FEBS Letters 583 (2009) 953–959

FEBS Letters

journal homepage: www.FEBSLetters.org

Minireview

Insight into "insoluble proteins" with pure water

Jianxing Song*

Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 27 January 2009
Revised 14 February 2009
Accepted 16 February 2009
Available online 20 February 2009

Edited by Peter Brzezinski

Keywords:
Water
Insoluble protein
Protein folding
Protein solubility
Ionic strength
Circular dichroism (CD)
NMR

ABSTRACT

Many proteins are not refoldable and also insoluble. Previously no general method was available to solubilize them and consequently their structural properties remained unknown. Surprisingly, we recently discovered that all insoluble proteins in our laboratory, which are highly diverse, can be solubilized in pure water. Structural characterization by CD and NMR led to their classification into three groups, all of which appear trapped in the highly disordered or partially-folded states with a substantial exposure of hydrophobic side chains. In this review, I discuss our results in a wide context and subsequently propose a model to rationalize the discovery. The potential applications are also explored in studying protein folding, design and membrane proteins.

1. Water, life and protein folding

Water covers two-thirds of our planet. At least to our experience, where there is water, there is life [1]. Water is widely regarded as the 'matrix of life' which is not just a passive scaffold but also has many active roles in molecular biology [2]. The absolutely-essential role of water for life is associated with its very extraordinary properties which are believed to be irreplaceable by another single molecule system. For example, ice floats over liquid water, thus allowing fish to survive even the top is frozen in rivers, lakes and oceans. Despite the apparent simplicity of the water molecule, water is probably the most mysterious substance in our world. For example, in the late 1960s, Russian chemist Boris Deryaguin and his colleagues claimed that polywater, a gel-like form of water, could be observed in small capillary tubes [2]. Even more unbelievably, in the 1980s Benveniste and his collaborators proposed the notion of the 'memory of water', whereby the liquid water could allegedly be imprinted by biomolecular information which is still wheeled out to justify homeopathy [2,3]. On the other hand, the deeper we probe into water by new scientific techniques, the more problems and puzzles rise. This is elegantly exemplified by the recent report claiming that instead of the classic tetrahedral structure, molecules in liquid water can only have two strong hydrogen bonds, thus resulting in water chains and rings [4,5].

All biomolecules function in water and in particular, water is central to their structure, stability, dynamics and function. So far, no other liquid has been demonstrated to be able to substitute water and meanwhile still compatible with the existence of complex biomolecules. As a consequence, water can not be simply treated as an inert environment, but rather should be considered to be an integral and active constituent of biomolecular systems [5–10].

Proteins are the most important player that implements the most difficult but essential tasks in living cells [11]. A protein is a linear heteropolymer of 20 common amino acids connected by peptide bonds. Although different amino acids have distinctive characteristics, they can be briefly divided into two groups: hydrophobic (or non-polar) and hydrophilic (or polar). To be functional, many proteins need to fold via intermediates into unique three-dimensional structures composed of spatially organized polypeptide fragments that assume different secondary structures, including the α-helix, β-sheet, reverse turns, and loops [12–15]. On the other hand, some proteins may remain highly unstructured, without stable secondary and tertiary structures, but are fully functional, thus called "intrinsically unstructured proteins" [16–19]. Nevertheless, to fold or to remain intrinsically unstructured mainly depends on the interplay between water and the protein sequence characteristic of amino acid composition, complexity, hydrophobicity, charge, flexibility, etc. In general, the

* Address: Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. Fax: +65 67792486.
E-mail address: bchsj@nus.edu.sg

0014-5793/$36.00
doi:10.1016/j.febslet.2009.02.022

intrinsically-unstructured proteins are significantly lacking of bulky hydrophobic (Ile, Leu, and Val) and aromatic (Trp, Tyr, and Phe) amino acid residues but are dramatically enriched in polar (Arg, Gly, Gln, Ser, Pro, Glu, and Lys) and structure-breaking (Gly and Pro) residues [16–18]. As such, despite being highly unstructured, they are soluble in buffer.

For a typical well-folded protein, its folding process is extensively thought to be dictated by the aqueous solution (55 M concentration of water). More specifically, the conflict between the hydrophobic side chains in the protein and the polar nature of the water represents the main driving force for protein folding. This conflict forces the hydrophobic groups to collapse into a tightly-packed core in which more than 80% of the nonpolar side chains are buried. Consequently, in the three-dimensional structure of a well-folded native protein, the majority of hydrophobic residues are buried in the internal core, thus being shielded from water while most hydrophilic residues are exposed to the bulk solvent, substantially contributing to the formation of the hydration shell by hydrogen bonding to water molecules. It has been even proposed that the protein motions and water dynamics are highly coupled. The fluctuation of the hydration water is able to slave the protein dynamics, thus mediating its function [8,20].

Proteins are only marginally stable. Even for a well-fold protein, its thermodynamic stability resulting from a balance of large opposing forces is relatively small, with ~5–20 kcal/mol in free energy more stable than unfolded states under physiologic conditions [21–25]. As a consequence, proteins undergo inherently dynamic fluctuations among different conformations and become partially unfolded even under modest denaturing conditions such as acidic pH [14,26,27]. In some cases the free energy gap separating the native from the denatured state is small enough to allow spontaneous denaturation. In a partially folded or denatured protein, the hydrophobic side chains are anticipated to be more exposed to the bulk solvent, and consequently they have a high tendency to be clustered together to form intermolecular aggregates.

One commonly-encountered problem in protein research and application is their insolubility. Recently protein misfolding/aggregation becomes a hot topic of both fundamental and practice interest. In vivo protein misfolding/aggregation have been extensively demonstrated to be responsible for a large spectrum of the neurodegenerative diseases including Spongiform encephalopathies, Alzheimer's and Parkinson's diseases [28]. In vitro protein aggregation presents a large challenge to the bio-pharmaceutical industry, where aggregation occurs frequently for a therapeutic protein [21].

