Water rotation barriers on protein molecular surfaces *

K. Tompa[1], M. Bokor[1], T. Verebélyi[1], P. Tompa[2,3,*]

1 Institute for Solid State Physics and Optics, Wigner RCP of the HAS, Budapest, Hungary

2 Institute of Enzimology, Research Center of the Natural Sciences of the HAS, Budapest, Hungary

3 VIB, Department of Structural Biology, Vrije Universitei Brussel, Brussels, Belgium

*corresponding author, tel. +32-26291962, email: ptompa@vub.ac.be

Abstract The experimental characterization of hindered-rotation barriers and mapping the energetic heterogeneity of water molecules bound to the molecular "surface" of proteins is critical for understanding the functional interaction of proteins with their environment. Here, we show how to achieve this goal by an original wide-line NMR procedure, which is based on the spectral motional narrowing phenomenon following the melting (thawing) process of interfacial ice. The procedure highlights the differences between globular and intrinsically disordered proteins and it enables to delineate the effect of solvent on protein structure, making a distinction between point mutants, monomeric and oligomeric states, and characterizing the molecular interactions taking part in different cellular processes. We put this unique experimental approach introducing novel physical quantities and quantifying the heterogeneous distribution of motional activation energy of water in the interfacial landscape into a historical perspective, demonstrating its utility through a variety of globular and disordered proteins.

1 Introduction

Understanding protein function at the molecular level is one of the best proxies to understanding life itself. Proteins are the workhorses of living systems, being involved in countless types of activities and interactions orchestrating the action of other types of molecules: partner proteins, substrates and other macromolecules, which collectively make up a living cell. For a long time, a dominant approach to deciphering protein function has been to solve 3D structures, primarily by X-ray crystallography, as witnessed by more than 100,000

structures deposited in the Protein Data Bank (PDB) [1]. So much insight in protein function has been offered and actually generated by this approach that the exclusivity of solving and analyzing static structural pictures of proteins has been practically unchallenged in protein science until very recently.

Of course, it has always been appreciated that protein structure can change, which forms an essential part of the functioning of some (e.g. allosteric) proteins [2], but it was only with the advent of the funnel view of the energy landscape of proteins [3] that it began to be appreciated that all proteins are better described as a dynamic ensemble of conformations. Whereas this transition suggests that function actually comes from the dynamics of structure [4], it also infers that the subtleties of function are tightly linked with its heterogeneous "surface" (interfacial region), which mediates its interactions with the environment. Interaction of the protein with one partner always occurs at the expense of interaction with others, including hydrate water molecules, which means that the description its heterogeneous energetic surface is indispensable for understanding its function [5,6]. It appears that this is one of the most neglected areas in molecular protein science.

The closely related subject of distinguishing ordered from disordered nature (state) of proteins arose some twenty years ago. Since the first comprehensive overview [7] on the concept and the techniques applied for studying intrinsically disordered (unstructured) proteins (IDPs/IUPs) [8,19] a rapid expansion of the field occurred and a continuously growing demand of more exact approaches and solutions, to the underlying structural concepts can be seen [10]. The categorization of ID/IU proteins and variety of terms and categories appearing in the literature [11] (e.g. pliable, flexible, mobile, vulnerable, chameleon, dancing) complemented by an endless combinations of „natively, naturally, inherently, intrinsically" with „unfolded, unstructured, disordered, denatured" indicate that their description by quantitative measures instead of simply contrasting them with structural order (as marked by terms disordered and unstructured) is highly desirable.

We suggest that part of the problem is that no well-defined, single order parameter is established in the field, which would be derived from a single, measurable physical quantity, and which could help classify a protein in quantitative terms as ordered, disordered or a given mixture of both. Our goal is to clear this issue through addressing the following questions: *i)* Can we define and measure one or more order-parameter-like quantities for quantifying the extent of the ordered or disordered state, and *ii)* Can the quantitative changes of such an order

parameter be followed in physiological processes? We have laid previously the ground of answering these questions by introducing the *HM* parameter, which characterizes the mobility of the hydrogen-hydrogen (proton-proton) radial vectors in lyophilized proteins [12]. Solid state (crystalline and/or lyophilized) proteins are motionless i.e. considered as rigid only at very low (at liquid nitrogen or much rather at liquid helium) temperatures. Although proteins shown earlier as disordered by several experimental methods score higher on the *HM* scale of mobility than globular lysozyme, they are far from the *HM* value of some small-molecule reference compounds, such as hydrocarbons. However, the mobility of the polycrystalline, globular lysozyme is not zero at room temperature, and it is an open question if there is any connection between the temperature at which the motion starts and the (extent of the) ordered or disordered state [13].

One may object that proteins cannot be classified by a single order-disorder parameter derived from gross characteristics of the geometrical structure, rather by the distribution of some physical and chemical properties and/or the attributes measured, and its relationship to physiological function(s). In accord, it is reasonable to ask what properties, beyond the geometrical structure (architecture), determine the physiological function of proteins. We suggest that surface hydration, which mediates the interactions of the protein with its environment, is a critical and characteristic parameter worth taking under closer scrutiny. To this end, the temperature-induced changes of wide-line ^{1}H-NMR spectra are studied for frozen and stepwise-heated aqueous protein solutions, focusing on the thermal effects on the interfacial regions of proteins and, particularly, the differences in the dynamical properties of interfacial water molecules. We attempt the rigorous descriptions of the thermodynamic processes (e.g. phase transitions) induced by changes in temperature and relate it to the static properties (e.g. architecture) of the proteins. The heterogeneity of binding properties makes this task difficult in a multiphase system, but revelational for quantifying "order and disorder" of proteins.

During freezing and melting the protein solutions, we observe the temperature at which atomic motions of water molecules in the interfacial region ("hydrate layer") of the protein become active, to be denoted as "melting start"[1] (MS). This start of mobility and the

---

[1] *footnote* The term "melting" means here the appearance of the motionally narrowed component in the NMR-spectrum. The motion involved should be fast enough on the NMR time-scale. The relative weight of the narrow spectrum component is plotted in the "melting diagram" as a function of temperature. Quotation marks were

following "melting process" (MP) were studied for several proteins in seeking a comprehensive interpretation of the whole process from the start of the mobility up to 0°C. To be more precise, our goal is the quantitative interpretation of the wide-line NMR spectra, especially the motionally-narrowed spectrum components measured in protein solutions. The phenomenon of motional narrowing plays a key role in NMR spectrometry [14], because it distinguishes rigid and mobile nuclear-spin ensembles and therefore it gives the physical basis for the interpretation of the "melting diagrams" (MD).

The fundamental goal of our work is to interpret the MDs in terms of measurable quantities. We deal with the reinterpretation of such results in literature, which were achieved when the potential of the MDs were yet not foreseeable. We attempt here to show the road taken to arrive at the present state. On this road, such ideas were made use of (e.g. fundamental temperature, or Waugh and Fedin's working hypothesis; see details in section "4 Small steps..."), which have already been known in other fields, but we have for the first time applied and even corrected them in this field. Furthermore, new quantities have been suggested such as order parameters, which characterize the heterogeneity of the potential barrier, or the derived MD to describe the "melting rate" of the interface. We believe that this study comprises not a historical description only, but perspective hints also, which point beyond past and present. Such hints are e.g. the possibilities to study molecular interactions, the critical value for the salt content of solutions. In the case of some particular proteins, these hints can be considered as not fully beaten but promising paths besides the already covered ones. Briefly, the following sequence of the chapters lead us from the old-style melting diagrams and the qualitative conclusions deduced to the new diagrams and their quantitative interpretations.

We suggest that the extent of heterogeneity in the measurable physical-chemical properties can help classify proteins, and the introduction of an order parameter ranging from zero to one provides a meaningful and continuous metric of order and disorder.

---

used because the only types of motion to be accounted for are libration and rotation, in contrast with the classical melting (solid to liquid phase transition) which results in both translational and rotational motions in the liquid phase. To put it more precisely, the vectors connecting two protons are in motion and their angles with the static magnetic field change, but the proton-proton distances remain constant. The detailed formalism can be found in refs. [12,14].

2 Water molecules map the interfacial space

The proton-proton spin-pairs located in the phases of bulk ice, protein, and interface can be separated by virtue of large differences in their mobility. These differences result in components of the free induction decay (fid) NMR signal characterized by different decay constants. The fid signal is the basis of the method applied to measure the fraction of the mobile water component at $T < 0°C$ (Fig. 1). The melting of the hydration shell can be tracked on this basis and it can be distinguished from the melting of bulk ice. By the phrase "hydration shell", we mean water molecules in the protein-water interface, whereas "melting" means the start of their reorientation. The water phase, which melts at about 0°C, is termed here as bulk water and bound water is the water phase, which melts well below 0°C (see also [15]). The water molecules in the bound phase interact with the protein molecules and with several other water molecules too. This separation of water into two phases is quite unambiguous at subzero temperatures, when the two phases are thermodynamically/ physically separated, which infers that they can also be separated experimentally.

