tamás Vajda* and András Perczel

The essential involvement of water in most fundamental extra-cellular and intracellular processes of proteins is criticallyreviewed and evaluated in this article. The role of water in protein behavior displays structural ambivalence; it can protectthe disordered peptide-chain by hydration or helps the globular chain-folding, but promotes also the protein aggregation, as well (see: diseases). A variety of amyloid diseases begins as benign protein monomers but develops then into toxic amyloid aggregates of fibrils. Our incomplete knowledge of this process emphasizes the essential need to reveal the principles governing this oligomerization. To understand the biophysical basis of the simpler *in vitro* amyloid formation may help todecipher also the *in vivo* way. Nevertheless, to ignore the central role of the water's effect among these events means toreceive an uncompleted picture of the true phenomenon. Therefore this review represents a stopgap role, because most published studies—with a few exceptions—have neglected the crucial importance of water in protein research. Thefollowing questions are discussed from the water's viewpoint: 1. interactions between water and proteins, 2. protein hydration/dehydration, 3. folding of proteins and miniproteins, 4. peptide/protein oligomerization, and 5. amyloidosis.

Introduction

There is a growing demand to study protein interactions in aqueous solution because of their great importance of a biophysical, biological, and medical point of view. Maintenance of the proteome in a functional, non-pathogenic state (proteostasis) is a critical component of health [1,2]. Many diseases are associated with amyloidogenesis, a process where folded, globular proteins misfold and then misassemble and also unfolded, disordered proteins misassemble into soluble and insoluble toxic oligomer/polymer cross-β-sheet fibrillar aggregates, that is, amyloids [3–7]. Sporadic, genetic, and infectious diseases are of major medical concern, where the outbreak of the infectious bovine spongiform encephalopathy even more increased the general interest. The genetic human diseases are the results of malfunctioning of the cellular machinery [8,9]. On the other side, diseases are caused also by changes of the extracellular conformations coupled to the aggregation of misfolded proteins, without the influence of intracellular control systems. Today, about 30 different proteinmisfolding diseases are known, as for example, Alzheimer's disease (AD) and Parkinson's disease, Creutzfeldt–Jakob disease, type II diabetes, prion, and various other forms of amyloidosis [10–13]. Studies on some of these diseases have suggested that the prion-like behavior may be a general feature of misfolded proteins. Nowadays, researchers are able linking more and more amyloid-forming proteins to a given disease, such as tau-protein to AD (and related tauopathies), α-synuclein to Parkinson's disease, poly-glutamine to Huntington's diseases, prion protein to Creutzfeldt–Jakob's disease, and amylin to type 2 diabetes [14–19]. Several of these diseases originate from intrinsically disordered proteins (IDPs) such as Alzheimer's, Parkinson's (see later). There are also some findings of 'benign' amyloids, that is, proteins that function normally in their amyloid state, for example, some pituitary hormones. Thus, some functional amyloids in the pituitary and other organs can contribute to the normal cell and tissue physiology [20], these are the

exceptions. In general, different experimental and theoretical data have proven that the polypeptide-chain of proteins is of a β-pleated sheet structure in the neurodegenerative and other diseases [21–23].While the structures of amyloid fibrils are very similar, the polypeptides they are formed from highly dissimilar in their native states; that is, the formation of amyloid aggregates is independent of the original amino acid sequence of the given protein [24,25]. The sequence of a globular protein determines its 3D, biologically active structure [26], while the transformation of the globular form into an aggregated cross-β-sheet structure indicates the conversion of a physiologically functionaling native structure into a disfunctional, pathogenic, oligomer-polymer one [27]. Interactions of proteins with solvating water are fundamental to their structure, stability, dynamics, and function [28]. Proteins' surface hydration is essential for their 3D structure and activity, while in the absence of hydrating water, they lack their activity. Soluble proteins keep their backbone amide-carbonyl hydrogen bonds watertight to maintain their structural integrity, often protected them by 'wrapping' with surrounding nonpolar sidechain groups. On the other side, however, the loss of hydrogen-bonded partners implies a significant thermodynamic cost for a water molecule [29]. Proteins evolved in water for billions of years, and it is a ubiquitous solvent indispensable for life.

Perczel and his team [21] have shown by first principle calculations that the β-pleated sheet conformation of the peptide backbone is the thermodynamically most stable structure of all possible polypeptides both in vacuum and in aqueous environments, or even in a crystalline state, too. Therefore, from view point of thermodynamics the secondary structures of the polypeptide- chains can be considered as being of proteins in a metastable state of high Gibbs free energy. According to the kinetics, however, the α-helix, hairpins, γ-turns, etc., conformations are well existing entities. They are in local minima isolated by energy barriers, i.e., kinetically trapped, as has been shown by Gazit [30]. In the first part of this review, the following types of proteins will be addressed: (i) globular ones of ordered secondary-tertiary structures, (ii) recently discovered and identified type of proteins as having a novel structural motif named of 'charged single alpha-helix' (CSAH). In general, α-helices of naturally occurring proteins are not stable in solution without tertiary interactions. These CSAH type of proteins, however, having sequences of charged residues are stable monomeric molecules in water, and their protecting hydrate shell is 'kept strongly' on their surface, and (iii) proteins disordered fully or in some regions called as IDPs which exist and function without a given time average 3D structure in their disordered parts. The second part of this article contains the discussion of miniproteins, a peculiar type of oligopeptides with protein-like behavior. The smallest and most interesting one of all of them is a tryptophan cage (TC5b) miniprotein, with an oligopeptide of only 20 residues. (For both parts of this text, the references see at the related chapters.)

Protein Hydration

Contribution of Water to Protein Structure and Function

The aqueous structure (hydrate shell) around proteins is extending to around 5 Å from the surface (Figure 1), and the dynamic hydration layer extends to more than 10 Å [31] between the neighboring proteins. In solution, the proteins possess a conformational flexibility, which encompasses a broad range of hydration states. The structural and functional behaviors of proteins depend on the formation of a 2D hydrogen-bonded network spanning the protein surface and connecting the surface hydrogen-bonded water clusters. Such a water network can transmit information around the protein and also control the protein's dynamics. Water molecules may bridge the carbonyl oxygen

atoms and amide protons of different peptide bonds to help the formation of long-living linkages of protein–ligand and protein–protein interfaces. It has been suggested that pressure waves formed from flickering water clusters may link protein molecular vibrations, so carrying 'information' through the intracellular milieu [32,33]. Among the important roles of the water molecules, it should be concerned that the proton-transfer processes and the electrontransfer reactions may be facilitated by the ordered water molecules connecting the donor to the acceptor site.