2. Solubilizing "insoluble proteins" with pure water

Previously, insoluble proteins were not assessable by the high resolution biophysical methods including NMR spectroscopy and crystallography. Although X-ray crystallography collects data on a protein in crystal, a high protein concentration in solution is needed first for crystallization. On the other hand, based on the structure genomics projects, it has been estimated that ~35–50% of the proteins expressed in *Escherichia coli* cells were in inclusion bodies, a large portion of which were not refoldable in buffer systems with any currently-available methods [29,30]. Despite extensive efforts to reduce formation of inclusion bodies in *E. coli cells* by fusing with a highly-soluble tag, coexpressing folding catalysts and chaperones, reducing culture temperature, and modifying culture media, these approaches do not always work [31]. This observation implies that the intrinsic property of a protein may be responsible for its insolubility. Therefore, structural characterization of insoluble proteins would provide valuable insights into their properties and aggregation mechanism, and moreover offer rationales for enhancing protein solubility. Unfortunately, previously no general method was available to solubilize these proteins without addition of detergents or/and denaturants such as sodium dodecyl sulfate, urea, and guanidine hydrochloride at high concentrations.

About 6 years ago, we started to work on the 66-residue extracellular domain of the human Nogo proteins designated as Nogo-66, in an attempt to determine its NMR structure for further design of mimetics to enhance the central nervous system (CNS) regeneration. Unfortunately, despite an extensive optimization of buffer systems by varying the salt type, concentration and pH, the isolated Nogo-66 protein was found to be highly insoluble. As such we were only able to determine the NMR structure of the soluble Nogo-40 with the C-terminal 26 residues of Nogo-66 deleted [32]. However, Nogo-40 was highly disordered in the aqueous buffer and assumed a two-helix conformation only in the presence of 50% trifluoroethanol (TFE). Intriguingly, later on in further optimizing buffer conditions we found that the RP-HPLC purified Nogo-66 with the estimated p*I* of 7.7 could be solubilized in salt-free water at a high protein concentration at pH 4.0 at which Nogo-66 had estimated charges of 15.2. In salt-free water, Nogo-66 had a CD spectrum typical of a helical conformation. To understand this phenomenon, we further generated several differentially-truncated forms of Nogo-66 and studied them by CD and NMR. The results led to the identification of Nogo-60 with the C-terminal 6 residues deleted which had a CD spectrum highly similar to that of Nogo-66 (Fig. 1a) as well as a well-dispersed HSQC spectrum in salt-free water (Fig. 1b). Although Nogo-60 still remained totally insoluble in buffer, we succeeded in determining its solution structure in salt-free water by use of heteronuclear three-dimensional NMR spectroscopy [33]. The NMR structure immediately reveals why Nogo-40 is highly disordered while Nogo-60 is helical. As seen in Fig. 1c, Nogo-60 is composed of three helices 1–3 while the corresponding region of Nogo-40 can only form two helices 1 and 2, which are very similar to those of Nogo-40 in the presence of 50% TFE. Interestingly the middle helix is unusually long, containing ~20 residues. Remarkably, the detailed analysis revealed that the formation of the stable helical conformation of Nogo-60 appeared to require the packing interaction between the long middle helix and the third helix which was absent in Nogo-40 [33]. This packing interaction may function to shield the unusually-long middle helix from being highly exposed to the bulk water, thus stabilizing the intrinsic helix-forming propensity of the whole molecule. In contrast, the absence of this packing in Nogo-40 would result in a severe solvation of the middle helix, thus resulting in a highly-disordered state. This also implies that one mechanism for TFE to stabilize the helical structure might be to protect the helix-forming residues from a severe solvation.

On the other hand, the Nogo-60 structure in salt-free water also sheds light on why it is high insoluble in buffer. As seen in Fig. 1c, the C-terminal 6-residues Leu55-Phe56-Leu57-Val58-Asp59-Asp60 were highly unstructured, in particular with the four unburied hydrophobic residues Leu55-Phe56-Leu57-Val58 forming a large and exposed surface. Therefore we speculated that this exposed hydrophobic surface might be mainly responsible for its insolubility in buffer. Indeed, the designed Nogo-54 with the last six residues removed (Fig. 1d) suddenly became soluble in buffer and had very similar conformations in both salt-free water and buffer [33]. Interestingly, thermal unfolding results indicate that Nogo-54 in salt-free water was much more stable than that in buffer. On the other hand, recently we also demonstrated that buffer-soluble Nogo-54 indeed had the inhibitory activity of Nogo-66 in vivo [34], suggesting that the protein conformation in pure water has no fundamental difference from the active form in buffer.

Although I have been deeply puzzled by this discovery, at that time I still considered that the successful solubilization of Nogo-60 with pure water should be resulting from some unique property