Details of the experiments, including protein preparation and NMR experiment set up can be found for each protein in the references cited. A more extensive description of and also a practical guide to the NMR experiments can be found in chapter 12, ref. [8] and chapter 13, ref. [9], respectively.
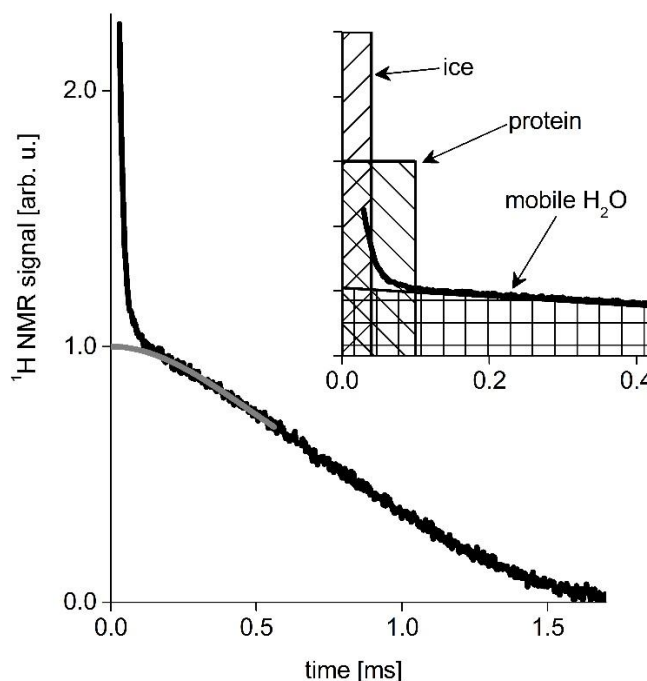
Figure 1. Illustration of the method applied to measure the fraction of the mobile water component. The slowly decaying part of the fid signal is extrapolated to $t = 0$ (by a stretched exponential, narrow line). The extrapolated signal intensity is then normalized to the signal intensity measured above 0°C where the whole sample is in liquid state. Inset: typical time-spreads of fid signals produced by ^{1}H nuclei in ice, in protein and in mobile water.

3 Melting diagrams of old fashion and qualitative conclusions

Our earlier results outline the evolution of approaches toward quantitative descriptions of protein hydration. In one of our first studies on IDPs (calpastatin and microtubule-associated protein 2 [16]), we demonstrated that the more open and largely solvent-exposed character a protein has, the larger the amount of water in its hydrate layer is. We have also investigated the interactions between the protein molecules and Na^+ and Cl^- ions in solution (BSA and ERD10) [17]. We compared and related these findings to similar studies on globular Ubiqitin [18], further IDPs, like ERD14 and p53TAD, and members of the casein family (unpublished results). These latter examples are in full agreement with the general conclusions drawn. The reasons of selecting the particular proteins, and the details of the applied experimental methods were outlined earlier [16-18,8,9]. It is important to note that the NMR measurements were done in the direction of heating, to avoid the problems caused by hysteresis.

The earliest investigations were carried out on buffered protein solutions of physiological NaCl concentration (usually Tris-buffer: 150 mM NaCl, 50 mM Tris, 1 mM EDTA, pH = 7.4). It was found that the "melting diagram" (MD), which represents the MP of a particular protein solution, depends on the protein-to-ion concentration ratio [17,18] and this called for the investigation of solutions prepared in distilled water solely (see e.g. human nuclear phosphoprotein p140 (hNopp 140) [19]).

We could provide evidence that the protein Df31 (*Drosophila melanogaster* histone chaperone involved in chromatin decondensation and stabilization) is intrinsically disordered along its entire length [20], the entire proline-rich region of CASK-interactive protein1 is intrinsically disordered [21], and that the results depend also on the protein structure (primary to quaternary level) [22] and its propensity to form a complex [23]. It was also found that a living organ, the eye lens, which contains various kinds of crystallins [24] in high concentrations, has characteristics similar to the protein solutions concerning the "non-

freezing water phase" [25]. The examples given not only report on properties of individual proteins, but also illustrate the new phenomenon revealed by the applied measurement methods.

Five typical MDs selected from our earlier works demonstrate that the mere visual comparison of the diagrams can reveal remarkable features (Fig. 2). (The mobile water fractions presented here in the melting diagrams are normalized to the protein concentration 50 mg/ml. The actual protein concentrations in the measured samples are given in Table 1).

One representative example for the globular and the ID proteins are compared in Fig. 2. BSA is used here as a globular standard (circles, Fig. 2) and ERD10 is an IDP (stars, Fig. 2) [18]; the MD of double distilled water is shown as the universal standard (squares, Fig. 2A). The effects Na^+ and Cl^- ions make on the hydration properties can be observed (Fig. 2B) as the difference relative to the MDs of the same proteins dissolved without the ions (Fig. 2A).
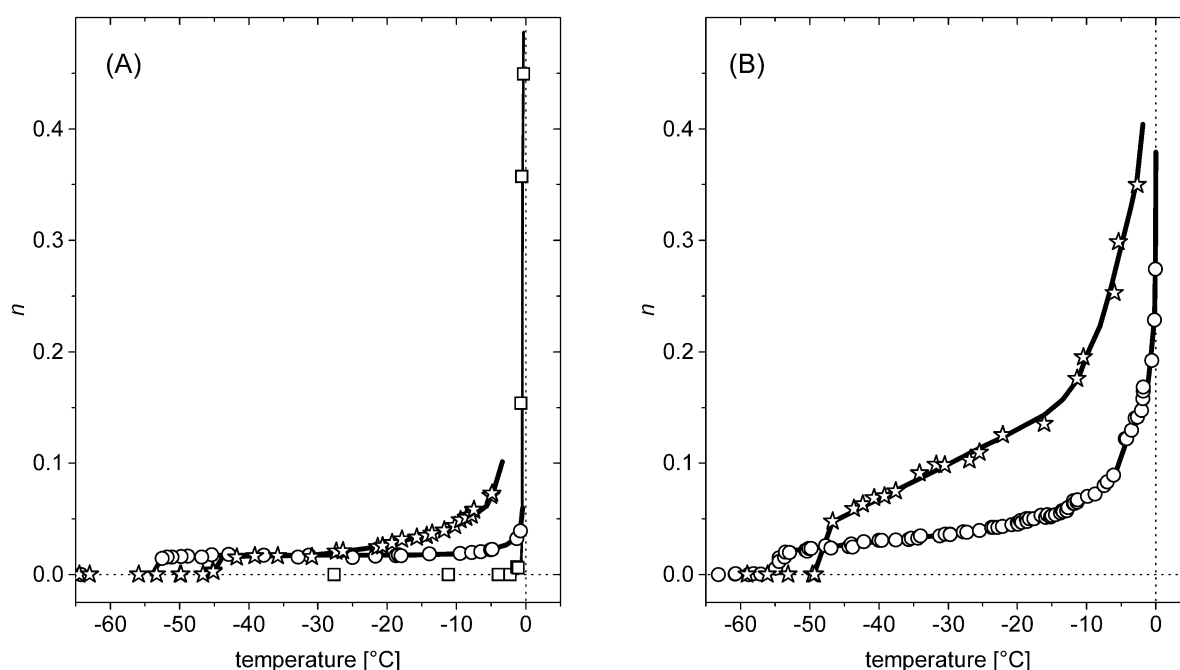

Figure 2. Comparison of the temperature dependencies of mobile water fraction n as measured by ^1H NMR for the globular protein BSA (circles) and the intrinsically disordered protein ERD10 (stars) dissolved in (A) double distilled water and (B) in Tris-buffer [18]. The mobile water fraction measured for double-distilled water is also shown (squares, panel A). The lines are guides to the eye.

In these and similar MD studies, the following observations were made and qualitative conclusions drawn:

- The (temperature dependent) number of the mobile water molecules in the interfacial region can be determined by analyzing MDs.
- The beginning of the motion of the water molecules (MS), which interact with the protein as first neighbors and the whole MP show a varied picture. A few common characteristic features can be found, however. The onset of the molecular motions can be considered as a melting-type transition. Certain motional processes start at significantly lower temperatures in the interfacial "ice" than in the bulk ice (0°C). This is a general finding for every protein studied to date.
- The attributes of MDs measured on globular and ID proteins show important differences: globular proteins have lower MS, and an MP almost independent of temperature, while the MD of IDPs is highly dependent on the temperature.
- In the case of proteins in solutions containing NaCl, the properties definitely shift towards "disorder" relative to salt-free solutions. The combination of the NMR and DSC results highlighted protein-ion interactions [17,18].
- More information on the dynamics of water molecules, mentioned here only marginally, is also clearly apparent. The wide-line NMR spectrum of interfacial water molecules becomes narrower as the reorientation of the molecules accelerates with increasing temperature [18]. As a result, the mobility (*HM*) of the hydrogen-hydrogen (proton-proton) nuclear dipole-pairs increases and contributes to the specific heat [18].
- The rotational and translational diffusion of water molecules can be studied as a function of temperature by looking into the actual presence and type of NMR echo-signals [25]. Information on the characters of the homogeneous or heterogeneous local hyperfine-field contributions of magnetic origin can also be gained this way.

