

SCIENTIFIC REPORTS



OPEN

Hydration shell differentiates folded and disordered states of a Trp-cage miniprotein, allowing characterization of structural heterogeneity by wide-line NMR measurements

Nóra Taricska¹, Mónika Bokor³, Dóra K. Menyhárd², Kálmán Tompa³ & András Perczel^{1,2}

Hydration properties of folded and unfolded/disordered miniproteins were monitored in frozen solutions by wide-line ¹H-NMR. The amount of mobile water as function of T ($-80^\circ\text{C} < T < 0^\circ\text{C}$) was found characteristically different for folded (TC5b), semi-folded (pH < 3, TCb5(H+)) and disordered (TC5b_N1R) variants. Comparing results of wide-line ¹H-NMR and molecular dynamics simulations we found that both the amount of mobile water surrounding proteins in ice, as well as their thaw profiles differs significantly as function of the compactness and conformational heterogeneity of their structure. We found that (i) at around -50°C ~ 50 H₂O/protein melt (ii) if the protein is well-folded then this amount of mobile water remains quasi-constant up to -20°C , (iii) if disordered then the quantity of the lubricating mobile water increases with T in a constant manner up to ~ 200 H₂O/protein by reaching -20°C . Especially in the $-55^\circ\text{C} \leftrightarrow -15^\circ\text{C}$ temperature range, wide-line ¹H-NMR detects the heterogeneity of protein fold, providing the size of the hydration shell surrounding the accessible conformers at a given temperature. Results indicate that freezing of protein solutions proceeds by the gradual selection of the enthalpically most favored states that also minimize the number of bridging waters.

Polypeptide and protein drugs are becoming ever more common in the pharmaceutical industry (over 130 different products now approved for clinical use by FDA) but the purification, formulation and storage of such drugs is a continuous challenge. Both cooling to subzero temperatures and heating above physiological temperatures results in the unfolding of globular proteins, in a great variety of solvents^{1,2}. While the loss of global fold at high temperatures has been studied in detail, the significance and mechanism of cold denaturing are still under debate^{3–8}. One of the major difficulties of studying the process is ice formation, even though a few examples can be found where proteins manifest detectable unfolding already above water freezing temperatures^{9,10} – this, however is not the general case. Proteins frozen in ice undergo a dynamical transition when heated above $\sim -70^\circ\text{C}$, leading to the functionally relevant forms with the simultaneous onset of water motion within the first hydration shell¹¹, which has long been shown to be required for activity^{12–14}. Without a free and mobile water shell, functionally important protein motions cannot take place and there is a strong coupling between protein and water dynamics which can only be uncoupled at cryogenic temperatures^{15–17}. However, the extent of inter-connectedness of the two processes is still unclear.

Studies aimed at understanding the mechanism of cold denaturing are usually either theoretical, where accessible simulation times (and the shortcomings of the available water models) render ice-formation unlikely^{18–20} or

¹Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Budapest, 1117, Hungary. ²MTA-ELTE Protein Modelling Research Group, Pázmány Péter st. 1A, 1117, Budapest, Hungary. ³Institute for Solid State Physics and Optics, Wigner RCP of the HAS, 1121, Budapest, Hungary. Correspondence and requests for materials should be addressed to A.P. (email: perczel@chem.elte.hu)

pH = 7		pH < 3	
Sequence and localized charges	net charge	Sequence and localized charges	net charge
TC5b: = H ⁺ ₂ -NLYIQWLK ⁺ D ⁻ GGPSSGR ⁺ PPPS-O ⁻	+1 (+0.9) ^d	TC5b(H+): = H ⁺ ₂ -NLYIQWLK ⁺ D ⁽⁰⁾ GGPSSGR ⁺ PPPS-O ^(½-)	+2.5 (>2.5) ^d
TC5b_N1R: = H ⁺ ₂ -R ⁺ LVIQWLK ⁺ D ⁻ GGPSSGR ⁺ PPPS-O ⁻	+2 (+1.9) ^d	TC5b_N1R(H+): = H ⁺ ₂ -R ⁺ LVIQWLK ⁺ D ⁽⁰⁾ GGPSSGR ⁺ PPPS-O ^(½-)	+3.5 (>3.5) ^d