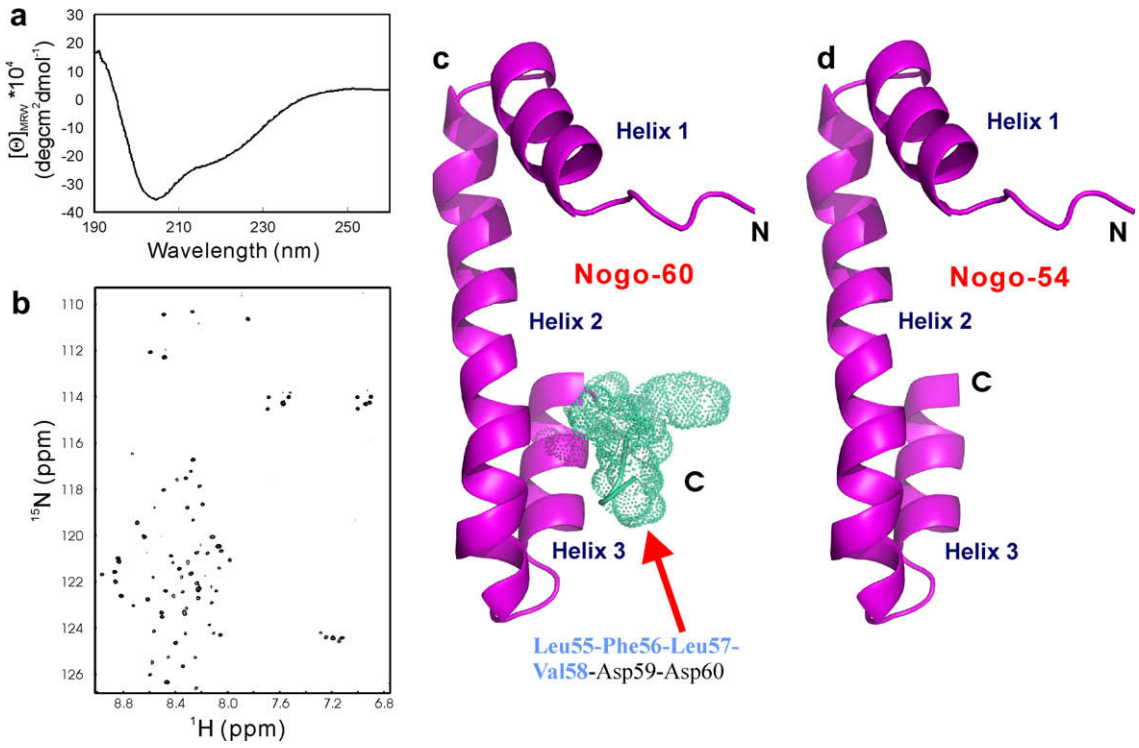

Fig. 1. Nogo-60 representing Group 3 of insoluble proteins. (a) Far-UV CD spectrum of Nogo-60 in salt-free water at pH 4.0. (b) ^{1}H–^{15}N NMR HSQC spectrum of Nogo-60 in salt-free water at pH 4.0. (c) NMR structure of Nogo-60 determined in salt-free water at pH 4.0 and a protein concentration of 600 μM. The C-terminal 4 residues forming a large hydrophobic surface are displayed in spheres. (d) The designed Nogo-54 with the last six residues removed is soluble both in salt-free water and buffer.

encoded in the Nogo-60 molecules because the relationship between protein solubility and ionic strength has been well established which contains two phases: salting-in and salting-out. In other words, solubility of proteins should be increased with the addition of small amounts of salt and then decreased in more concentrated solutions of electrolytes [35–39]. However, despite intense examination of Nogo-66 sequence and structure, I failed to find any unique property which might account for the discovery. Because of this, the results with Nogo-60 were retained for quite a while before submission for publication. However, in the flight back from the visit to Bethlehem, Birthplace of Jesus, I suddenly realized that before moved to Singapore, I have encountered myself several other proteins which were also totally insoluble in buffer but could be dissolved in salt-free water after HPLC purification. At that moment, I also thought that these proteins carried some unique properties thus ignored the observations. Nevertheless, by comparing those proteins to Nogo-66 again and again, I still failed to identify any correlation. Therefore I was tempted to think that solubilization of insoluble proteins with salt-free water might represent a general phenomenon, rather than due to the uniqueness of those proteins. As such it is certainly worth to test on more insoluble proteins despite being apparently incompatible with the established notion. Eventually I succeeded in convincing all members in my lab to test on their insoluble proteins.

To my great surprise, all 11 buffer-insoluble protein fragments/domains we had, with a great diversity of cellular function, location, and molecular size, could be readily solubilized in salt-free water at high protein concentrations [40]. For all those 11 proteins, we have devoted significant efforts to optimize buffer systems by varying the salt type, concentration and pH, in an attempt to solubilize them but all failed. Their successful solubilization in pure water thus offered us an unprecedented opportunity to investigate the structural properties of these previously-thought insoluble proteins by use of circular dichroism (CD) and NMR ^{1}H–^{15}N heteronuclear single quantum correlation (HSQC) spectroscopy. The CD and NMR results led to the classification of these proteins into three groups: Group 1, with no stable secondary structure as detected by CD, and with narrowly-dispersed but sharp HSQC peaks; Group 2, with stable secondary structure by CD but with HSQC peaks broadened and, consequently, only a small set of peaks detectable; and Group 3, with stable secondary structure by CD, and with narrowly-dispersed but sharp HSQC peaks. Group 3 proteins can be nicely represented by Nogo-60 which has been discussed above in details. On the other hand, Groups 1 and 2 can be exemplified respectively by the Avian Influenza receptor-binding domain (RBD) and the reduced Nogo-66 receptor (NgR).

The 152-residue RBD representing Group 1 was isolated from the 561-residue hemagglutinin of the Avian Influenza A virus, with the estimated p*I* of 9.67. In the context of the full-length hemagglutinin, RBD is well-folded and adopts a β-sheet dominant structure (Fig. 2a). However, upon being separated from hemagglutinin, it was totally insoluble even in 5 mM buffer and also not refoldable by a variety of refolding protocols. Remarkably, it could be dissolved at a protein concentration of 400 μM in salt-free water at pH 6.2 at which it had estimated charges of 14.8. In the salt-free water, RBD has a far-UV CD spectrum (Fig. 2b) typical for a highly disordered protein. On the other hand, although it has a narrowly-dispersed HSQC spectrum, the NMR resonance peaks are sharp and visible for almost all no proline residues, as evidenced from the well-separated HSQC peaks for all 9 Gly residues in the sequence (Fig. 2c). RBD appears to be highly absent of stable secondary structure and tight tertiary packing although the existence of the dynamic residual structures cannot be ruled out. Very recently, we also found that the truncated form of a transcriptional activator ApLLP was soluble in buffer but the full-length could only be dissolved in salt-free water [41]. More interestingly, the truncated ApLLP was able to bind to the DNA fragment in buffer but the full-length showed no binding ability to DNA in salt-free water.