Our results also emphasize the necessity of precision in setting experimental conditions and of high accuracy in measurements. *The question is open whether the molecular "surface" or the "structure" plays the more important role in the biological/physiological reactions. In other words, should we rather search the answer to the order/disorder issue by approaching the question from the homo- or heterogeneity of the potential energy distribution of the molecular surface?* Certain physical properties also deserve special interest. Such candidates are e.g. the

density of the interfacial phase in comparison to bulk water [26], and the electric and thermal properties of the hydration shell.

4 Small steps (milestones?) on the way of quantitative interpretation and a working hypothesis

As suggested above, we could only provide partial and incomplete interpretation of melting diagrams (MDs) in our previous publications, which has led to largely qualitative conclusions. *We find it necessary to make a move toward more quantitative interpretations, to reach a deeper understanding by introducing clearly defined and experimentally determined features, in a way similar to that achieved previously for the mobility/dynamics of hydrogen-hydrogen (proton-proton) vectors in solid (crystalline and lyophilized) protein samples* [12]. MDs shown in Figure 2A and 2B qualitatively differ for globular proteins and IDPs, which infers that particular "features" quantitatively defining order and disorder could be defined based on the MDs.

First we address the seemingly naive question if it is advisable to present and interpret our results on a Celsius temperature scale? It seems generally accepted that thermodynamics of water – and life itself – manifest themselves on the Celsius scale. In consequence, on the horizontal axis of our coordinate system we traditionally put not a dimensionless number but a dimensional physical quantity, i.e. degrees in Celsius. On the vertical axis, however, the line-narrowed NMR fid amplitude values are presented on a relative scale of simple numbers, showing what portion of water molecules are partially bound to the protein molecules. More precisely, we indicate what portion of water molecules are in certain potential energy wells generated partly by protein and partly by the neighboring water molecules (hindering the water movement) in the interfacial region. These water molecules can be termed "bound" [16]. In advance, these water molecules are bound weaker than those in bulk water, i.e., they have a lower melting temperature, between $-40\ldots-70^{\circ}$C This environment is principally different from the potential field generated by bulk water molecules in pure water, defined by H-O$\cdots$H bonds of tetrahedral geometry.

This also means that no matter what approach of series expansion we apply for the mathematical description of MDs, the parameters obtained by fitting will have dimensions. In

addition, the parameters of different order will have different dimension, which renders them directly incomparable. Our impression is that life scientists cannot operate with these "amplitudes", especially if they try to interpret them as physical-chemical quantities.

In accord, it is desirable to present a dimensionless (normalized) physical quantity on the horizontal axis too. In their book, Thermal Physics, C. Kittel and H. Kroemer [27] gave the possible temperature scales and various approaches of describing and measuring temperature. It appears that for our purpose the use of *fundamental temperature*[2], an energy/temperature scale defined in energy units, is the most expedient. In this scale, by definition

$$T_f = k_B T \qquad (1)$$

where, $k_B = 1.381 \cdot 10^{-23}$ J K^{-1} is the Boltzmann constant, and T is absolute temperature in K. In molar units

$$T_f = RT \qquad (2)$$

where $R = 8.314$ J K^{-1} mol^{-1} is the universal gas constant.

These scales are useful if we think in terms of thermal activation energy units. On these energy scales, the melting temperature(s) of the interface (MS) or particular points of MD one can relate to the melting point of bulk ice (water), thus introducing the T_{fn} *normalized fundamental temperature* as

$$T_{fn} = k_B T / (k_B \cdot 273.15 \text{ K}) = RT / (R \cdot 273.15 \text{ K}) = T / 273.15 \text{ K} \qquad (3)$$

On this relative T_{fn} scale, the melting point of bulk ice (water) is 1, whereas that of the interfacial ice (water) is a positive number that falls between 0 and 1. If we apply these dimensionless axes, and work on relative NMR amplitude and normalized fundamental temperature scales, the fitting of MDs can be achieved correctly, for example with the polynomial series

$$n = A + BT_{fn} + CT_{fn}^2 + \ldots \qquad (4a)$$

or for practical point of view as

---

[2] *footnote* We will not use the usual symbol τ, because in NMR spectroscopy it is generally used to denote correlation time.

$$n = A + B \cdot (T_{fn} - T_{fno}) + C \cdot (T_{fn} - T_{fne})^2 + \ldots , \qquad (4b)$$

where the parameters T_{fno} and T_{fne} are the lower limits for the first and second order terms to be active (we shall give their physical meanings in Section 6).

In this representation, n is the mobile water (bound to protein) content, and the amplitudes A, B, and C are (dimensionless) numbers, i.e. they can be directly compared. Their relative value can relate either to the total water content of sample or to the total interfacial water.

As a next step some further ideas are cited from literature before we try to provide a quantitative evaluation and interpretation of the experimental results: J. Waugh and E. Fedin outlined and critically analyzed the relationship between NMR spectral motional narrowing and the potential barrier (V_0) inhibiting the rotation or/and torsion-oscillation of molecules or molecular groups in solid state [28]. Melting is the feature most relevant with respect to the relationship of starting of rotational movement and the height of the potential barrier. They applied their concepts to solid state materials composed of simple species (e.g. NH_4Cl, NH_4Br, NH_4I, and adamantane), for which the temperature (T_c) where rotation of the NH_4^+ ion or the entire adamantane molecule starts is connected to the height of the energy barrier as

$$V_0 \, [\text{kcal/mol}] \approx 37 \cdot T_c \, [\text{K}] \qquad (5)$$

We will use their original formula in a corrected form, because the original one is dimensionally incorrect. The proportionality factor (37) is not dimensionless, because it relates two quantities of different dimensions (kcal/mol and K, i.e. its dimension is truly identical to that of the universal gas constant!).

*As **a working hypothesis**, we use the model connecting the height of potential energy barrier and the energy of thermal activation. We intend to use it for describing the melting of the interfacial ice, i.e. the mobilization of water molecules bound to the protein.* We assume and use the following relationships corresponding to Eq. (1)

$$V_0 \, [\text{erg}] \approx c T_f \, [\text{erg}] \qquad (6a)$$

or, in molar units, see Eq. (2)

$$V_0 \text{ [kJ/mol]} \approx cT_f \text{ [kJ/mol]}, \qquad\qquad (6b)$$

where the quantity c is dimensionless, probably a parameter that depends on the number (degree) of motional (rotational) freedom of water molecules.

The value of c can be determined by applying Eq. (6b) to bulk ice/water and considering the molar heat of fusion of water, which is 6.02 kJ/mol (= 1.439 kcal/mol, or the specific heat of fusion is 334 kJ/kg) [29]. This is the thermal activation energy required to overcome the potential energy barrier that hinders the movement of water molecules in bulk ice. The fundamental temperature T_f corresponding to 273.15 K is 2.271 kJ/mol, which means that the proportionality factor in Eq. (6b) for our working hypothesis is $c = 2.65!$ [3]

If we wish to relate this result to one water molecule instead of one mole by Eq. (6a), and interpret the $T_f = k_B T$ fundamental temperature as the thermal activation energy corresponding to two degrees of freedom in accord with the equipartition theorem, Eq. (6a) then provides approximately 5.3 degrees of freedom per one molecular unit, which seems reasonable for a water molecule (or, even better, for a rigid electric dipole, that is, for the accepted electric model of a water molecule).

It is easier to address certain issues if we relate the rotational energy barriers in the interface to those in the bulk water: this way the value related to the etalon will be proportional to T_{fn}. By our result thus far, the potential energy barrier of the interface is smaller, differing only by 20-25% from that in the bulk. One might ask how it relates to the energy of H···O-H bonds in bulk water. Clearly, the melting energy barrier of 6.02 kJ/mol and the energies reduced by protein binding give a reasonable energy scale, and they are significantly smaller than the energy required to break four H-O···H bonds upon evaporation of bulk water.

As far as the character of weak interactions, the mathematical formalisms used, and thoughts that arise with regards to the order of magnitude of potential barriers related to water solutions are concerned, it is probably not superfluous to remind the reader to a few handbooks and

---

[3] *footnote* Eq. (6a), apart from the value of parameter c, corresponds to basic statement of statistical physics, which links the thermal energy of one atom of an ideal gas with $k_B T$ (that is with the fundamental temperature), considering the proportionality factor 3/2 corresponding to three degrees of freedom.

works. For the H-bond, the work of M. Chaplin [29], for weak interactions in general, the book *The molecular biology of gene* by J. D. Watson [30], for the theoretical description of possible interactions the book *Elements of physical chemistry* by Atkins and de Paula [31], and *"Van der Waals Forces*: *A handbook for biologists, chemists, engineers and physicists"* by V.A. Parsegian [32], finally the *Molecular dynamics of water and the protein-solvent interfaces* [34] are worth to mention. These works help to oversee the present status quo and to lay down future tasks.