Table 1. Amino acid sequences and net-charges of (mini)proteins studied by wide-line NMR. ^dCalculated by PROTEIN CALCULATOR v3.4.

experimental, where the application of pressure or additives interferes with ice formation^{21–24}. Therefore, these studies disregard the interaction between the freezing/un-freezing of the hydration layer and of the solute protein molecules. It is now generally thought that during cooling – as opposed to heat denaturation – the protein 3D-fold preserves some of its compact native structure, resulting in a state better described as water-penetrated, than an unfolded one. The process seems to involve the extended hydration of both hydrophobic and hydrophilic residues, with most recent results pointing to the significance of the latter phenomena^{10,18,25}.

In this work our aim was to deduce the structural consequences of freezing in case of similarly sized but conformationally very different proteins in the –80 °C to 0 °C temperature range, through monitoring their hydration properties by wide-line ¹H-NMR spectroscopy without interfering with the freezing of their solution. The applied NMR technique relies on the detection and analysis of the limited but significant amount of water that remains mobile around proteins – hydration waters¹⁴ – way below 0 °C. This approach also allows for the experimental characterization of hindered-rotation barriers and mapping of the energetic heterogeneity of water molecules bound to the molecular surface of proteins^{26,27}. We have previously studied larger proteins with this technique^{26,27} and have identified differing hydration properties of globular and intrinsically disordered proteins (IDPs). While in case of globular proteins, a wide temperature range exists where the amount of mobile water doesn't change with increasing *T*, after melting of the first hydration layer of IDPs the thaw is continuous as *T* increases. The applied method allows for quantification of structural heterogeneity (the ratio of heterogeneous binding interface values (*HeR*)) based directly on the melting diagrams (MD) that report the amount of molten water measured by NMR as a function of temperature²⁶.

TC5b, a miniprotein, was chosen as the test case for the present cold-temperature studies because by small changes in its sequence great structural alterations can be evoked. TC5b consists of 20 residues and folds autonomously (under 4 μs in water at neutral pH)²⁸. It was produced - by truncation and designed mutations - from of Exendine-4²⁹, a protein drug used as a key GLP-1 receptor agonist³⁰ that potentiates insulin secretion in pancreatic β-cells^{31,32} and is used in the daily clinical practice to treat type II *Diabetes Mellitus*^{33–36}. The TC5b shortened variant has a well-defined 3D structure under physiological conditions: a hydrophobic core clustered around Trp⁶, surrounded by multiple secondary structural elements (α-helix, β-turn and polyprolin-II helix). TC5b has a cooperative melting profile with a protein like melting temperature: (~40–50 °C) with folding propensities like those of globular proteins. Its internal dynamics and folding pathways were studied in detail^{28,29,37–44}. The Asn¹→Arg switch in its sequence (TC5b_N1R) leads from folded to a largely unfolded state (Table 1)⁴⁵. Shifting the pH below 3 (TC5b(H+)) results in the appearance of partly-folded I-states^{37,38,46} characterized by elevated internal dynamics. Thus, three systems of nearly identical chemical composition representing three very different classes of proteins' fold (Folded (F)-, Intermediate (I)- and Unfolded (U)-states) could be studied and compared directly. Besides being easy to tune, these proteins are small enough to be characterized thoroughly by molecular dynamics simulations (MDSs) which allow, for the first time, enhancement of the NMR results of ice trapped macromolecules with atomistic detail.