Dynamics of Hydrated Proteins

Proteins are dynamic molecular systems that strongly interact with their aqueous environment; that is, protein motions are intimately linked to the motions of their environment, or in other words, protein folding is slaved to solvent motions [34–39]. The dynamics of water molecules within the hydration shell can have a crucial influence on the protein's biochemical and biological functions [32–40]. The proteins, even their important biological role, are dynamically passive, and all fluctuations and shape changes are controlled by the solvent. The bulk solvent fluctuations control the shape changes and global motions of the proteins, while the hydration shell fluctuations are related to internal motions. Proteins exhibit a very large number of confomational substates, where the dominant conformational motions are often related to the hydration shell as well as to the bulk solvent. The hydration shell of proteins consists of ≈2 layers of water as shown in Figure 1. Protein functions depend on the hydration degree, h, which sign symbolizes the weight ratio of water to protein and h>1 is required to the full function. Because the interior of proteins is fluid-like, their intrinsic viscosity is small, similar to water [41].

Relaxation Dynamics of Hydrated Proteins

Khodadadi and her co-workers have studied picosecond to nanosecond dynamics of hydrated protein powders by dielectric spectroscopy data and molecular dynamics (MD) simulations, completed with the data of neutron scattering [42–45]. This dielectric method is based on the measurement of the change of the 'relative permittivity' at a given frequency range. These spectroscopic experiments were carried out with hydrated lysozyme powders at different hydration levels, where the hydration levels were determined by thermogravimetric analysis. The dielectric relaxation spectra in the frequency range from 10^{-2} to 10^7 Hz were measured. On the other hand, the MD simulations were performed with RNase protein. The dielectric spectra of hydrated proteins contain two components: a fast (20–50 ps) relaxation related to the bulk water dynamics and a slower one (0.5–10 ns) ascribed traditionally to hydration water tightly bound to the protein. However, Khodadadi et al. argued the latter interpretation, by saying that the protein hydration water exchanges with the bulk water on a time scale faster than 100 ps. The experimental and simulation data suggest that the nanosecond relaxation is due to motions of the protein atoms themselves, including backbone, side chains, turns, surface and the core part of the protein. Nevertheless, it needs further studies to reveal the details of contribution of segments and secondary structures of the protein motions at the ns time scale of motion.

Protein Dynamics Studied by NMR. Excellent reviews appeared of protein dynamics as reported by NMR-techniques [46–48]

Dynamics Detected in a Reverse Micelle

Wand and his colleagues [49] carried out NMR measurements on water–protein interaction with protein encapsulated in an AOT [(sodium bis(2-ethylhexyl)sulfosuccinate] reverse micelle. They determined that confining a protein within a reverse micelle of nano-scale interior slows down the water dynamics. Complications that normally arise from hydrogen exchange and long range dipolar coupling are overcome here by the nature of the reverse micelle medium. Characterization of the hydration of human Ubiquitin demonstrates that encapsulation within a reverse micelle allows detection of several hydration waters. As it is well known, the biological cells are crowded with macromolecular surfaces that confine proteins to nanometer-scale volumes, and this can result in a potentially highly heterogeneous intracellular hydration environment [50,51]. A schematic model of hydrated Ubiquitin is shown in Figure 2.

Relaxation Studies on the Hydrate Layer of Intrinsically Unstructured/Disordered Proteins

Hydration of IDPs has been investigated by Tompa P & K and their colleagues with NMR technique [52]. It was demonstrated that the hydration of IDPs can be detected by solid-state NMR relaxation measurements [53]. To study the hydration of these proteins and to separate the different water phases of the solutions, samples are frozen (T ~ 200 K or below) to separate the free induction decays (FIDs) of the hydrate layer from those of the overwhelming bulk water by freezing the non-hydrate or bulk water out. On this manner, it was possible to measure the FIDs pertaining to protons of bound water, from which, it could be calculated the amount of hydrate shell, its activation energy, and correlation times. As the primary sequences of IDPs contain a large number of charged amino acid residues, they bind a significantly larger amount of water than globular proteins do. Furthermore, the activation energy of dynamics of the most strongly bound part of the hydration shell is about 50% more for IDPs when compared with globular proteins [54].

Protein Hydration/Dehydration

Folding of Globular Proteins

An important review of particular interest on translational water entropy on self-assembly processes like protein folding was published by Kinoshita [55]. Another review of special interest deals with the progress of folding studies in the last half-century [56]. Folding of globular proteins is one of the most fundamental and universal example of biological self-assembly. Native globular proteins fold and unfold continuously, where the distribution and strength of contacts are influencing their folding kinetics. The folding time scale varies from microseconds to seconds. It is an astonishing surprise and questions the fact that in spite of the enormous large number of the conformational possibilities, the given protein molecule can fold to its only precisely defined native structure in a really short time [56]. This question has been first time posed by Levinthal already in 1968 and cited nowadays very often as Levinthal's paradox [56,57]. The solution of this problem is in the funnel-shaped energy landscape of proteins, with several unfolded structures of high energy and only a few low-energy folded structures. Folding occurs energetically downhill via alternative microscopic trajectories, mainly toward the deepest valley. Nevertheless, some molecules need more time to reach the native structure, as arriving to a higher kinetically trapped local minimum. Although the motions of folded proteins follow the solvent fluctuations, it can be reduced significantly. Slowing may occur because a broad range of motions proceeds in several small steps. Considering the transient unfolded state, the unfolded protein makes a Brownian motion [58] in the conformational folding funnel. Whereas the activation enthalpy of folding is mainly determined by the solvent, the number of steps is controlled

by the number of substates of the unfolded protein and interacting solvent with. In this folding process, the principal driving force is the entropic effect arising from the translational movement of water molecules, which means a large gain in water entropy. Beside the protein folding, the role of water is important also in other self-assembling and ordering processes in biological systems, such as molecular recognition and ordered aggregation. The great entropic loss for the biomolecules accompanying these processes is largely compensated by the gain of the water's entropy. The other side of the coin is the role of the hydrophobic interactions, the unfavorable forming a surface area of non-polar groups with water, which results in a surface minimization together with the release of water and by this manner maximizing the water's entropy. An interesting review on the translational motion of water molecules in sustaining life was written also by Kinoshita [59].