Finally, the comparison of MS and MD with the "solvent-driven dynamic transition at 180-220 K (see for instance Fig. 3 in [33] Fig. 4 in [34], and Fig. 3 in [35]) *is an "invitation for a dance".* In addition, it is critical to compare the wide-line spectral motional narrowing, i.e. *HM* parameters [14], which describe the movement of proton-proton radial vectors. All these works point to the abrupt change in the molecular-dynamic characteristics of protein molecules within the narrow temperature region of 200-220 K. In all our measurements, the beginning of MS falls in this region, which infers a clear correlation between the change of protein dynamics and the melting of interfacial ice (water). If this link is proven to be correct, one might characterize protein dynamics within the given region through studying hydration water melting.

Next, we will focus on the interaction of protein and water molecules, through the results of homogeneity/heterogeneity of respective potentials and their energy distributions. We will show their unique character and the differences between globular and disordered proteins.

5 Melting diagram of new form for protein-water solutions

We present MD for the pure (and buffered, in a few cases) water solution of some select proteins to show how the novel view outlined in Section 4 enables to reinterpret these observations. We present our results on the suggested T_{fn} temperature scale. We also give the scale of rotational energy barrier (based on 2.65 RT kJ/mol, eq. 6b) in the figures, providing support to our working hypothesis. The two scales enable the immediate assessment of the amplitude of the energy barrier that hinders the rotation of water molecules with thermal activation.
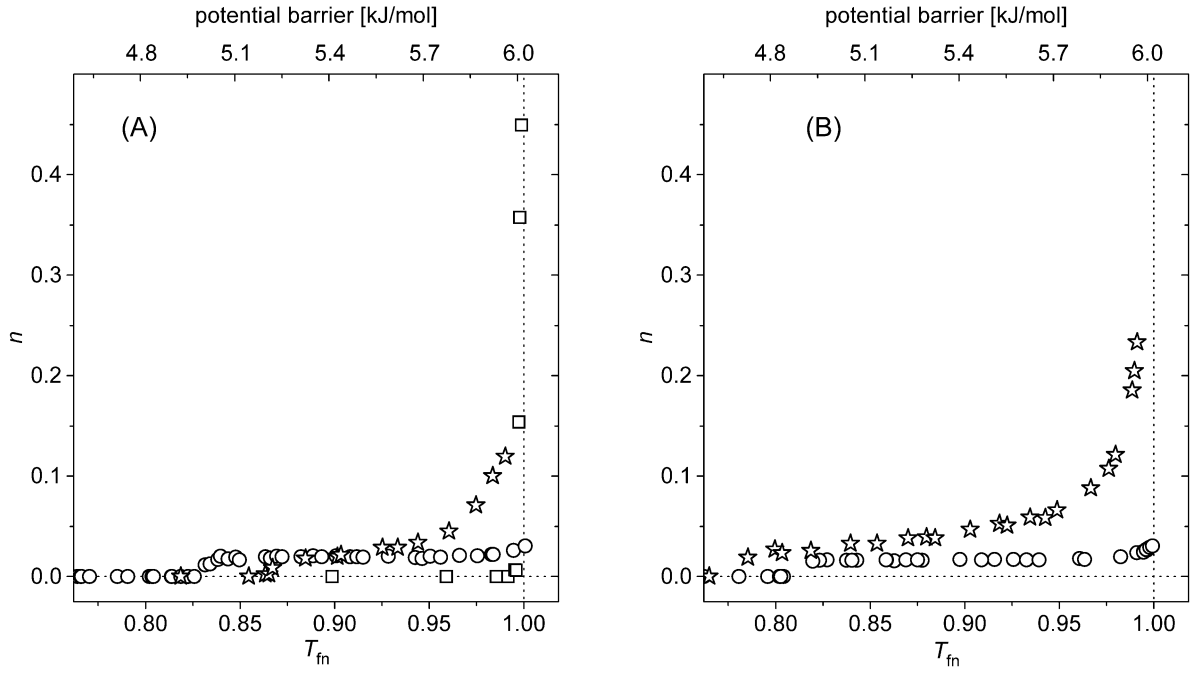
Figure 3. Comparison of the temperature dependencies of mobile water fraction n as measured by ^{1}H NMR. (A) The globular protein ubiquitin (circles), the intrinsically disordered wild-type α-synuclein (stars) dissolved in double distilled water and the mobile water fraction measured for double distilled water is also shown (squares, panel A). (B) β-casein dissolved in double distilled water (circles) and in Tris-buffer (stars). The protein concentrations were 41 mg/ml ubiquitin, 25 mg/ml wild-type α-synuclein, and 50 mg/ml β-casein. The mobile water fraction data were calculated for 50 mg/ml protein concentration in each case

In the figures, it is interesting to note that the melting of reference water (square MS points) is a few degrees below $T_{fn} = 1$, which might come from inaccuracies of setting and recording the temperature, the inhomogeneity of sample temperature, contaminations of the sample (e.g. dissolved oxygen), and the large relative surface of the small sample. Apparently, the local surrounding of water molecules in this environment differs from that in pure bulk water.

In accord, we relate the melting that occurs immediately below $T_{fn} = 1$ in our protein containing samples also to the contamination of our sample and the effect of its surface. Therefore, we leave these MD points of protein solution samples out of consideration. In the MS initial phase, the MD curves reach within a few degrees (an energy range of ~0.2 kJ/mol) the value associated with hydration by classical chemistry. In Figure 3A, it is clear that on the T_{fn} scale this value is ~0.02. One might assume that this value carries valuable information on the heterogeneity of the potential energy distribution of the first hydration shell. We should not forget the experimental problems referred to in the previous section, which will have to be solved in the future. Until this is done, we will describe this melting with a single step, we

consider the potential energy barrier homogeneous, and assume that the number of water molecules, which start moving in this region is constant (parameter A in polynomial fitting).

Making these caveats, the analysis and interpretation of the rest of the MD of UBQ is rather straightforward. Water in the interface melts at a certain temperature (MS ~ T_{fn} = 0.81-0.83 [4]), and from here the <u>number of mobile water molecules does not change in a broad range, up to T_{fn} = 0.94</u> (unlike their mobility, as demonstrated already [16, 27]). It is to be emphasized <u>that the MD "plateau" indicates that no breakable water bonds exist within this energy range (T_{fn} = 0.83 - 0.94)</u>. That is, the approximate activation energy range 5.04-5.7 kJ/mol is excluded in the spectrum accessible for water molecules. For an ideal globular protein, the entire molecule should behave like this. In the current case, further water molecules start moving in the range T_{fn} = 0.94 - 0.99, in increasing numbers as temperature rises. It seems that part of the molecules behave different: they appear as disordered. What fraction shows this irregular behavior? Let's assume (as a Delphi forecast) it is about 23% (which will be rationalized later). In advance, we can state that each globular protein studied displays this irregular behavior.

The water solution of α-synuclein confirms the points raised with regards to UBQ (Figure 3A), which shows a basically different behavior. Here, the number of moving water molecules increases steadily from MS on, without reaching a plateau. Each different thermal activation energy value can move a new fraction in the interfacial population, i.e., melting (and the corresponding potential energy surface) is heterogeneous. The whole molecule shows "irregular" behavior. By its suggested ID character, its "disorder" spreads over the entire molecule (the Delphi forecast to its heterogeneity is approximately 100%). The MS behavior of the two molecules is clearly distinct, the hydration shell of the globular protein melts at a lower temperature (T_{fn} = 0.83), whereas that of the ID/IU protein is closer to the bulk water (T_{fn} = 0.88)

In Figure 3B, we show the melting curve of β-casein, which is a more ordered protein in water (circles) than in buffer of 150 mM NaCl (stars) The NaCl co-solute makes the protein display a "surface" behavior more similar to that of disordered proteins, and its MD shows similarities to that of α-synuclein (Figure 3A). The number of mobile water molecules

---

[4] *footnote: T_{fn} ~ MS, the proportionality is meant to express that whereas T_{fn} means a single temperature, MS is a narrow temperature (activation energy) range, and knowing the MD enables to map the motional energy barrier distribution.*

measured in a protein solution cannot be reproduced as a sum of the MD measured for the protein-free Tris-buffer and the MD of the protein dissolved in pure water [17]. The melting of water molecules occurs (MS) at T_{fn} = 0.81-0.82, a plateau can be detected from T_{fn} = 0.82 to T_{fn} = 0.97, where the MD changes into an increasing phase. The "forecast" gives a value of 14% for the irregular portion of the protein. When dissolved in a buffer solution, the protein interacting with co-solutes makes the curve rise constantly, and the MD can be fit the best with an initial linear function followed by a quadratic function. The initial temperature of the two states, here that of the "irregular" phase becomes lower, the whole molecule changes significantly due to the co-solutes.