Results and Discussion

TC5b has a compact, well-folded structure. Preservation of the folded, globular state of TC5b (neutral conditions) – even at low temperatures – was reflected by our wide-line NMR measurements (Fig. 1). Heating of the frozen sample from –80 °C affords a melting diagram (MD) characteristic for well folded proteins: after the thaw of first hydration layer ($T = -53 \pm 1$ °C, $E_a = 4.85 \pm 0.02$ kJ/mol), a plateau can be seen in the -48 ± 1 °C ≤ T ≤ -26 ± 1 °C temperature range (corresponding (as described in the experimental section) to a 5.01 ± 0.02 – 5.44 ± 0.02 kJ/mol potential energy range Figs 1 and 2S) indicating a temperature region where no additional water molecules melt. This means that the potential energy barriers corresponding to this temperature range are insurmountable for further waters present in the system. In other words, the protein and its lubricating shell of unfrozen waters remains intact; the energy of heating is absorbed in form of enhanced motions of these hydration water molecules. The melting of the first hydration layer at ~–50 °C (4.92 ± 0.02 kJ/mol) is in good agreement with our previous experience concerning a great variety of proteins^{15,16,26,27}, but also with the temperature where the water molecules specifically bound to the 2D surface of a polymer (with regularly spaced –OH groups to model solvated protein surfaces) start percolation – spreading – on the surface from previously formed low density water patches⁴⁷. This is also the temperature range where the movement of water associated to the protein surface starts in wetted powder samples¹¹ and where molecular oxygen trapped in the hydration shell of hemoglobin is liberated to diffuse into the protein matrix¹⁷.

The number of waters that become mobile during the first thaw can be estimated using the *n* value derived directly from the FID signal. Here, this corresponds to approximately 50 water molecules. Comparing this value to the MDS derived solvent structure (see Fig. 4S) we can conclude that only waters very close to the protein surface (within 2.6 Å) melt during this first melting step.

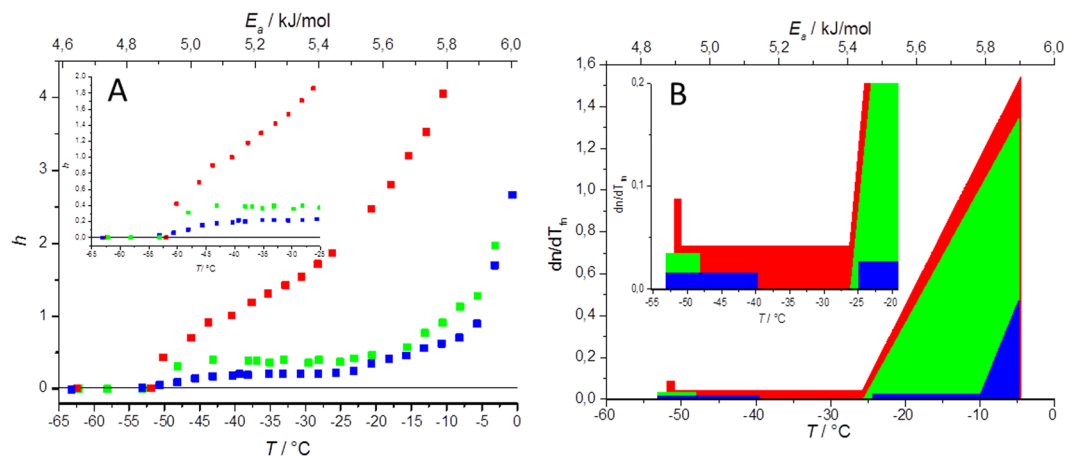


Figure 1. (A) Melting diagram – where h is the mass-normalized extent of hydration (in g water/ g protein units) - in TC5b in neutral (green) and acidic condition (blue) as well as in case of TC5b_N1R dissolved in water at pH = 6.9 (red). (B) Derivative melting diagram of TC5b in neutral (green) and acidic condition (blue) as well as that of the TC5b_N1R dissolved in water at pH = 6.9 (red).