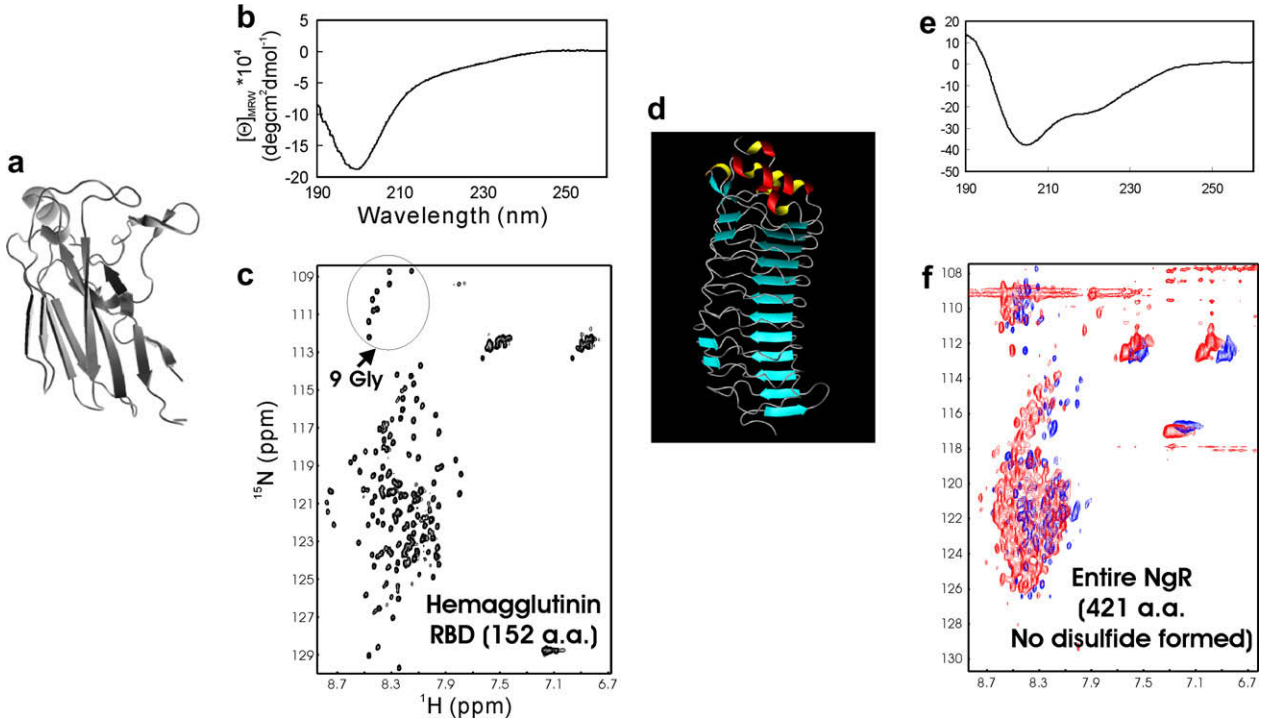

Fig. 2. RBD and NgR representing Groups 1 and 2 of insoluble proteins. (a) Three-dimensional structure of the Avian Influenza RBD in the context of the entire hemagglutinin (PDB code 1JSN). (b) Far-UV CD spectrum of the 152-residue RBD at pH 6.2. (c) ^{1}H–^{15}N NMR HSQC spectrum of RBD at pH 6.2 and a protein concentration of 400 μM. The HSQC peaks for nine Gly residues (including those from His-tag) are indicated by the arrows. (d) Three-dimensional structure of the N-NgR previously determined by X-ray crystallography (PDB code 1OZN). (e) Far-UV CD spectrum of the 421-residue NgR at pH 4.2. (f) ^{1}H–^{15}N NMR HSQC spectra of NgR in the absence (blue) and presence (red) of 8 M urea at pH 4.2 and a protein concentration of 120 μM.

The 421-residue NgR with the estimated p*I* of 8.82 representing Group 2 was derived from the 473-residue Nogo-66 receptor with its signal peptide and GPI-anchor sequences removed. Its Nogo-66 binding domain adopts a leucine-rich repeat fold (Fig. 2d) which usually gives rise to helix-like CD spectra probably due to its unique solenoid-like fold [40]. However, when expressed in *E. coli* cells, it was not refoldable and completely insoluble in a variety of buffer systems because of lack of the formation of six disulfide bridges. However, it could again be solubilized in salt-free water at pH 4.2 at which it had estimated charges of 53.8 thus were assessable by CD and NMR spectroscopy. Very amusingly, if judged from its far-UV CD spectrum (Fig. 2e), it might be concluded that the native-like secondary structure might already form in the disulfide-lacking NgR [40]. On the other hand, NgR had an HSQC spectrum with a narrow dispersion as well as very broad NMR peaks (Fig. 2f). Addition of 8 M urea resulted in no significant change in the near-UV CD spectrum, indicating that the side-chain packing was largely disrupted even without urea. However, introduction of 8 M urea did lead to the appearance of many new HSQC peaks (Fig. 2f). These results clearly indicate that the disulfide-lacking NgR undergoes intermediate conformational exchanges or/and dynamic aggregation, largely owing to its fluctuating side-chain packing [14,26,42,43]. It appears that the absence of six disulfide bridges leads NgR to being trapped in the molten globule state with both native-like secondary structure and tertiary topology, but without a tight side-chain packing. In this regard, although the disulfide formation appears not essential for folding, it is vital for maintaining protein stability and tight side-chain packing [44,45]. Recently we also demonstrated that before the formation of two disulfide bridges, the ephrin-B2 ectodomain could only be solubilized in salt-free water which underwent intermediate conformational exchanges or/and dynamic aggregation. However, once the two disulfide bridges were correctly formed by in vitro refolding of the same sample in the buffer containing both oxidized and reduced glutathione, the ephrin-B2 ectodomain suddenly became well-structured as well as functional in binding to its receptor EphB2 [46]. More incredibly, we recently also discovered that the full-length Nogo-B receptor with ~30 residue transmembrane fragment could also be solubilized in salt-free water with molten globule like properties [47].