In Figure 4A, we compare the MD diagram of the wild-type and the E46K α-synuclein (published [18] and unpublished data). It is to be mentioned that two other mutants have also been studied, the A53T variant behaves similar to the WT, whereas A30P looks more like E46K.

The two α-synuclein variants in Fig. 4A show the typical MD behavior of IDPs. Both have an MS of high temperature (T_{fn} = 0.87) and MDs of somewhat different slope, with WT being steeper. It is to be emphasized that wide-line NMR spectrometry can distinguish point-mutants based on the binding heterogeneity of the protein surface, similarly to our prior measurements on lyophilized samples based on proton-proton vector mobility [13]. We will return later to the issue of the quantitative definition of the "slope".

We compare the monomeric and amyloid state of WT α-synuclein in Figure 4B. The two MDs show clear differences in the initial melting temperature (T_{fn} = 0.87 for the amyloid, in contrast with 0.82 for the monomer state), which corresponds to a less heterogeneous energy distribution of hydrate water molecules in the case of the amyloid.
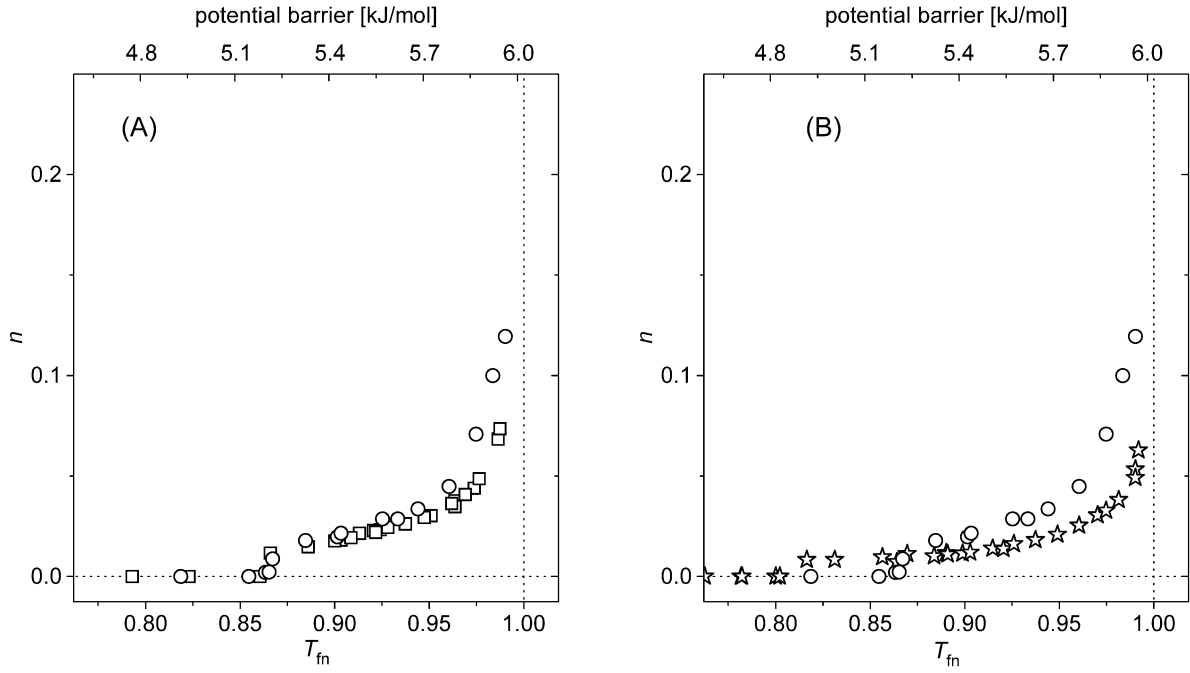
Figure 4. Comparison of the temperature dependencies of mobile water fraction n as measured by ^{1}H NMR of (A) the E46K (squares) and the WT (circles) α-synuclein variants dissolved in double distilled water, (B) the monomer (circles) and amyloid (stars) WT α-synuclein dissolved in double distilled water. The protein concentrations were 25 mg/ml. The mobile water fraction data were calculated for 50 mg/ml protein concentration in each case.

6 The ratio and measure of water-protein binding heterogeneity

The MDs shown in Sections 3 and 5 provide the number of water molecules that start moving at a given temperature, and help distinguish regions which do not change, change slowly or faster as a function of temperature. By observing the beginning and slope of these characteristic regions, we can try to estimate the magnitude (ratio) of homogeneous (non-changing) and heterogeneous (changing) populations.

As an example, let's consider the first MD measurements [16] on the buffered solution of the fragment PRD2-His (the proline-rich region of Caskin1 protein) (Fig. 5). We have the most complete results with this protein. We measured the number of moving water molecules at several temperature values and fitted it with a polynomial Eq. (4). The variable A is a temperature-independent factor, which characterizes the homogeneous component, whereas B and C characterize the number of water molecules connected to heterogeneous potential surface regions. The fitted parameter values are collected in Table 1.

The <u>number of water molecules which just start moving at a given thermal energy</u> can derived from the rate of the change (derivative curve, Fig. 5B) in the MD. The amount of the water molecules, which just get on the move at the actual temperature, relates to the bonds broken at the given thermal energy value. So, the number and distribution of energy barriers can be obtained. If the MD can be approached by an analytical function, then the rate of change is given by the derivative, i.e. by the differential quotient dn/dT_{fn}

$$dn/dT_{fn} = B + 2C\, T_{fn} + \ldots \tag{7}$$

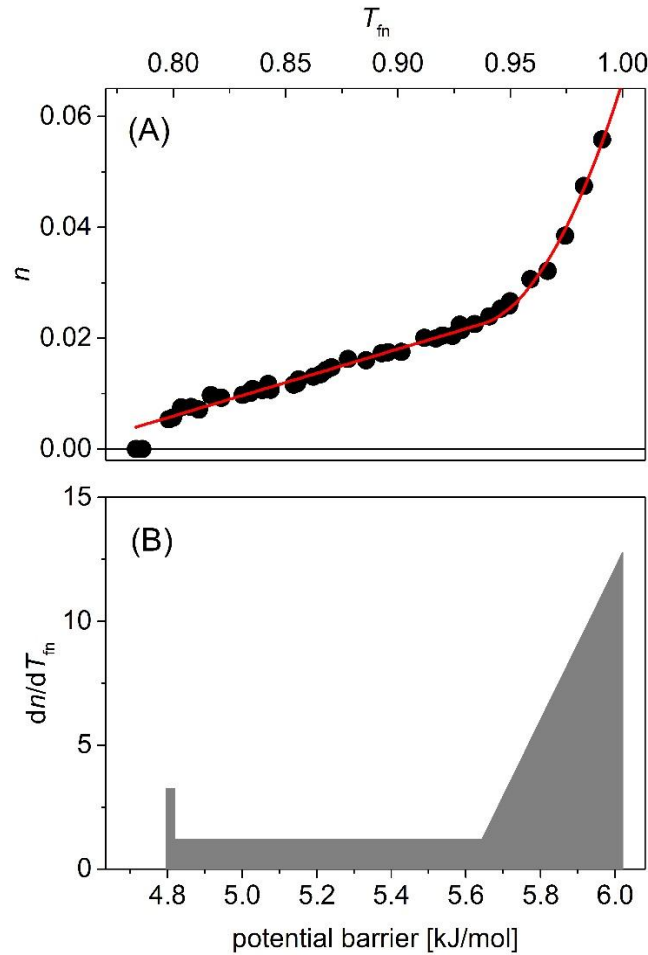or else the difference quotient $(\Delta n/\Delta T_{fn})$ gives the results.

Figure 5. (A) Mobile water fraction as a function (circles) of normalized fundamental temperature fitted with Eq. (4) (line) and (B) its differential quotient (number of water molecules which just start moving) for the fragment PRD2-His of the proline-rich region of cascin1 protein dissolved in Tris-buffer. The protein concentration was 25 mg/ml. The mobile water fraction data (A) were calculated for 50 mg/ml protein concentration.

We define the quantities required to determine the ratio of heterogeneity to homogeneity (*HeR*) as follows. Let T_{fno} be the normalized fundamental melting temperature of a **ho**mogeneous protein region (surface), and T_{fne} that of a heterogeneous protein region. We suggest that the following definition of the ratio of heterogeneous binding interface is a useful measure of the structural disorder of proteins.

$$HeR = \frac{1 - T_{\text{fne}}}{1 - T_{\text{fno}}} \qquad (8)$$

The numerator and denominator in Eq. (8)[5] denotes the possible energy range (or the lower limit of the energy range), which is characteristic of disordered and ordered (globular) protein molecules. Disorder corresponds to heterogeneous energy interface and ordered means homogeneous energy interface from this perspective. This way we relate the water-protein interfaces to the potential energy barriers normalized to that of bulk water. If the two normalized fundamental temperature values are equal (*HeR* = 1), the protein has heterogeneous energy surfaces only and it can be assumed to be completely disordered. If the two values differ (*HeR* is between 0 and 1), the protein molecule is energetically partly homogeneous and partly heterogeneous.