Charged Single Alpha-helix

This highly charged novel structural motif was recently discovered and identified in proteins [60,61]. In naturally occurring proteins, α-helices are rarely appearing as independent structural elements stable in aqueous solution without of tertiary interactions. Nevertheless, there are some few but important cases existing, for example, Caldesmon [62], Myosin VI, and Myosin X [63,64]. These CSAH types of proteins are stable monomeric molecules in solution as primary sequence comprises charged residues keeping strongly their hydrate shells. Although it was proposed that these α-helix conformers are stabilized by electrostatic interaction, we emphasize also the role of hydration that the pattern of the hydrated charged residues confers stability on α-helices. Spudich and his team [65] found that amino acid sequences of proteins containing alternating repeats of negatively charged glutamic acid (E) residues and positively charged combinations of arginine (R) and lysine (K) form stable α-helices. Gáspári et al. published [66] a consensus prediction method suitable for predicting CSAHs from primary sequences. By using this consensus method, they found that although the CSAH motif is rare in general, it is quite abundant in proteins of symbiosis and RNA binding/processing. Nyitray, Perczel et al. arrived to an important and interesting conclusion, concerning the not deep conservation pattern of CSAH-containing proteins, that they might take part in relatively rapid molecular evolution, and therefore, they may contribute to the emergence of novel functions [67,68].

Intrinsically Disordered Proteins

More and more IDPs or at least large segments of them become known, which lack a well-structured 3D fold [69]. The occurrence of disordered regions of significant size (>50 residues) is common also in functional proteins. It points to the fast growing field of these IDPs, which several reviews appeared recently, including also the Tompa's textbook [70]. The intrinsic disorder means the low content of bulky hydrophobic amino acids and a high one of polar and charged amino acids. Although there are different experimental techniques to study disordered proteins (CD, VCD, fluorescence and Raman spectroscopy, and also hydrodynamic measurements), the most important method for obtaining systematic information on unstructured proteins is the NMR spectroscopy. In spite of these sophisticated methods, *in silico* predictions still have an important role in studies of disorders. Many of the disordered segments fold partly on binding to their targets ('coupled folding'), while other types constitute flexible linkers for the assembly of macromolecular arrays [71–76]. The disordered segments never become fully ordered in the bound state, and often some remaining disordered regions can take part in function and/or recognition. The extension of this partial disorder of the

bound state is termed 'fuzziness', and this phenomenon may be general in function [77]. Zagrovic and his group published that the ID regions of proteins can lower the hydration free energy for the whole protein [78]. They studied with MD simulations the dimeric enzyme Nudix hydrolase, a member of the proteome of the radiation—and desiccation—resistant bacterium D. radiodurans. The hydration free energy of the Nudix hydrolase dimer together with the disordered tails is significantly more negative than that of the globular proteins of equivalent size, which means an average difference of more than/about 1200 kcal/mol. In general, the addition of extra hydrophilic amino acids would lower the hydration free energy of any protein, and therefore, the hydration shell of the protein will be held more strongly. Zagrovic et al. hypothesize that the disordered tails increase the possibility of the protein to be located in the water patches of the desiccated cell, and by this protection, it can function normally. Beside Zagrovic's team, other reseach groups also studied the Nudix hydrolase family [79,80].

Thermodynamic Calculations of Peptide/Protein Aggregation

Perczel et al. [21,27,81] carried out thermodynamic quantum level calculations for the reasons of oligomerization/aggregation: amyloidosis. In their work, the following two questions were investigated: (i) why is the thermodynamically most stable dimer structure preferred as the 'double stranded antiparallel β-pleated sheet' (DSAS) conformation and (ii) what is the role of the entropy in peptide aggregation? Concerning on the first issue, it has been shown that the β-pleated sheet conformation of the peptide backbone is the thermodynamically most stable dimer structure. These calculations are based on the For-(L-Ala)$_4$-NH$_2$ molecule hydrogen bonded dimerization with a yield of 31.13 kcal/mol, while the value of the parallel β-pleated sheet structure (DSPS) is a slightly less of this. Nevertheless, it should be kept in mind that these results refer to model peptides. These relative energy calculations of folding or oligomerization/aggregation started from the isolated extended single stranded β-sheet conformation (Extβ), with respect to the 'dead-end street' [21]. Beside the thermodynamics, however, all conformations of folding are in a stable state, according to the kinetics, as discussed before. Furthermore, these secondary structures are inhabitants of the hydrophobic core of the globular proteins and stabilized also by hydrophobic interactions. The calculations were carried out for aqueous environment and with simulation of physiological conditions. In the second question, the calculations concentrated on the thermodynamic parameters, that is, on the relative enthalpy that contains the relative Gibbs free enthalpy (ΔG), the measure of driving force of peptide dimerization and the relative entropy (ΔS) related to the extent of order or disorder [82–84]. These calculations were carried out in the exciting hope to reveal the entropy changes upon dimerization, which data may help to get closer to predict the possible way that can turn the dimerization– aggregation into the dissociation of plaques. In accordance with the previous results, these calculations have proven also thatthe dimerization and oligomerization processes are thermodynamically favored, where the antiparallel β-pleated sheet conformation is more stable than the parallel one. The antiparallel arrangement has staggered conformation shown along the longer peptide chains. The results indicate also, that for longer peptides, the relative changes of the thermodynamic functions are constant. The parameters normalized per hydrogen bonds (n) point to a slight difference between the data yielded in vacuum versus in water; (antiparallel) ΔG kcal mol^{-1} =4.34 and 2.20, which represents a slight aqueous destabilization effect on the sheets, or from the opposite viewpoint, it demonstrates a slight effect in the direction of dissociation of the sheets. The difference between the thermodynamic parameters of the antiparallel and parallel conformations are small, as the difference is of ΔS cal K^{-1} mol^{-1} =0.73 andTΔS kcal mol^{-1} = +0.22. (However, the

results of experimental studies of amyloid plaques, like Aβ 1–42, show the opposite data; namely, these aggregates contain parallel β-pleated sheets [83].) From these quite small differences of parameters, Perczel and his colleagues arrived at an important conclusion that only a minor change in entropy could direct to the dissociation of plaques. It can be seen on Figure 3 that the stability increases monotonically with the increase of the polypeptide chain length and with the number of β-strands, as well. The schematic diagram of the same figure shows the more stable S14 and the less stable S10 hydrogen-bonded antiparallel pseudo-rings, where each S14 subunit adds to the system a stability of ≈10 kcal mol^{-1} [21], while a single S12 unit stabilizes a parallel β-layer by ≈7 kcal mol^{-1}. (S14, S10, and S12 denote the members' number of the pseudo-rings.