So far we have tested on 30 proteins insoluble and not refoldable in a variety of buffer systems, but all of them could be solubilized in salt-free water, manifesting a diverse spectrum of structural states covering all possible conformations previously observed for highly or partially unfolded proteins. However, we failed to find any insoluble protein fragment/domain with a tight tertiary packing. This observation strongly implies that proteins are insoluble in buffers, probably because they lack an intrinsic propensity to reach or/and maintain the well-packed native state and consequently were trapped in the highly-disordered states, or molten-globule which result in a significant exposure of hydrophobic side chains, thus owning a tremendous tendency to aggregation in the presence of salt ions.

3. Model for solubilizing insoluble proteins with pure water

So what is the mechanism by which insoluble proteins can be solubilized in pure water? In general, protein solubility can be regarded as the outcome of the complex interplay among various interactions between protein–protein, protein–water, protein–ion and ion–water. Protein solubility has been extensively studied since the beginning of the last century, particularly the 1930s and in the last decade. The key work of Green offered the experimental basis to the rules that protein solubility is minimal at its pI, followed by the first increases (salting-in) and then decreases (salting-out) with the increase of ionic strength [35–38]. Despite

extensive studies, the molecular mechanisms underlying protein solubility and salt effect still remain poorly-understood and in particular very controversy [48,49]. Nevertheless, it has been widely accepted that the protein aggregation is mainly mediated by electrostatic and hydrophobic interactions between individual molecules. In general, the inter-molecular hydrophobic interaction will trigger the clustering of hydrophobic side chains of individual molecules, thus leading to precipitation/aggregation. On the other hand, electrostatic interaction contains two terms: the repulsive force between protein molecules carrying the same net charges; and dispersive/attractive force between protein molecules with permanent and induced dipoles.

Here I would propose a schematic model to rationalize why an insoluble protein can be solubilized in pure water. As extensively documented, the isoelectric point (p*I*) of the majority of globular proteins is in the pH range of 5–8. On the other hand, most previous examples of salting-in were established from well-folded proteins and their solubility data were acquired from experiments performed near the p*I* of these proteins, such as carboxyhemoglobin and lactoglobulins [38]. As shown in Fig. 3a, for a well-fold globular protein, the majority of the hydrophobic side chains are buried in the core while on the surface are mainly hydrophilic side chains. In this regard, the inter-molecular hydrophobic interaction is expected to be relatively weak. If the protein is in salt-free water with pH close to its p*I*, the net charge of the protein is very small and consequently the repulsive electrostatic interaction is also weak. As a result, the main force in mediating the intermolecular interaction will be the attractive electrostatic portion. The individual molecules of a well-folded protein have the strongest tendency to come together, mainly driven by the attractive electrostatic force in salt-free water with the pH close to the pI of the protein, and thus the solubility is the lowest under this condition. However, if a small amount of salt ions are introduced into this system (Fig. 3b), the charged ions will screen out the attractive electrostatic interaction to a certain degree, which will reduce the attractive intermolecular interaction, thus leading to the increase of the solubility (salting-in). If the salt concentration is further increased, the protein solubility will decrease, thus resulting in precipitation (salting-out). Consequently, for a well-folded protein in aqueous solution with pH close to its p*I*, the classic curve is obtained for protein solubility versus salt concentration (Fig. 3c).

By contrast, as we recently revealed [33,34,40,41,46,47], insoluble proteins represent a special class of proteins which are trapped in the highly-disordered or molten globule states without a tight tertiary packing. Because of this, these proteins have a substantial amount of hydrophobic side chains exposed to the bulk water. As seen in Fig. 3d, if a protein of this family is dissolved in the salt-free water with pH deviated from its p*I* (for example, pH 4.0 as we usually used), the individual molecules will bear a significant amount of net charges. In the regard, the repulsive electrostatic interaction or/and large protein hydration shell will constitute an energy barrier unfavorable for inter-molecular interactions in pure water. As a result, in pure water the hydrophobic interaction and attractive electrostatic force will be significantly suppressed, and the precipitation/aggregation is prevented. However, if even a small amount of salt is introduced, the repulsive electrostatic interaction will be screen out or/and the protein hydration shell may be disrupted to some extent. As a sum the hydrophobic interaction will become dominant, thus leading to immediate precipitation/aggregation (Fig. 3e). These will give rise to a non-classic curve of protein solubility versus salt concentration with a very steep slope (Fig. 3f). For example, at pH 4.0 Nogo-60 would precipitate immediately at a NaCl concentration of 5 mM while at pH 4.2 NgR would precipitate rapidly at a NaCl concentration of 3 mM. According to this model, for an insoluble protein, the solubility is the highest in pure water and the presence of even a small amount of salt ions will trigger immediate precipitation. Furthermore, this model is also able to explain our observation that in the salt-free water, the binding interaction between the biomolecules was dramatically blocked [41] because the repulsive electrostatic force or/and large hydration shell will significantly reduce the probability for individual protein molecules to collide.