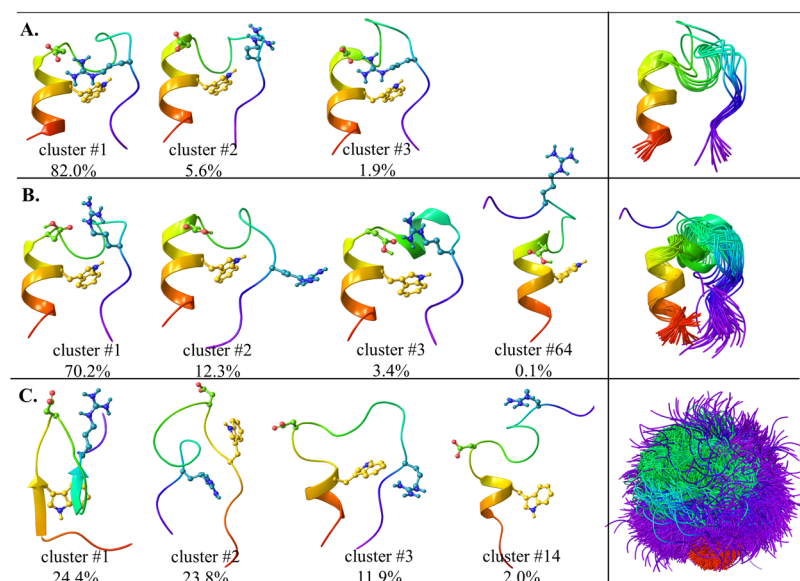


Figure 2. Left side: Characteristic backbone scaffolds of TC5b (A), TC5b(H⁺) (B) and TC5b_N1R (C) (showing Trp⁶ (in yellow), Asp⁹ (in green) and Arg¹⁶ (in blue) explicitly). Right side: Superimposed mid-structures of the most populated clusters (using a 1 Å cutoff) accounting for at least 90% of all snapshots of the last 800 ns of the molecular dynamics simulations.

From $T = -26 \pm 1^\circ\text{C}$, the amount of melted water continuously rises as temperature increases. This steep final section of the melting diagram (5.45 ± 0.02 – 5.90 ± 0.02 kJ/mol) can be attributed to water molecules in energetically heterogeneous interactions with the accessible protein residues of the surface (Fig. 2S).

MDS also indicated that the fold of neutral TC5b is composed of nearly homogenous and conserved scaffolds over time, as expected (Fig. 3S)^{29,38}. Grouped based on their backbone structural similarities, the mid-structure of the most populated cluster (cluster #1), accounting for over 82% of the snapshots, could be fit with a 0.7 Å rmsd to the backbone of the NMR determined solution state structure (PDB 112y) (Fig. 2), also confirming the applicability of the presently used MDS-protocol.

It was shown in previous ambient temperature solution-state NMR measurements that under acidic conditions (affording TC5b(H⁺)), local charge shifts (Table 1) of TC5b(H⁺) result in the loss of the key fold stabilizing salt-bridge of Asp⁹-Arg¹⁶ as the Asp side chain becomes protonated⁴⁶ leading to a dynamic ensemble of several semi-folded I-states. In the present wide-line NMR measurements we found that below pH = 3 the thaw of first hydration layer of TC5b(H⁺) becomes more extended, taking place in the temperature range of $-53 \pm 1^\circ\text{C} \leq T \leq -40 \pm 1^\circ\text{C}$ (corresponding to 4.82 ± 0.02 – 5.13 ± 0.02 kJ/mol potential energy barriers). This is followed by a temperature independent plateau ($-40 \pm 1^\circ\text{C} \leq T \leq -25 \pm 1^\circ\text{C}$) (Fig. 1) where the thaw

	$T_{\text{fno}} - T/^\circ\text{C}$	$T_{\text{fne}} - T/^\circ\text{C}$	HeR
TC5b	$0.842 \pm 0.008 - 43 \pm 1^\circ\text{C}$	$0.915 \pm 0.008 - 23 \pm 1^\circ\text{C}$	0.536 ± 0.007
TC5b(H+)	$0.850 \pm 0.008 - 41 \pm 1^\circ\text{C}$	$0.897 \pm 0.008 - 28 \pm 1^\circ\text{C}$	0.687 ± 0.003
TC5b_N1R	—	—	1

Table 2. The determined T_{fno} , T_{fne} and HeR of three miniproteins. T_{fno} and T_{fne} are the beginning and end point of the $dn/dT_{\text{fm}} \sim 0$ plateau region. The ratio of heterogeneous binding interface (HeR) was defined as $HeR = (1 - T_{\text{fne}})/(1 - T_{\text{fno}})$.