Influencing Factors and Pathways

The aggregation depends on intrinsic and extrinsic factors, as well. Intrinsic ones mean several characteristics of the polypeptide chain, such as charged, polar, and/or apolar residues, the propensity to adopt secondary, tertiary, or quaternary structure elements. The extrinsic factors signify physico-chemical properties, that is, the environmental conditions in which the polypeptide chain is present. Such factors are hydration degree, protein concentration, pH, temperature, ionic strength and species, denaturants, and so on [85–89]. It is important to consider that the environmental conditions ultimately affect the kinetics of nucleation of amyloid fibrils or the denaturation of the protein in extreme conditions. It is therefore crucial to investigate the factors that modulate amyloid formation/denaturation of the given protein.

The Water's Paradox: Aggregation or Folding?

Globular proteins aggregate on different pathways: (i) through partially or fully unfolded intermediates, (ii) through self-association, (iii) through chemical linkages, and (iv) by chemical degradation. Native folded, dissolved proteins are in equilibrium with a small amount of unfolded intermediates or with some already completely unfolded molecules. In concentrated aqueous solution, the protein conformation often turns to phase separation and then to precipitation. Considering the free energy landscapeof aggregation, its driving force is a favorable free energy yield because of reduction of the solvent-accessible area (surface?) In other words, it means an entropy increase of the water molecules tightly bounded in the proteins' hydration shells, as escaping into the bulk aqueous part [55]. It depends, however, that the loss of the binding enthalpy of the water molecules could be paid by the entropy increase (see also the indent in the section on Folding of Globular Proteins). The second station of the folded proteins via the most stable β-pleated sheet formation to the aggregation/ amyloidosis means their unfolded/disordered state. However, these partly or fully intrinsically disordered regions increase the 'surface-properties'; that is, the whole protein will be more hydrophilic, which suggests a strongly bounded hydration shell and the protection against the aggregation becomes increased. On the other side, if the hydrophobic groups of the molecule are exposed to the water, then paradoxically, kinetic competition will appear between the aggregation and folding of the proteins [55,84,85]. It seems worthwhile therefore to consider an example of the energy landscape of amyloid formation, which has been carried out by isothermal titration calorimetry at 25 °C [90] (Table 1).

Calculation of Aggregation-prone Regions

Dobson and Vendruscolo et al. [91,92] elaborated a method for calculating the intrinsic amyloid aggregation-prone regions of the polypeptide sequences of the unfolded or partially unfolded proteins. This computational method is based on the aggregation propensities of individual naturally occurring amino acids, where the aromatic residues may play an important role in aggregation, while the polar and charged, much stronger hydrated residues are considered as anti-aggregant factors. Some of the aggregation-prone residues are for example, Cys, Phe, Trp, Tyr, and the anti-aggregant ones: Asn, Asp, Lys, Glu, and Arg. This approach has been used to investigate natively disordered polypeptide domains associated with neurodegenerative diseases, Aβ42, α-synuclein, and tau-protein. The authors differentiated two types of sensitive regions, where large changes in the propensities for aggregation can be predicted as a result of mutations. These aggregation-susceptible (S) regions are according to their response to single amino acid mutations: regions of low propensity that can become of high propensity (type S+) and regions of high propensity that can become of low propensity (type S-). Regions of type S- are suitable targets for strategies to find amino acid replacements to combat diseases associated with amyloid formation. There are prominent examples of type S; the residues 30–42 of Aβ42, NAC (non-amyloid-beta peptide component) region of α-synuclein and 306–311 regions of the tau-protein.

Protein Dehydration

Because the expulsion of interfacial water molecules of proteins is the initiator of oligomerization / amyloidosis, it seems interesting to investigate the correlation between the water content and oligomerization's propensity of a given peptide [93,94]. Gai and his colleagues studied the dependence of the aggregation kinetics of peptides on their hydration degree [95]. The results indicated that the aggregation rates increased significantly with decreasing number of water molecules: w_o= 20, 10, and 6. These experiments have proven that the dehydration promotes aggregation by reducing the free energy barrier between the protein molecules by decreasing the water shell protection. The sequences of the two studied peptides were: KLVFFAE and GNNQQNY, which despite their short chains have been shown to exhibit strong amyloidogenic propensity [95]. The Gai's team hypothesizes that their results have also implication for the aggregation of in vivo (in cell), where crowding can be expected to reduce the local density of water.

Amyloid Fibrils *In Vitro*

Formation of amyloids appears to be structures that were not selected for in molecular evolution. But in contrast with the traditional conception, recent studies have shown that the ability to form amyloid fibrils is not only given by some specific sequences of proteins, but it can be considered as a generic feature of polypeptide chains, short and long ones, as well. Fibrils can be formed also by not disease associated proteins, for example, myoglobin or by homopolymers such as poly-lysine and other poly-amino acids. Amyloid fibrils are highly organized structures of an unfolded polypeptide chain behaving as a typical polymer, where the essential features are determined by the physicochemical properties of this polymer chain. The proportion of the 'core structure' of a polypeptide chain can vary substantially. 'Core structure' means the sticky part of the polypeptide sequence responsible for the aggregation. Nevertheless, it is crucial to emphasize again the role of water in the driving force for addition of a solvated monomer to a preformed fibril [55,96]. These 'amyloidogenic' segments, according to Eisenberg's team [97], have a self-complementary 'steric-

zipper' structure that lets them mesh very tightly with an identical segment located on another protein. An example of this 'steric zipper' can be seen in Figure 5 [98].