This relationship provides the foundation of the heterogeneity values suggested in Chapter 5 (23% in the case of UBQ, and 14% in the case of β-casein). All variants of α-synuclein and all three fragments of the proline-rich region of Caskin1 are approximately 100% heterogeneous, i.e. they are fully disordered by the definition used.

To return to the results shown in Figs. 2A and 2B, water itself presents a homogeneous H-O···H binding structure (with the exception of the narrow range below 0°C), globular BSA presents smaller interface energy barriers, which are homogeneous to a large extent (72%), heterogeneous to a smaller extent (28%). A fully heterogeneous binding energy surface distribution characterizes ERD10 in water, and both ERD10 and BSA in buffer. This fully heterogeneous behavior can be termed as an IDP, because, by our working hypothesis, fully

---

[5] *footnote:* Eq.(8) is equivalent with the formula $HeR = \dfrac{273.15 - T_{\text{e}}}{273.15 - T_{\text{o}}}$, where $T_{\text{e,o}}$ denote absolute temperature, but it is not clear in Eq. (8) that it is used for comparing binding energy regions, i.e. it only helps calculations.

disordered proteins have a completely heterogeneous potential energy interface. It appears that the presence of buffer, or rather the sodium and chloride ions in it, makes the binding energy distribution of all proteins more heterogeneous.

One needs to take a closer look at the parameters of the polynomial fitting function to answer the following questions. Why do we use the fundamental temperature scale only for categorization, and why is the ordinate made use of only to the extent that to check if its value changes with temperature. By looking at the curves of WT and E46K α-synuclein in Figure 4A and 4B, and also the differences of monomeric and amyloid states, it is clear that the melting temperatures themselves do not provide all the information on IDPs and their parts. In other words, the melting temperature does not give quantitative information on how many water molecules (what part of the interfacial region) melt at the given thermal activation energy level. This number depends on several things (apart from trivial protein concentration, which can be eliminated by normalization), such as the primary structure of the protein, the form of surface potential energy (known only at model level) and its dependence on distance, etc. It appears that the details of the way in which the measured hydration depends on melting temperature is important from this point of view. The measured and analyzed MD curves provide a unique opportunity to map the potential energy interface of proteins.

Coming back to WT α-synuclein and its point mutants, they only differ in a single amino acid and are disordered in the monomeric state, yet their MDs differ and their fits deviate in the factor C of the second order term in Eqs. (4).

Next, we suggest to introduce a single "parameter" to describe heterogeneity, which might characterize nearly-full "irregular" proteins (HeR ≈ 1) and the irregular part of the "globular" ones. Homogeneity can be associated with the constant (zero exponent) A term in the Eqs. (4), whereas heterogeneity can be characterized with the sum of the B and C terms (non-zero exponent). More correctly, we use the sum of the derivatives (Eq. 5) $B + 2C$, which can be done because B and C are dimensionless due to having introduced fundamental temperature and normalization to the melting of bulk ice (water). We also demand from the heterogeneity parameter that it represents energy regions of hydrogen bonds formed between the protein and the water molecules, which are close to the H···OH energies present in bulk water. We suggest the following parameter as the measure of heterogeneity.

$$HeM = (B + 2C)/(1 - T_{\text{fne}}). \tag{9}$$

The dimensionality of the formula is OK, and takes a value of zero for ideal (equipotential) proteins ($B + 2C = 0$), and $1 - T_{\text{fne}}$ marks the energy range within which $B + 2C$ pieces of heterogeneous protein-water bonds are broken. *HeM* can serve as a measure of binding heterogeneity (which could be related to structural disorder), but that is not an "order parameter" in traditional sense, because the upper limit (namely the 1) is missing at present . *HeM* redefines the value we tried to reach by referring to the slope of MD previously in [18], without success.

Select examples from our results and the novel concept are demonstrated through a globular and a disordered protein in Fig. 6. The preliminary results are summarized for most of our experimental results in Table 1. These were evaluated and are interpreted in the proposed way. Parameter *A* gives the amount of water molecules bound to homogeneous surficial protein residues and T_{fno} is the melting point of this water fraction. Parameters *B* and *C* describe the fraction of the water molecules bound to heterogeneous surficial protein residues and T_{fne} is the melting point of this water fraction. The heterogeneity ratio is calculated as *HeR* = $(1 - T_{\text{fne}})/(1 - T_{\text{fno}})$ and the heterogeneity measure is *HeM* = $(B + 2C)/(1 - T_{\text{fne}})$. The details of the work on the mentioned protein families are to be published later.
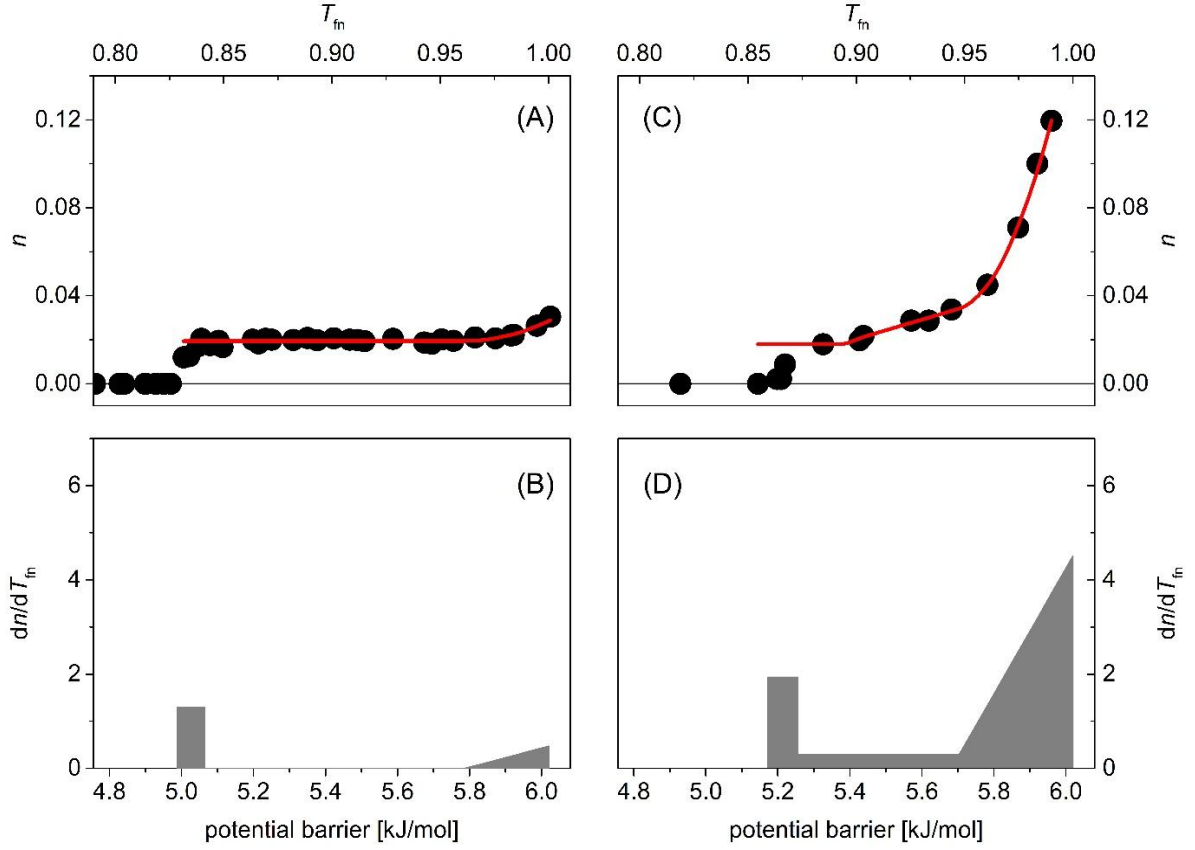
Figure 6. (A,C) Mobile water fraction (circles) as a function of normalized temperature fitted with Eq. (4) (line) and (B,D) its differential quotient (number of water molecules which just start moving), for the proteins ubiquitin (A,B) and wild-type α-synuclein (C,D) dissolved in water. The protein concentrations were 41 mg/ml ubiquitin and 25 mg/ml wild-type α-synuclein. The mobile water data were calculated for 50 mg/ml protein concentration in each case.