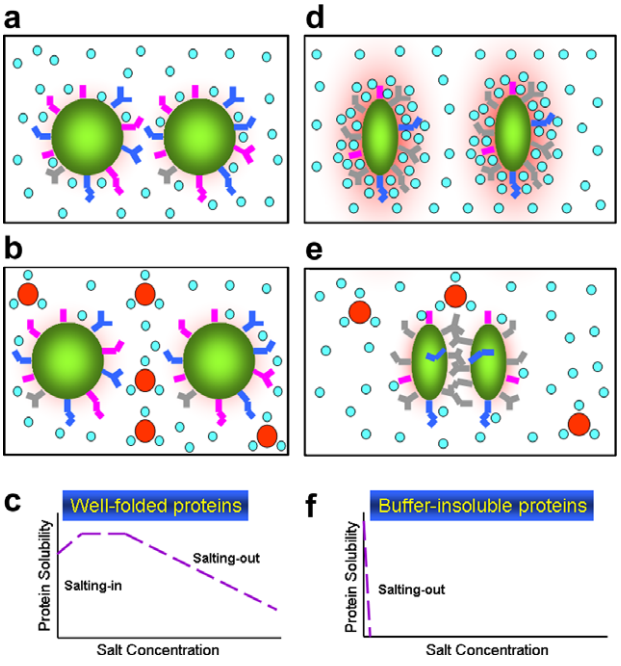

Fig. 3. Proposed model for salt effect on a well-folded and an insoluble proteins. (a–c) Salt effect on a well-folded protein. (a) Well-folded protein molecules in salt-free water with the pH value close to its p*I*. The green balls stand for the protein molecules. The hydrophobic side chains are represented by grey sticks. Purple and blue sticks are used to represent respectively the hydrophilic side chains positively-charged and negatively-charged. The red dot cloud is utilized to symbolize the net charge. The small-size cyan balls are water molecules. (b) Well-folded protein molecules in the presence of a small amount of salt. The medium-size red balls are salt ions with some water molecules clustered around them. (c) The classic curve of protein solubility versus salt concentration for a well-folded protein. (d–f) Salt effect on an insoluble protein. (d) Insoluble protein in salt-free water. The green ellipsoids stand for insoluble protein molecules with a large amount of hydrophobic side chains exposed. (e) Insoluble protein molecules in the presence of a small amount of salt. (f) The unique curve of protein solubility versus salt concentration for an insoluble protein.

4. Implications and potential applications

Almost all religions decree the magic power of pure water. On the other hand, probably all biological reactions occur in salty water and consequently pure water seems irrelevant to the biological systems. Now our studies reveal that pure water does indeed own a unique ability to allow insoluble proteins to manifest their full conformation states in aqueous solution. Our discovery also sheds light on a previously unknown regime associated with proteins: proteins appear so designed that in pure water their intrinsic repulsive interactions are sufficient to suppress the attractive forces, thus preventing them from severe precipitation/aggregation. Interestingly, as shown by our analysis of Nogo-54, the conformations of a protein are very much similar in both pure water and buffer. This implies that the driving forces for protein folding may remain fundamentally unchanged in aqueous solution regardless of the presence or absence of salt ions. Our discovery also holds several potential applications. First of all, it may largely facilitate our experimental understanding of protein folding and

further rationalizing protein design. Also, by adjusting salt concentration, the aggregation rate may be manipulated and thus monitored by high-resolution biophysical methods such as NMR spectroscopy. This will allow our insight into the molecular details of the aggregation events of significant biological relevance.

As we have demonstrated on Nogo-66 fragments [33,34], the determination of the NMR structure of buffer-insoluble Nogo-60 in salt-free water offered key rationales to finally design of buffer-soluble Nogo-54 which retained both solution conformation and biological activity. More generally, the discovery may be valuable to experimentally assessing the relationship between the protein sequence and three-dimensional structure by the structural analysis of protein mutation or fragment libraries. One severe problem encountered in these approaches is that many mutants or fragments are highly insoluble and consequently their structural properties can not be characterized [50,51]. In fact, we have tested the approaches by using the first hNck2 SH3 domain as a model system. This 58-residue small protein without any disulfide bridge shares a classic SH3 fold, namely the β-barrel fold comprising five β-stands, which are organized into two β-sheets [27,52]. We first dissected the 58-residue SH3 domain into a library of small fragments according to their secondary structure boundary. Despite insolubility of several peptide fragments, we succeeded in solubilizing and further characterizing them in salt-free water. Intriguingly, first detected by CD, then confirmed by detailed NMR analysis, the conformation of the fragment without the N- and C-terminal short β-strands were found to switch from the native β-dominant to a totally helical conformation. With further mutagenesis and NMR characterization, we provided the first high-resolution NMR evidence showing that the classic SH3 fold shared by more than 4000 SH3 domains owns the chameleon ability, which can form highly populated helical conformations upon becoming partially folded [27]. On the same SH3 domain, we also identified a group of single mutants which suddenly rendered the proteins into highly buffer-insoluble. However, our successful solubilization in salt-free water again allowed us to conduct an exhaustive NMR structural and dynamic characterization on one representative mutant. The results indicate that on the one hand it had a helical secondary structure; on the other it still owned a native-like topology with significantly-limited backbone motions. Furthermore, we were also able to follow the dynamic aggregation process of this mutant by a progressive addition of NaCl and monitored by NMR (Liu and Song, unpublished data). Interestingly, recently a report showed that by a gradual addition of NaCl and $CaCl_2$ salts in salt-free water, the muti-step assembly process of the nanosphere formed by Amelogenin could be mapped to a residue-level by use of NMR spectroscopy [53].