This generic structure type contrasts with the globular structures of natural proteins. It should be considered also that even a single amino acid change in the protein sequence could change dramatically the hydration of the polypeptide and, by this manner also, the aggregation rate of the unfolded polypeptide chain. Experiments have shown that the fibril formation has generally startedby a lag phase and followed then by a rapid growth, which behavior is typical of a nucleated process like crystallization. The transition of amyloid-forming peptides/proteins into well-ordered fibers and then to crystals means an important issue. For this purpose, a seven-residue peptide, GNNQQNY, the fragment of N-terminus of the yeast prion-like protein Sup35 was studied (as shown already in the preceding text) [95]. The goal of this research was to find out the structural changes accompanying fiber to crystal switch. The structures of ≤7 amino acids short peptides that form amyloid fibers and microcrystals share also a 'steric zipper' structure arrangement [97,98]. The GNNQQNY peptide can form fibrils and microcrystals, as well, depending on the concentration and incubation time [99]. It has been found in all tested concentrations that the crystal's formation was favored with the elapsed time which points to the stability of crystals. Recent research has revealed that the monomer to oligomer transition of proteins is the pathogenic transformation; these oligomers are the most toxic entities, which means also that the monomer to oligomer transition is not only the first step of aggregation but also the transformation of a benign protein to a toxic one. For example, the β-amyloid peptide (Aβ) monomer shows no neurotoxicity contrary to its oligomer. However, the monomeric state of this aggregation- prone peptide remained yet beyond of most experimental techniques, because the oligomers are thermodynamically unstable. The oligomer–monomer dissociation is strongly impeded by a large kinetic barrier of mostly entropic of origin, as it was found by Maiti's team [100]. Concerning of Aβ40 and Aβ42, both remain predominantly monomeric up to 3 μM at equilibrium, and above it forms large aggregates. This in vitro value is a much higher concentration than the estimated one in the cerebrospinal fluid of the normal or diseased brain (<1 μM) [100]. Eisenberg and his group found an amyloid-forming protein (αB crystallin; ABC) [101], which produces fibers more slowly than the most ones, for example, the Aβ. Therefore, in this case, the oligomeric state could be trapped before the fibrillation, and this oligomer has been recognized by an oligomer-specific 'conformational' antibody (A11) that binds only oligomers but not fibrils. This oligomeric complex shows similar properties to other amyloid oligomers: β-sheet rich structure and cytotoxicity. The 3D X-ray solved atomic structure of the oligomer pointed to a cylindrical barrel of six antiparallel chains, which received the 'cylindrin' name from the discoverers. Two segments of the cylindrin exhibited high amyloidogenic propensity, with KVKVLG and GDVIEV sequences. Together, the entire segment forms a hairpin loop in the ABC's 3D structure, where this segment may be further stabilized by strong hydration because of its polar residues (K, K, D, E). A sequence of cylindrin is similar with a sequence segment of the β-amyloid protein of AD.

Metal Ions and Proteins/Amyloids: Cu^{2+} and/or Zn^{2+} Binding

It is well known that several proteins and amyloid fibrils are inclinable to metal ion binding. These mean also that the metal ions play a serious role in the onset and development of neurodegenerative disorders [102,103]. However, a metal ion binding on protein represents also the protection of the given protein. From the viewpoint of energy landscape, two questions should be addressed: (i) how

much activation energy the strongly hydrated metal ion needs for binding, i.e., how much deficit means in free energy the partial loss of the metal ion's hydrate shell; (ii) could the metal ion's binding enthalpy and the water's translational entropy for the mentioned activation energy pay? Metal-binding proteins play several important roles, as maintaining the structure, effect of catalysis, recognition some substrate, signal transduction. In order to hold a given metal ion with high specificity and affinity, the proteins form multidentate binding pockets to satisfy the geometric and chemical bonding requirements of that metal [102–105]. The metal ion binds in a hydrophilic shell of the protein, and this shell is embedded within a larger hydrophobic shell. This paradox of metal–ligand interaction has been interpreted with different viewpoints [102]. The answer of a protein on the metal ion binding depends on the physico-chemical environment and also on the given salt ion. Concerning some of the most important cations, interesting results are described for the binding of the Cu^{2+} and Zn^{2+} ions. Moroder and his colleagues investigated the copper(II) binding on the prion protein inmembrane mimetic conditions, while Sóvágó et a. studied the interaction of the same ion with a prion peptide fragment [106,107]. Furthermore, Sóvágó's team studied also the competition between Zn^{2+}/Cu^{2+} ions for the available binding sites of an Aβ(1-16) polyethlenglycolyted peptide [108].

Proteins in Cell

The discussed considerations are valid mainly for in vitro condition; however, what is the proteins' situation in vivo (in cell)? The interior of a cell is a complex environment in which proteins together with other macromolecules are present at an extremely high concentration range of about 300–400mgml1 [109,110]. Because of this extreme crowding, there can be a special hydration; that is, it can be expected to reduce the local number of water molecules. It is known that within the cells, there are large numbers of auxiliary factors that protect the proteins against oligomerization–aggregation in this complex, concentrated and crowded milieu, and these factors including catalysts and molecular chaperones assist also in the folding process. The crowded environment can even accelerate the folding process, because a folded protein needs less volume than an unfolded one. Nevertheless, the macromolecular crowding increases the chances of proteins' aggregation if the intrinsic folding rate of the given protein is too slow [111].

Chaperone and Chaperonin Proteins

Chaperones are predominantly disordered proteins that assist the folding of proteins and prevent the oligomerization–aggregation both of the newly synthesized polypeptide chains and assembled subunits into nonfunctional/possibly toxic structures [112–116]. This is valid for chaperone molecules working intracellular or extra-cellular, as well. Nevertheless, it is surprising that this type of molecules without a 3D structure in their functional regions can help proteins' folding problems and prevent their aggregation. Chaperonins are a subgroup of chaperones that encapsulate their folding substrates. The molecular chaperonin's role in the folding acceleration correlates with the possibility that water can become confined inside the chaperonin, for example, in GroEL a 60-kDa chaperonin [117]. It might be expected therefore that this confined water behaves differently than it does in bulk. Pande's team concluded from their results [118] that accelerated folding occurs because the polar residues on the inner surface of the chaperonin accumulate the water in their vicinity and providing so a stronger drive for the decrease of substrate surface hydrophobicity, for protein folding. Folding of a polypeptide chain can be hampered by kinetic traps to halt the process for a

physiologically significant time. In this case, the chaperone with its disordered regions loosens the structure of the trapped folding intermediate via transient binding and unfolding to offer another chance for a folding of the substrate protein [112]. Considering the inhibition of aggregation, on the other hand, the transient-formed protein chaperon complex destroys the nuclei of aggregates.