stops and no further waters can be mobilized by increasing temperature (or the available potential energy). At $\text{pH} < 3$, the side chain of Asp is dominantly ($>90\%$) neutralized $\text{Asp}^{(0)}$ (with a C'-terminal of main chain being half-protonated/deprotonated (50%/50%)) (Table 1), so MDS of TC5b was repeated with a protonated Asp^9 to describe TC5b(H+). The folded-like structure and the α -helical segment was conserved, however in only 12% of the structures we did find an H-bond between Asp^9 and Arg^{16} , as opposed to the $>92\%$ of the structures of the TC5b simulation (Fig. 2). In spite of this, the most populated cluster of TC5b(H+) (cluster #1), encompassing 70.2% of all structures, comprises globularly folded forms, similar to that of TC5b (backbone rmsd: 0.6 \AA). Furthermore, cluster #2 of TC5b(H+) (12.3% of the structures) also shows quite a similar 3D-fold, with slightly more opened backbone structures. However, elements of the above 2 clusters are less stringently packed compared to those of TC5b(neutral): the relative (average) backbone mobility increased by a factor of 1.9 and 2.2, in case of clusters #1 and #2 of TC5b(H+), as compared to that of cluster #1 of TC5b. The sampling of such loosened arrangements potentiates various unfolding routes resulting in the appearance of conformers where – for example – the 3_{10} -helix of residues 10–14 is turned into an α -helix (3% of the structures), or where the Trp-cage opens up completely (0.1%) (Fig. 2). These findings allow elaboration of the melting profile: because of the different conformers contributing to the “folded” ensemble which all possess a unique hydration layer, the first melting event is prolonged, and due to the higher number of low energy open forms, the temperature independent plateau ($-40 \pm 1^\circ\text{C} \leq T \leq -25 \pm 1^\circ\text{C}$ or $5.13 \pm 0.02 \text{ kJ/mol} \leq E_a \leq 5.46 \pm 0.02 \text{ kJ/mol}$) is shortened (Figs 1 and 2S) but is still present and characteristic. This means that at $\text{pH} < 3$, the TC5b carries some characteristics of both folded and disordered proteins.

Although TC5b_N1R differs from TC5b only by its first amino acid, this mutation has a large and global conformational effect. In principle, the side chain of the positively charged N-terminal Arg(+) could hook its C-terminal via an $\text{Arg}^{1(+)} \leftrightarrow -\text{COO}^{(-)}$ salt bridge, but this does not take place because the α -helix macro-dipole clashes with the positive charge of Arg^1 and the backbone unfolds into a highly dynamic U-state⁴⁵. In line with this, the MDS derived equilibrium structural ensemble of TC5b_N1R consists of greatly varied/perturbed conformers. Only about $\sim 12\%$ of the snapshots have a folded structure (Fig. 2), the remaining 88% forms the ensemble of a U-state. Accordingly, melting of the hydrate layer as recorded by wide-line NMR is continuous (Figs 1 and 2S) once the melting process starts (at $-50 \pm 1^\circ\text{C}$ or $4.91 \pm 0.02 \text{ kJ/mol}$), and from this point the number of hydration water molecules that become mobile around the protein increases at a constant rate with increasing temperature, which results in a constant derivative in the 4.88 ± 0.02 – $5.45 \pm 0.02 \text{ kJ/mol}$ energy region. The n value of the initial thaw signals the presence of compact conformers lubricated by ~ 50 water molecules, similarly to that seen in case of TC5b. As heating progresses, more and more H_2O s become mobile, clear characteristics of an IDP. At -35°C ($5.24 \pm 0.02 \text{ kJ/mol}$; about the middle of the plateaus of TC5b and TC5b(H+)) there are more than 120 mobile water/protein present, thus more than twice as many as in its more ordered counterparts.