Our discovery also bears some practical applications. In fact, we found that the efficacy in in vitro refolding to form disulfide bridges could be enhanced by incubating the HPLC-purified protein sample in slightly-acidic salt-free water for a while before addition of buffering salt and reduced and oxidized glutathiones (unpublished observations). This approach is especially useful in refolding proteins with a relatively weak propensity to form correct disulfide bridges and its efficacy may be further improved if the refolding reaction is performed in the anaerobic chamber [42]. Furthermore, our discovery may provide a powerful tool to study membrane proteins without needing the presence of lipid molecules. Last year, we showed that the buffer-insoluble Nogo-B receptor with ~30-residue transmembrane fragment could be solubilized in salt-free water at a high protein concentration [47]. Very recently, a study reported that an integral membrane channel protein could be solubilized in salt-free water without membranes, and more surprisingly it still adopted a three-dimensional structure very similar to that in the membrane environment [54].

Acknowledgements

I would like to thank all the lab members for their important contribution. This study is supported by Ministry of Education (MOE) Tier 2 Grant R-154-000-388-112, and Biomedical Research Council of Singapore (BMRC) Young Investigator Award, R-154-000-217-305 to Jianxing Song.

References

[1] Ball, P. (2005) Water and life: seeking the solution. Nature 436, 1084–1085.
[2] Ball, P. (2008) Water: water – an enduring mystery. Nature 452, 291–292.
[3] Davenas, E., Beauvais, F., Amara, J., Oberbaum, M., Robinzon, B., Miadonna, A., Tedeschi, A., Pomeranz, B., Fortner, P., Belon, P., Sainte-Laudy, J., Poitevin, B. and Benveniste, J. (1988) Human basophil degranulation triggered by very dilute antiserum against IgE. Nature 333, 816–818.
[4] Wernet, P., Nordlund, D., Bergmann, U., Cavalleri, M., Odelius, M., Ogasawara, H., Näslund, L.A., Hirsch, T.K., Ojamäe, L., Glatzel, P., Pettersson, L.G. and Nilsson, A. (2004) The structure of the first coordination shell in liquid water. Science 304, 995–999.
[5] Zubavicus, Y. and Grunze, M. (2004) New insights into the structure of water with ultrafast probes. Science 304, 974–976.
[6] Despa, F. (2005) Biological water: its vital role in macromolecular structure and function. Ann. NY Acad. Sci. 1066, 1–11.
[7] Finney, J.L. (2004) Water? What's so special about it! Philos. Trans. Roy. Soc. Lond. B: Biol. Sci. 359, 1145–1163.
[8] Levy, Y. and Onuchic, J.N. (2006) Water mediation in protein folding and molecular recognition. Annu. Rev. Biophys. Biomol. Struct. 35, 389–415.
[9] Ball, P. (2008) Water as an active constituent in cell biology. Chem. Rev. 108, 74–108.
[10] Ball, P. (2008) Water as a biomolecule. Chemphyschem 9, 2677–2685.
[11] Pace, C.N., Trevino, S., Prabhakaran, E. and Scholtz, J.M. (2004) Protein structure, stability and solubility in water and other solvents. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 59, 1225–1234.
[12] Kim, P.S. and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu. Rev. Biochem. 51, 459–489.
[13] Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins Struct. Funct. Genet. 6, 87–103.
[14] Song, J., Jamin, N., Gilquin, B., Vita, C. and Menez, A. (1999) A gradual disruption of tight side-chain packing: 2D ^{1}H-NMR characterization of acid induced unfolding of CHABII. Nat. Struct. Biol. 6, 129–134.
[15] Udgaonkar, J.B. (2008) Multiple routes and structural heterogeneity in protein folding. Annu. Rev. Biophys. 37, 489–510.
[16] Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.
[17] Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. 579, 3346–3354.
[18] Uversky, V.N., Oldfield, C.J. and Dunker, A.K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu. Rev. Biophys. 37, 215–246.
[19] Li, M. and Song, J. (2007) The N- and C-termini of the human Nogo molecules are intrinsically unstructured: bioinformatics, CD, NMR characterization and its functional implications. Proteins 68, 100–108.
[20] Fenimore, P.W., Frauenfelder, H., McMahon, B.H. and Parak, F.G. (2002) Slaving: solvent fluctuations dominate protein dynamics and functions. Proc. Natl. Acad. Sci. USA 99, 16047–16051.
[21] Chi, E.Y., Krishnan, S., Randolph, T.W. and Carpenter, J.F. (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm. Res. 20, 1325–1336.
[22] Privalov, P.L. and Gill, S.J. (1988) Stability of protein structure and hydrophobic interaction. Adv. Protein Chem. 39, 191–234.
[23] Dill, K.A. (1990) Dominant forces in protein folding. Biochemistry 29, 7133–7155.
[24] Jaenicke, R. (1991) Protein folding: local structures, domains, subunits, and assemblies. Biochemistry 30, 3147–3161.
[25] Bowler, B.E. (2007) Thermodynamics of protein denatured states. Mol. Biosyst. 3, 88–99.
[26] Wei, Z. and Song, J. (2005) Molecular mechanism underlying the thermal stability and pH-Induced unfolding of CHABII. J. Mol. Biol. 348, 205–218.
[27] Liu, J. and Song, J. (2008) NMR evidence for forming highly populated helical conformations in the partially folded hNck2 SH3 domain. Biophys. J. 95, 4803–4812.
[28] Chiti, F. and Dobson, C.M. (2006) Protein misfolding, functional amyloid, and human disease. Annu. Rev. Biochem. 75, 333–366.
[29] Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H. and Edwards, A.M. (2000) Structural proteomics: prospects for high throughput sample preparation. Prog. Biophys. Mol. Biol. 73, 339–345.
[30] Pedelacq, J.D., Piltch, E., Liong, E.C., Berendzen, J., Kim, C.Y., Rho, B.S., Park, M.S., Terwilliger, T.C. and Waldo, G.S. (2002) Engineering soluble proteins for structural genomics. Nat. Biotechnol. 20, 927–932.