Concerning the Water's Role in Functions of the Chaperone and Chaperonin

The protein-bound chaperone efficiently solubilizes its protein substrate because of the hydrate shell dominance of the chaperone's disordered regions. This means that in the free energy landscape, the gain of the solvation's binding enthalpy of the protein-chaperon complex can pay for the loosening of the trapped intermediate and also for the water's entropy decrease by the protein hydration. The destroying effect of the aggregates' nuclei may also be related to these nuclei dissolution [112,114,118] (consider also the sections of Protein Hydration/Dehydration and The Water's Paradox: Aggregation or Folding?).

Amyloids *In Vivo*

Although the amyloid precursor peptide (APP) is degraded by α-secretase, β-secretase, and γ-secretase, an imbalance between production and clearance of Aβ fragments leads to accumulation of Aβ peptide monomers, oligomers, and at last to the diseases [119,120]. Verdier and Penke studied the binding of β-amyloid (Aβ) peptides to the cell plasma membranes in AD, as regards the toxicity on neurons. They discuss the membrane proteins that can mediate the interaction between the β-amyloid (Aβ) peptides and the cell plasma membranes [121]. Aβ peptides can bind a variety of biomolecules such as lipids, proteoglycans, and also proteins. According to the suggestion of Verdier and Penke, the binding of β-amyloid peptides to plasma membranes could be a promising possibility for intervention in the events leading to the development of AD.

Miniproteins: The Trp-cage Motif (TC5b)

Oligopeptides of <40 amino acid residues and including a Pro-Trp sequence are a priori considered as miniproteins, the truncated derivatives of exendin-4 (Ex-4), an intestinal hormone, and natural drug. The artificial construct of TC5b miniprotein actually can be considered to a molecule of natural origin, as being a truncated and mutated derivative of the 39-residue long Ex-4 peptide. This 20-residue construct 'Trp-cage' motif is an excellent finding of Andersen's team [122], where the 20-residue peptide represents the smallest folded tertiary structure as known to date. It is >95% time average folded in water at280 K and in physiological pH, that is, it is the smallest fold that can be viewed as a globular protein. This protein-like 3D-fold molecule affords prominent possibilities for experimental studies as well as for MD simulations of protein unfolding/folding [123]. An MD simulation article appeared as showing the successful possibility of stepwise elongation of TC5b's peptide chain with locally driven folding [124]. Accordingly, the interactions of the successively evolving Trp-cage were Trp^6-Pro^{12}, then Trp^6-Pro^{18}, and finally Trp^6-Tyr^3.

The TC5b molecule has the following sequence: N L^2YIQWLK^8 DGG P^{12}SS14 G R^{16}PPP^{19} S (see Table 1 of [122]). The numbered sequence parts represent the given conformations: the L^2–K^8 range means an α-helical segment, while the P12–S14 one displays a short 310-helix, and the C-terminal R^{16}–P^{19} part indicates a PPII helix structural motif, together with a hydrophobic stacking between the aromatic rings of Y^3 and W^6. All these structural segments provide a watertight seal for the

tryptophan residue. The central W^6 and Y^3, L^7, G^{11}, P^{12}, and P^{18} amino acid residues form the hydrophobic core. It is noteworthy also a natural example: the avian pancreatic polypeptide of 36-residue, which contains a PPII-helix followed by an α-helix and an aromatic-pro interaction [125].

TC5b Cooperativity and Stability

The guarantee for stability of this miniprotein roots in its hydrophobic cooperativity around the central Trp6 and in a salt-bridge between Asp^9 and Arg^{16}. These interactions have been studied in detail [126,127]. It was an interesting observation that the Arg16 residue stabilizes more the miniprotein molecule than the Asp^9 does it, according to the NMR measurements at pH values of 3.2 and 6.9. This effect may be caused by Arg's hydrated positive charge kept at both pH values, together with the interaction between the arginine side chain and the hydrophobic core of the miniprotein [127]. Among the synthesized salt-bridge variants, the TC5b-D9E derivative showed the best properties. The one methylene group longer side chain of the Glu^9 residue results more efficient contacts in the hydrophobic parts and by this manner a more compact 3D-structure, where the PPII segment shields the indole ring of the Trp better [127]. However, these conclusions are argued by Andersen's group; they suggested that the D9E substitution means an intrinsic helicity increase and not an improvement in salt-bridging [128].

Crucial Reveal of the Trp-cage Folding/Unfolding

The Unfolding Route

Perczel's group elucidated the multistate folding scenario, by temperature-induced unfolding procedure. It has been shown by electronic circular dichroism technique that beside the two states, there appears at acidic pH and at 45–55 °C a molten globule-like or intermediate state, too [127]. In their second approach, they used NMR techniques for ^{15}N-labeled and $^{13}C/^{15}N$-labeled proteins, at neutral and acidic pH values and at different temperatures [129]. They succeeded also in developing a deconvolution technique, which made possible to 'see' the invisible fast exchanging states. Because of these fine-tuned methods, it was possible to detect and discover the intermediates of the multistate unfolding process of this miniprotein, where at neutral conditions, a fast-exchanging intermediate, while at acidic pH, a slow-exchanging intermediate pair could be determined. The fast-exchanging intermediate of a native-like structure contains a short α-helix in its G11–G15 segment. Beside the detection of the structural characteristics at atomic detail of these three-step processes, the folded, intermediate, and unfolded states, also the thermodynamic parameters could be obtained by nonlinear fitting methods.

The Folding Route

Kuhn and his team investigated with residue-specific, or relaxation NMR techniques the atomic-level structural forms of the TC5b miniprotein (Figure 6) denatured in 6M urea [130,131], and 13C-labeled and 15 N-labeled peptides were used for this purpose. It has been shown by another team that the Trp-cage molecule is an ultra-fast folding protein: 4 μs folding time (laser temperature-jump spectroscopy) [132]. Although the 6-M urea-unfolded TC5b lacks any regular secondary structure, it is a dynamic ensemble of species that contains a distinct cluster of locally interacting residues, comprising both aromatic and aliphatic side chains, and it corresponds to the native miniprotein's region that forms an α-helix. This hydrophobic clustering in the denatured state may be considered

as a 'preparation' for the ultra-rapid folding to TC5b, where it means a pre-existing native and/or non-native interaction [130].