Table 1. Measured and fitted parameter values used in the evaluation of MDs (Eqs. 8 and 9). The values of the parameters A, B and C are given for 50 mg/ml protein concentration. The star denotes globular or ordered nature.

protein, its concentration, solvent	A	B	T_{fno}	T_{fne}	HeR	$B+2C$	HeM
ubiquitin 41 mg/ml in water (∗)	0.019	0	0.834	0.961	0.23	12	0.47
lysozyme 50 mg/ml in water (∗)	0.017	0	0.727	0.929	0.26	3.3	0.24
bovine serum albumin 50 mg/ml in water (∗)	0.017	0	0.780	0.938	0.28	5.7	0.36
β-casein 50 mg/ml in water (∗)	0.017	0	0.820	0.975	0.14	46	1.14

WT α-synuclein (oligomer) 50 mg/ml in water (∗)	0.018	0	0.808	0.935	0.34	15.9	1.03
WT α-synuclein 25 mg/ml in water	0.018	0.30	0.867	0.893	0.81	81	4.23
WT α-synuclein amyloid 25 mg/ml in water	0.008	0.08	0.816	0.850	0.80	7.4	0.67
E46K α-synuclein 40.5 mg/ml in water	0.012	0.24	0.866	0.876	0.93	40	1.97
thymosin-β4 25 mg/ml in water	0.015	0.08	0.852	0.870	0.91	13	0.84
stabilin-2 CTD 25 mg/ml in water	0.011	0.14	0.806	0.809	0.99	12	1.19
thymosin-β4 + stabilin-2 CTD 25 mg/ml in water	0.017	0	0.863	0.910	0.66	19	1.67
β-casein 50 mg/ml in buffer	0.019	0.24	0.785	0.795	0.95	20	1.70
caskin1 PRD1, 25 mg/ml in Tris-buffer	0.063	0.54	0.813	0.810	1.00	127	8.48
caskin1 PRD2, 25 mg/ml in Tris-buffer	0.055	1.21	0.798	0.796	1.00	186	11.6
caskin1 PRD3. 25 mg/ml in Tris-buffer	0.070	1.24	0.828	0.854	0.85	44	3.0
estimated error	0.001	0.02	0.002	0.002	0.02	20%	

7 Summary, Conclusions, Outlook

In our previous studies [16-23], we raised that a motional-narrowed wide-line NMR spectroscopic approach can be used for recording MD curves. The novelty of this study is to interpret the MD curves through energy potentials, which did not occur to us in the initial studies. In accord, the density of experimental points in prior works is often insufficient to carry out the present analysis, in particular in the initial (MS) phase of the curves. This section of the curve carries important information, because the energies of the water molecules, which start moving here, are the farthest from those of the water molecules in the bulk, engaged in tetrahedral H⋯OH bonds. Therefore, the water molecules, which fall in a very narrow energy range, are the direct embodiment of the "first" hydrate layer of the protein. Although we approach this region in the current work with a single energy value (a single A in Table 1), it deserves a close attention in future studies.

a) Globular proteins (rows 1-4, Table 1)

In the case of globular proteins, water in the interface melts at a certain temperature (MS ~ T_{fn} = 0.81-0.83), and the number of mobile water molecules then does not change up to as high as T_{fn} = 0.94 (unlike their mobility, as demonstrated already in ref. [18]). It is to be emphasized that the MD "plateau" means that no breakable water bonds exist in this energy range (T_{fn} = 0.83 - 0.94). The approximate activation energy range of 5.04-5.7 kJ/mole is excluded from the spectrum accessible for water molecules. We found that this behavior is characteristic of all globular proteins studied to date. If this type of MD applies to the entire protein, it appears of diagnostic value for the highly ordered globular state. In the current case, further water molecules start moving in the range of T_{fn} = 0.94-0.99. It seems, that a part of the molecules behave differently, they appear as disordered, i.e. all proteins show some tendency for bimodal distribution. This might be interpreted as intrinsically disordered segments of the proteins, but in the case of well-structured proteins, this behavior might indicate a tendency for local transient unfolding emerging from breathing motions of the protein. The heterogeneity-ratio *HeR* is 0.14-0.28 for the proteins in the first four rows in Table 1, and T_{fno} (~MS) is generally lower than that of IDPs.

b) WT and E46K α-synuclein (rows 5-8, Table 1)

WT α-synuclein (WT-aS) can assume structural states which are placed at far away regions on a virtual order to disorder scale. The oligomer behaves very similarly to the traditionally globular proteins, namely it has a plateau ($B = 0$) and *HeR* = 0.34. The MDs of the monomeric and the amyloid forms show markedly different properties compared to the oligomeric form. The monomeric WT-aS shows stronger and the amyloid has weaker ID properties. Accordingly, the monomer is of more heterogeneous bonding character than the amyloid or the (mostly ordered) oligomer. The differences between the MDs of E46K- and WT-aS demonstrate, that wide-line NMR is able to make clear distinction between point mutants (for the experimental details see [22]; detailed interpretation will be published separately).

c) Disordered proteins/states (rows 6-15, Table 1)

As globular proteins can have disordered parts (paragraph a), the opposite behavior is also apparent in our studies, because some disordered proteins show definite signs of structural order. This observation probably attests to the unique capacity of our approach in determining the tendency of IDPs for transient structure formation. It is speculated for many IDPs, that they transiently sample structured states in the disordered ensemble, and the particular regions

are involved in functional interactions with partner proteins or other macromolecules [36]. This might be the case with E46K-aS, which carries a mutation that promotes aggregation, probably by shifting the ensemble to more ordered states [37]. In accord, the aggregated state of WT-aS appears entirely ordered, which is in agreement with its folded, amyloid state. This state can be characterized by the absence of any "plateau", a small (0.08), nonzero B can be detected instead and a broad (HeR = 80%) slightly heterogeneous HeM = 0.67 character is dominant. For comparison, the monomeric WT-aS gives for HeR the same value, but B = 0.30 and HeM = 4.23.

d) Thymosin-β4, stabilin-2, and their complex (rows 9-11, Table 1)

As suggested in the Introduction, molecular interactions of proteins have to include changes in their hydrate layer by definition. A local dehydration must ensue to allow the physical contact of two proteins, which infers that the local binding potential of hydrate water molecules is directly related to the capacity of the protein to bind partners. This is formally demonstrated in the case of the interaction between T-β4 and stabilin. These proteins have been shown to be functionally related [38], and here we demonstrate that our approach can prove their interaction. As follows from molecular logic, the sum of hydrate water molecules of the two proteins is significantly larger than the hydrate layer of their complex, which is the first example [24] in the literature that wide-line NMR can detect and characterize molecular interactions.

e) The effect of solvent (rows 4-9and 12 Table 1)

In the case of proteins in solutions containing NaCl, the properties definitely shift towards "disorder" relative to salt-free solutions. The combination of the NMR and DSC results highlighted protein-ion interactions [17,18]. At the same time, it is reasonable to ask the optimal ratio of protein to salt concentration to have a single-phase system (i.e. to get rid of the NaCl eutectic phase in our case).

f) Proline-rich region of Caskin1 protein (last three rows, Table 1)

Segments of the proline-rich region of Caskin1 protein were among the first proteins we studied by wide-line NMR [16], at a time when the influence of buffers (ions) on the structural status of a protein was not yet expected. This, however, contributes to a higher level of disorder, and probably distinguishes between the three segments, albeit to an unknown

magnitude. On the other hand, these later findings stress the importance of experimental accuracy in subsequent studies (cf. Figure 5).

g) Summary and Outlook

The uniqueness of our novel method stems from its capacity to provide a quantitative description of the energetic landscape of the hydrate layer of protein molecules. Since local dehydration events are inescapably linked with protein-protein interactions, the function of protein critically depends on this feature. Since protein folding is also an interaction phenomenon (between parts of the same polypeptide chain), it is also intimately linked with hydration. In accord, our method is uniquely sensitive to the folding state of proteins, and it can provide a quantitative measure of the folding state (might be called disorder) of a protein. This enables to position a protein along the order/disorder continuum correctly, i.e. in a physically meaningful manner.

The quantitative order parameter developed in this work enables a rigorous treatment of the structural disorder phenomenon. Whereas it is agreed that structural disorder represents a continuum from fully ordered to fully disordered proteins, it has been a challenge to quantify its level in a model-free and experimentally defined way. Here, we provide a normalized and experimentally well-defined parameter, which seems to have the power to delineate important function-related features of (disordered) proteins. Because this measure derives form an actual energy scale, it has the potential to be directly related to the thermodynamic and kinetic features of protein function. It can give an energetic fingerprint of every protein molecules, which will enable one to explore proteins in a great variety of functional contexts. It is also descriptive of protein-protein interactions, a special case of which is amyloid formation in pathological processes. Importantly, these measures seem to be related to the actual behavior of proteins in live cells/tissues, as demonstrated by previous measurements on eye lenses, for example [25]. Alongside its time-effectiveness (a few hours is sufficient to map a protein), this makes the suggested technique uniquely effective in providing essential function-related details on the hydration behavior of proteins.

Finally, let us mention the short time span required for the measurement and evaluation procedure. A few hours is enough to map a protein, which shows promise for the investigation of processes and external effects too.