[31] Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., Philo, J.S. and Arakawa, T. (2004) Role of arginine in protein refolding, solubilization, and purification. Biotechnol. Prog. 20, 1301–1308.
[32] Li, M., Shi, J., Wei, Z., Teng, F., Tang, B. and Song, J. (2004) Structural characterization of the human Nogo-A functional domains. Solution structure of Nogo-40, a Nogo-66 receptor antagonist enhancing injured spinal cord regeneration. Eur. J. Biochem. 271, 3512–3522.
[33] Li, M., Liu, J. and Song, J. (2006) Nogo goes in the pure water: solution structure of Nogo-60 and design of the structured and buffer-soluble Nogo-54 for enhancing CNS regeneration. Protein Sci. 15, 1835–1841.
[34] Li, M., Li, Y., Liao, X., Liu, J., Qin, H., Xiao, Z. and Song, J. (2008) Rational design, solution conformation and identification of functional residues of the soluble and structured Nogo-54, which mimics Nogo-66 in inhibiting the CNS neurite outgrowth. Biochem. Biophys. Res. Commun. 373, 498–503.
[35] Green, A. (1932) Studies in the physical chemistry of the proteins. VIII. The solubility of hemoglobin in concentrated salt solutions: a study of the salting-out of proteins. J. Biol. Chem. 93, 495–516.
[36] Green, A. (1932) Studies in the physical chemistry of the proteins. IX. The effect of electrolytes on the solubility of hemoglobin in solutions of varying hydrogen ion activity with a note on the comparable behavior of casein. J. Biol. Chem. 93, 517–542.
[37] Green, A.A. (1932) Studies in the physical chemistry of the proteins. X. The solubility of hemoglobin in solutions of chlorides and sulfates of varying concentration. J. Biol. Chem. 95, 47–66.
[38] Retailleau, P., Ries-Kautt, M. and Ducruix, A. (1997) No salting-in of lysozyme chloride observed at low ionic strength over a large range of pH. Biophys. J. 73, 2156–2163.
[39] Zhou, H.X. (2005) Interactions of macromolecules with salt ions: an electrostatic theory for the Hofmeister effect. Proteins 61, 69–78.
[40] Li, M., Liu, J., Ran, X., Fang, M., Shi, J., Qin, H., Goh, J.M. and Song, J. (2006) Resurrecting abandoned proteins with pure water: CD and NMR studies of protein fragments solubilized in salt-free water. Biophys. J. 91, 4201–4209.
[41] Liu, J. and Song, J. (2008) A novel nucleolar transcriptional activator ApLLP for long-term memory formation is intrinsically unstructured but functionally active. Biochem. Biophys. Res. Commun. 366, 585–591.
[42] Song, J., Bai, P., Luo, L. and Peng, Z.Y. (1998) Contribution of individual residues to formation of the native-like tertiary topology in the alpha-lactalbumin molten globule. J. Mol. Biol. 280, 167–174.
[43] Bai, P., Song, J., Luo, L. and Peng, Z.Y. (2001) A model of dynamic side-chain–side-chain interactions in the alpha-lactalbumin molten globule. Protein Sci. 10, 55–62.
[44] Song, J., Gilquin, B., Jamin, N., Drakopoulou, E., Guenneugues, M., Dauplais, M., Vita, C. and Menez, A. (1997) NMR solution structure of a two-disulfide derivative of charybdotoxin: structural evidence for conservation of scorpion toxin a/b motif and its hydrophobic side chain packing. Biochemistry 36, 3760–3766.
[45] Luo, Y. and Baldwin, R.L. (1999) The 28–111 disulfide bond constrains the a-lactalbumin molten globule and weakens its cooperativity of folding. Proc. Natl. Acad. Sci. USA 96, 11283–11287.
[46] Ran, X., Qin, H., Liu, J., Fan, J.S., Shi, J. and Song, J. (2008) NMR structure and dynamics of human ephrin-B2 ectodomain: The functionally critical C–D and G–H loops are highly dynamic in solution. Proteins 72, 1019–1029.
[47] Li, M. and Song, J. (2007) Nogo-B receptor possesses an intrinsically-unstructured ectodomain and a partially-folded cytoplasmic domain. Biochem. Biophys. Res. Commun. 360, 128–134.
[48] Baldwin, R.L. (1996) How Hofmeister ion interactions affect protein stability. Biophys. J. 71, 2056–2063.
[49] Zhang, Y. and Cremer, P.S. (2006) Interactions between macromolecules and ions: the Hofmeister series. Curr. Opin. Chem. Biol. 10, 658–663.
[50] Davidson, A.R. and Sauer, R.T. (1994) Folded proteins occur frequently in libraries of random amino acid sequences. Proc. Natl. Acad. Sci. USA 91, 2146–2150.
[51] Yi, Q., Scalley-Kim, M.L., Alm, E.J. and Baker, D. (2000) NMR characterization of residual structure in the denatured state of protein L. J. Mol. Biol. 299, 1341–1351.
[52] Liu, J., Li, M., Ran, X., Fan, J. and Song, J. (2006) Structural insight into the binding diversity between the human Nck2 SH3 domains and proline-rich proteins. Biochemistry 45, 7171–7184.
[53] Buchko, G.W., Tarasevich, B.J., Bekhazi, J., Snead, M.L. and Shaw, W.J. (2008) A solution NMR investigation into the early events of amelogenin nanosphere self-assembly initiated with sodium chloride or calcium chloride. Biochemistry 47, 13215–13222.
[54] Ma, D., Tillman, T.S., Tang, P., Meirovitch, E., Eckenhoff, R., Carnini, A. and Xu, Y. (2008) NMR studies of a channel protein without membranes: structure and dynamics of water-solubilized KcsA. Proc. Natl. Acad. Sci. USA 105, 16537–16542.