The Role of Water in Miniproteins

The energy landscapes of the F/U routes of miniproteins are similar to those of the native globular proteins, although they bear some smaller sized energy funnels. Nevertheless, the folding kinetics depends also here on the distribution and strength of contacts, on the enthalpy gain together with the water's translational entropy gain, as paying for the 'folding order'. Namely, the hydrophobic amino acid residues inside and the hydrophilic ones outside guarantee the strong folding of the peptide chain. The extra fast folding speed points to the fast leaving of the water molecules from the unfavorable hydrophobic environment inside and the tightly bounded water molecules in the proteins' hydration shell outside. This truly cooperative folding behavior of the TC5b miniprotein is exhibited by the ultra-fast folding rate and the smallest tertiary structure of >95% time average folded in water at 280K.

Aggregation Propensity of Trp-cage Miniproteins

Because the TC5b and its derivatives are excellent models for F/U studies in vitro and in silico conditions, it is an important question to investigate the aggregation propensity of these molecules, becuase aggregation is associated with misfolded diseases. The aggregation propensity versus stability of a folded globular protein is in inverse relationship, and this suggests the idea to explore the behavior of Trp-cage proteins of different stability, that is, of different tightness of their cores [133]. For this purpose, the differently truncated derivatives of EX-4 were prepared with some mutations of the residues. E0 (20 aa), E5 (25 aa), and E 10 (30 aa), where the +5/+10 residues mean the α-helix forming (stabilizing) chain (Table 2). The applied techniques were FT-IR; electronic circular dichroism, and VCD spectroscopies. The experimental conditions were T = 5–65 °C; pH = 5–7; c ~ 30 μM, which were used for all three miniproteins. At this low concentration, not any of the E-constructs indicated β-structures. Nevertheless, E5 at c ~ 1–3mM and E10 at c ~ 30mM concentration demonstrated the α→native-β→β-sheet folding transitions.

It can be concluded that the E0 has a very dynamic fold, the least tight core conformation among the three foldamers, while the E5 polypeptide has been proven intermediate in stability. This means also that the E5 molecule is the most vulnerable entity against aggregation. The E10 with the longest α-helix chain segment, is the most stable Trp-cage fold with the tightest core. The water's role also supports the aforementioned conclusion, as concerning the relation of the increasing percentage of hydrophylicity to the tightness and stability of the miniproteins.

Conclusions

'The water molecules are the main actors and the proteins are only the figurants on the stage of Life'. This sentence points to the involvement of water in most fundamental extra-cellular and intracellular processes of proteins. Considering the conformation of a given protein, or the folding of a polypeptide chain into a globular protein molecule, the protection of a charged protein by hydrate shell, and aggregation of proteins, all of these phenomena are based on the interactions of water. Concerning the unstructured/disordered proteins or protein regions, some of them can be considered also as 'hydrators', that is, the ID regions increase the probability to remain solvated the

protein in case of dehydration. The Trp-cage miniproteins, like the TC5b of a 20 residue long polypeptide, and its longer analogs are excellent foldamer models, because their simplicity and as they are influenced by less factors like the large native proteins. They are good targets also for the exploration of the aggregation propensity.

Tables and Figures:

kcal/mol					
ΔH		ΔG	ΔG$_{app}$	-TΔS	
Native	Amyloid	Native	Amyloid	Native	Amyloid
25.6±1.2	13.8±0.7	5.7±0.2	10.0±0.04	-19.9±0.9	-3.8±0.5

Table 1. The unfolding thermodynamic parameters of amyloid formation by isothermal titration calorimetry[a]
([a]The native monomeric β$_2$-microglobulin (β2m) was compared with the amyloid form, where the β2m means a protein responsible for dialysisrelatedamyloidosis.)

P/A[a]	%	Name	Amino Acid Sequence
			LSKQMEEE AV RLFIEWLKNG GPSSGAPPPS-NH$_2$[b]
5/15	33	Tc5b	NLYIQWLKDG GPSSGRPPPS
5/15	33	E0	RLYIQWLKEG GPSSGRPPPS
8/17	47	E5	EEEAV RLYIQWLKEG GPSSGRPPPS
10/20	50	E10	LSKQMEEEAV RLYIQWLKEG GPSSGRPPPS

Table 2. Relative polarity differences between amino acids in miniproteins of different lengths
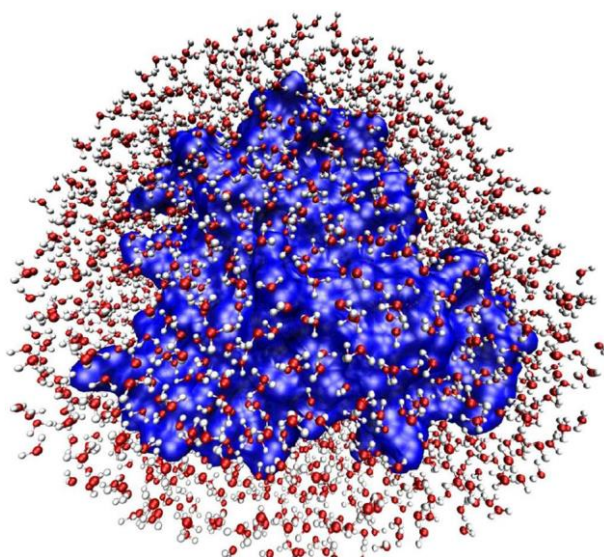