We found that to explain the basic features of the observed melting curves it is sufficient to consider the room temperature backbone-conformational heterogeneity of the studied proteins. Using a quite strict cutoff value of 1 \AA , 90% of all snapshots of the equilibrated MDS trajectory could be adequately described using the 10 most populated conformers of TC5b (cluster #1 of the folded form accounts for $>80\%$ of them). On the contrary, 73 clusters for the loose, but still folded TC5b(H+) (cluster #1 accounting for $>70\%$ of them) and as many as 6042 clusters are required for depicting the conformational ensemble of TC5b_N1R. For the latter molecule, the most folded-like cluster comprises only 12% of all structures (Fig. 2). The three systems studied here therefore represent three thermodynamically different states: (i) TC5b has a practically non-degenerate F-state and very high energy unfolded region. (ii) The folded state of TC5b(H+) is moderately degenerate while its unfolded region remains high in energy and scarcely populated. (iii) For TC5b_N1R the folded and unfolded ensembles overlap in energy and are degenerate.

These findings are in correlation with the ratio of heterogeneous binding interface values (HeR) derived from the wide-line NMR measured MDs (Table 2): 0.536 ± 0.007 , 0.687 ± 0.003 and 1 for TC5b, TC5b(H+) and TC5b_N1R, respectively, also indicating a $\text{TC5b} > \text{TC5b(H+)} \gg \text{TC5b_N1R}$ order of backbone scaffold homogeneity. The surface of TC5b is partly homogeneous (indicated also by zero derivative section) and partly ($\sim 54\%$) heterogeneous, probably due to the highly flexible C-terminal region of the protein. In case of TC5b(H+) the heterogeneous content increases ($\sim 69\%$) and TC5b_N1R is fully disordered (100%) based on this descriptor too.

With all the above considered, we can provide a “high resolution” picture of the melting process of these proteins starting from their encapsulated forms in ice at -80°C ($E_a = 4.26 \pm 0.02 \text{ kJ/mol}$). Approximately 30–70 water molecules per protein melt at $\sim -50^\circ\text{C}$ ($E_a = 4.92 \pm 0.02 \text{ kJ/mol}$) (determined from the three MDs) and start protein surface lubrication first (see Fig. 4S). This amount of H_2O corresponds to the first hydration layer, within 2.5 – 2.8 \AA of the protein surface. As these waters must bridge the irregular surface of the molecular scaffolds and the edges of ordered bulk ice, the amount of melted water is expected to be chiefly determined by the size of the solvent accessible surface of the conformers that are populated at a given temperature. Sequences that do not possess a dominant fold but comprise instead of a multitude of conformationally versatile and easily accessible