8. References

1 H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shingyalov, P.E. Bourne. The Protein Data Bank. Nucleic Acids Res. 28: 235-242. 2000. doi: 10.1093/nar/28.1.235

2 A.W. Fenton. Allostery: an illustrated definition for the 'second secret of life'. Trends Biochem. Sci. 33: 420-425. 2008. doi: 10.1016/j.tibs.2008.05.009

3 K.A. Dill and H.S. Chan. From Levinthal to pathways to funnels. Nat Struct Biol. 4: 10-19. 1997. doi:10.1038/nsb0197-10

4 H.N. Motlagh, J.O. Wrabl, J. Li, V.J. Hilser. The ensemble nature of allostery. Nature 508: 331-339. 2014. doi:10.1038/nature13001

5 M.F. Colombo, D.C. Rau, V.A. Parsegian. Protein solvation in allosteric regulation: a water effect on hemoglobin. Science 256: 655-659. 1992. doi: 10.1126/science.1585178

6 M. Tarek and D.J. Tobias. The dynamics of protein hydration water: a quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. Biophys J. 79: 3244-3257. 2000. doi: 10.1016/S0006-3495(00)76557-X

7 P. Tompa. Structure and Function of Intrinsically Disordered Proteins. Taylor and Francis Group, Boca Raton, 2010.

8 V.N. Uversky, Sonia Longhi, Eds. Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure And Conformation. John Wiley & Sons, 2010. doi: 10.1002/9780470602614

9 V.N. Uversky, K.A. Dunker, Eds. Intrinsically Disordered Protein Analysis. Book Series: Methods in Molecular Biology, Vol. 895. Humana Press, 2012.

10 P.W. Fenimore, H. Frauenfelder, S. Magazu, B.H. McMahon, F. Mezei, F. Migliardo, R. D. Young, I. Stroe. Concepts and problems in protein dynamics, Chem. Phys. 424: 2-6. 2013. doi: 10.1016/j.chemphys.2013.06.023

11 V.N. Uversky. A decade and a half of protein disorder: Biology still waits for physics. Protein Sci. 22: 693-724. 2013. doi: 10.1002/pro.2261

12 K. Tompa, M. Bokor, K.H. Han, P. Tompa. Hydrogen skeleton, mobility, and protein architecture. Intrinsically Disordered Proteins 1: 82-91. 2013. doi: 10.4161/idp.25767

13 T. Verebélyi, M. Bokor, P. Bánki and K. Tompa Hydrogen mobility in polycrystalline lysozyme. to be published in J. Magn. Reson.

14 C.P. Slichter. Principles of Magnetic Resonance. Springer Series in Solid-State Sciences Volume 1. 1990.

15 R. Cooke and I.D. Kuntz. The properties of water in biological systems. Ann. Rev. Biophys. Bioeng. 3: 95-124. 1974. doi: 10.1146/annurev.bb.03.060174.000523

16 V. Csizmók, M. Bokor, P. Bánki, É. Klement, K. F. Medzihradszky, P. Friedrich, K. Tompa, P. Tompa. Primary Contact Sites in Intrinsically Unstructured Proteins: The Case of Calpastatin and Microtubule-Associated Protein 2. Biochemistry. 44: 3955-3964. 2005. doi: 10.1021/bi047817f

17 P. Tompa, P. Bánki, M. Bokor, P. Kamasa, D. Kovács, G. Lasanda, K. Tompa, Protein-Water and Protein-Buffer Interactions in the Aqueous Solution of an Intrinsically Unstructured Plant Dehydrin: NMR Intensity and DSC Aspects, Biophys. J. 91: 2243-2249. 2006. doi: 10.1529/biophysj.106.084723

18 K. Tompa, P. Bánki, M. Bokor, P. Kamasa, G. Lasanda, P. Tompa, Interfacial Water at Protein Surfaces: Wide-Line NMR and DSC Characterization of Hydration in Ubiquitin Solutions, Biophys. J. 96: 2789-2798. 2009. doi: 10.1016/j.bpj.2008.11.038

19 A. Tantos, K. Szrnka, B. Szabo, M. Bokor, P. Kamasa, P. Matus, A. Bekesi, K. Tompa, K.-H. Han, P. Tompa. Structural disorder and local order of hNopp140. Biochim. Biophys. Acta. 1834: 342-350. 2013. doi: 10.1016/j.bbapap.2012.08.005

20 E. Szőllősi, M. Bokor, A. Bodor, A. Perczel, E. Klement, K. F. Medzihradszky, K. Tompa, P. Tompa. Intrinsic Structural Disorder of DF31, a Drosophila Protein of Chromatin Decondensation and Remodeling Activities. J. Proteome Res. 7: 2291-2299. 2008. doi: 10.1021/pr700720c

21 A. Balázs, V. Csizmók, L. Buday, M. Rakács, R. Kiss, M. Bokor, R. Udupa, K. Tompa, P. Tompa. High levels of structural disorder in scaffold proteins as exemplified by a novel neuronal protein, CASK-interactive protein1. FEBS J. 276: 3744-3756. 2009. doi:10.1111/j.1742-4658.2009.07090.x

22 E. Hazy, M. Bokor, L. Kalmar, A. Gelencser, P. Kamasa, K.H. Han, K. Tompa, P. Tompa. Distinct Hydration Properties of Wild-Type and Familial Point Mutant A53T of α-Synuclein Associated with Parkinson's Disease. Biophys. J, 101:2260-2266, 2011. doi: 10.1016/j.bpj.2011.08.052

23 A. Tantos, B. Szabo, A. Lang, Z. Varga, M. Tsylonok, M. Bokor, T. Verebelyi, P. Kamasa, K. Tompa, A. Perczel, L. Buday, S.H. Lee, Y. Choo, K.H. Han, P. Tompa. Multiple fuzzy interactions in the moonlighting function of thymosin-β4. Intrinsically Disordered Proteins. 1: 50-65. 2013. doi:10.4161/idp.26204

24 P.A. Wilmarth, J.R. Taube, M.A. Riviere, M.K. Duncan, L.L.David. Proteomic and sequence analysis of chicken lens crystallins reveals alternate splicing and translational forms of beta B2 and beta A2 crystallins. Invest. Ophthalmol. Vis. Sci. 45: 2705–2715. 2004. doi: 10.1167/iovs.04-0131

25 K. Tompa, P. Bánki, M. Bokor, P. Kamasa, P. Rácz, P. Tompa. Hydration water/interfacial water in crystalline lens. Exp. Eye Res. 91: 76-84. 2010. doi: 10.1016/j.exer.2010.04.005

26 F. Merzel, J.C. Smith. Is the first hydration shell of lysozyme of higher density than bulk water? Proc. Natl. Acad. Sci. USA 99: 5378-5363. 2002. doi: 10.1073/pnas.082335099

27 C. Kittel and H. Kroemer. Thermal Physics. W. H. Freeman, San Francisco. 1980. p. 445. Appendix B. Temperature Scales

28 J.S. Waugh and E.I. Fedin. On the Determination of Hindered Rotation Barriers in Solids, Fiz. Tverd. Tela (Leningrad). 4: 2233-2237. 1962. English translation Sov. Phys. Solid State. 4: 1633-1636. 1963.

29 M. Chaplin. Water's hydrogen bond strength. (2007) arXiv:0706.1355 [cond-mat.soft] http://arxiv.org/abs/arXiv:0706.1355

30 J.D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, R. Losick. Molecular Biology of the Gene. Cold Spring Harbor Laboratory Press; 7th edition (2013)

31 P. Atkins, J. de Paula. Physical Chemistry. 9th edition. Oxford University Press, Oxford. (2010)

32 V.A. Parsegian. Van der Waals Forces: A Handbook for Biologists, Chemists, Engineers, and Physicists. Cambridge University Press. New York (2006)

33 A.R. Bizzarri, S. Cannistraro. Molecular Dynamics of Water at the Protein-Solvent Interface. J. Phys. Chem. B, 106: 6617-6633. 2002. doi: 10.1021/jp020100m

34 J.C. Smith, F. Merzel, A.N. Bondar, A. Tournier, S. Fischer. Structure, dynamics and reactions of protein hydration water. Phil. Trans. R. Soc. Lond. B Biol Sci., 359: 1181-1190. 2004. doi: 10.1098/rstb.2004.1497

35 A.L. Tournier and J.C. Smith. Principal Components of the Protein Dynamical Transition. Phys. Rev. Lett. 91: 208106. 2003. doi: 10.1103/PhysRevLett.91.208106

36 P.E.. Wright, H.J.Dyson, Linking folding and binding. Curr Opin Struct Biol. 2009 19(1):31-38. doi: 10.1016/j.sbi.2008.12.003

37 V. N. Uversky, J. Li, A.L. Fink. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. J Biol Chem. 2001 276(14):10737-10744.

38 S.J. Lee, I.S. So, S.Y Park, I.S.Kim, Thymosin beta4 is involved in stabilin-2-mediated apoptotic cell engulfment. FEBS Lett. 2008 582(15):2161. doi: 10.1016/j.febslet.2008.03.058.