Figure 1. The hydration shell of myoglobin (Mb). Diagram of myoglobin (blue surface) with 1911 water molecules (CPK model), the approximate number needed for optimal function (h = 2). The waters form a shell ≈5Å thick around the protein. Approximately 200 water molecules are distinguishable from background with high-resolution X-ray crystallography. Reprinted from Frauenfelder et al. [34].
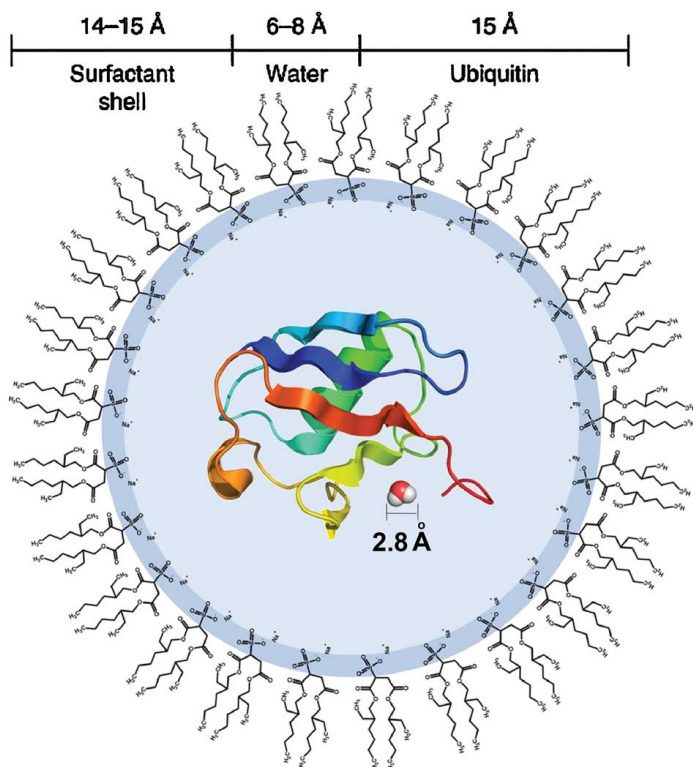
Figure 2. Schematic of Ubiquitin in an AOT reverse micelle. The surfactant, water pool, and protein are shown to scale, with a single water molecule pictured, also to scale, for comparison. Alkane solvent, which surrounds the reverse micelle particle, is not shown. Reprinted from Wand et al. [49] with permission from the publisher.
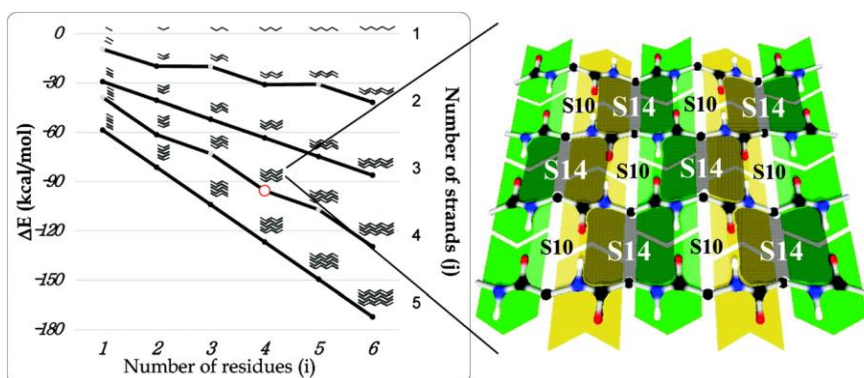

Figure 3. Total electronic stabilization effect (ΔE, kcal mol^{-1}) of the interchain hydrogen-bond network. For multiple stranded β-layers formed by {For- (L-Ala)$_i$-NH2}$_j$ at j = 2, 3, 4, the stability increases with the increase of the length of the polypeptide chain (i=1, 2, 3, 4, 5, 6). All ΔEs are with respect to the isolated and independent Extβ conformers of {For-(L-Ala)i-NH2}j where j=1 and $1 \leq i \leq 6$. The schematic diagram of a tetrapeptide tetramer arranged as a four-stranded antiparallel β-layer (i=4, j = 4) is enlarged. The most stable S14 pseudo-rings are shaded, and the less stable S10 subunits are brighter. The S14 and S10 symbols denote pseudo-rings of 14-membered and 10-membered, respectively. On Figure 4, strands from parallel β-pleated sheet superstructures are shown [84].
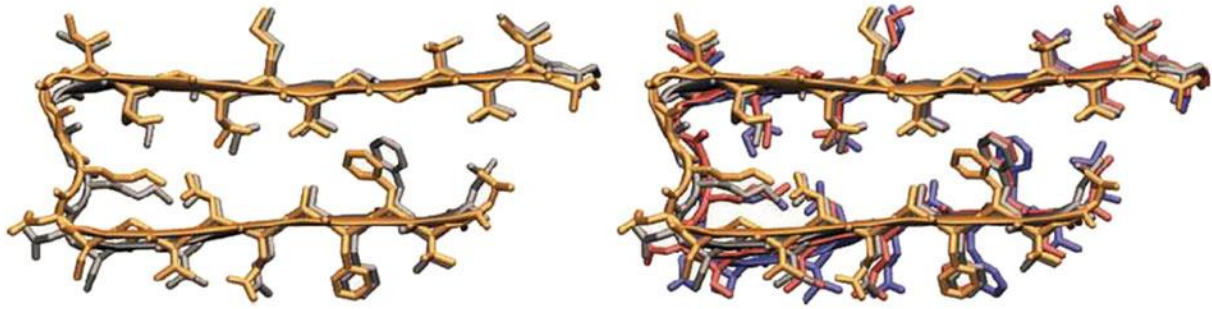
Figure 4. Structure of two stranded (left) and four stranded (right) amyloid aggregates (PDB ID: 2 BEG). Figures 3 and 4 reprinted from Perczel et al. [21,84] with permissions from the publishers.

Figure 5. Protein segments with a 'steric zipper' structure mesh tightly to form the spine of amyloid fibrils. Reprinted from Schnabel [98] with permission from the publisher.
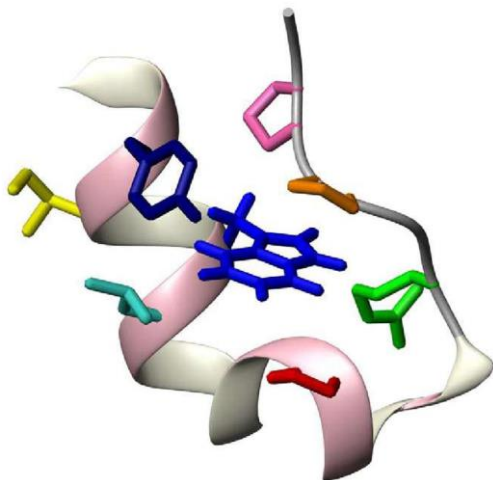
Figure 6. NMR-derived native state structure of TC5b. Highlighted is the Trp 6 residue (blue) together with the cage-forming side chains of Tyr 3 (dark blue), Ile 4 (yellow), Leu 7 (light blue), Pro 12 (red), Arg 16 (green), Pro 18 (brown), and Pro 19 (pink). Reprinted from Rogne et al. [130].