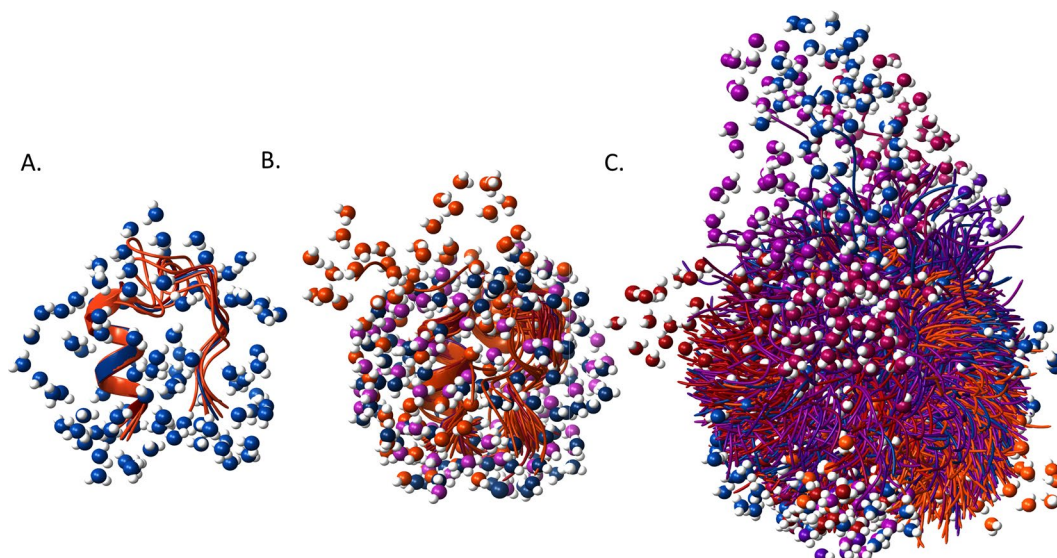


Figure 3. Backbone structures of TC5b (A), TC5b(H+) (B) and TC5b_N1R (C) conformers accounting for at least 90% of all snapshots of the last 800 ns of the molecular dynamics simulations with surrounding surface waters within 2.8 Å of the protein surface. Conformers are colored from blue to orange - with blue showing the most populated clusters' mid-structures, while lighter colors indicating those of the less populated clusters'. Liquid H₂O's are colored in line with the color of the backbone scaffold. For a more detailed analysis of the water-structure, see Fig. 4S.

backbone scaffolds, when heated from low temperatures will gradually populate all these states. Appearance of structurally distinct protein folds loosens the surrounding structure of the ice, since the gradual increase of the available thermal activation energy as T increases will allow for the melting of the hydration shell of less and less compact conformers, with growing number of mobile waters, as is the case of TC5b_N1R (Fig. 3).

In contrast to this, for a well-structured globular protein with no accessible low-lying states, after the melting of the first hydration shell, heating will only enhance the mobility of both the trapped protein and the liberated waters but for a considerable temperature range (5.01 ± 0.02 – 5.44 ± 0.02 kJ/mol) no increase in the number of freely-moving waters will be seen: here the rate of the introduction of newly mobilized water molecules and thus the derivative of the melting diagram (dn/dT_m) is zero (Figs 1 and 2S). In other words, the accessible potential energies in this temperature range are insufficient for inducing the transfer of any of the waters present from ice-contained to the melted state. This “zero-derivative” region is the most prominent feature of the melting diagram of TC5b at neutral pH. TC5b(H+) presents an example of the intermediate between the two extremes, having few low-lying states which provide a continual increase of freely-moving waters during the first melting thaws starting at $\sim -50^\circ\text{C}$ (4.92 ± 0.02 kJ/mol), but after this, $n(T_m)$ forms a plateau (5.13 ± 0.02 – 5.43 ± 0.02 kJ/mol) and $dn/dT_m = 0$. Clearly, these melting diagrams reflect on the conformational and energetic heterogeneity of the solvent accessible surface of the studied proteins.

Analyzing the first thaw of the melting process, we see that the amount of water liberated in this step is only enough for “covering” compact – well-folded – conformers of the protein but it is simply too few to create a closed solvent layer of even monomolecular thickness around an open – unfolded – conformation (Fig. 4). Thus, it seems that special hydration effects attributed to cold denaturing lose some of their significance in the presence of ice formation, where, according to our results, cooling results in the gradual selection of the enthalpically most favored state with as few bridging waters as possible (as also illustrated by the continual decrease of n with decreasing temperatures).

Cold denaturing of TC5b was the subject of a detailed replica exchange molecular dynamics study published recently¹⁸, which offers an interesting comparison with the present results. Authors found distinct signs of cold denaturing below -60°C ($E_a = 4.70 \pm 0.02$ kJ/mol), remarkably close to the temperature range where our studies and those of others^{11,47} indicate the universal (heterogeneity independent) unfreezing of the first hydration shell. Cold denaturing was accompanied by (i) the widening of the Trp-cage (increase of the Trp⁶ ↔ Ser¹⁴ distance), (ii) loss of the Asp⁹ ↔ Arg¹⁶ salt bridge, (iii) increase in the number of protein-water H-bonds and (iv) decrease of intramolecular contacts. However, all these changes have resulted in neither the significant increase of the solvent accessible surface area (SASA) nor that of the radius of gyration (R_g). This supports the concept that cold denaturing results in loosening of intramolecular interactions but not the vast unfolding of the protein scaffold. Heat denaturing, on the other hand, lead to characteristic increase of both SASA and R_g , with the reduction of 3D-fold dependent pairwise interactions. Similar tendencies could be captured at the room temperature MDSs of our three model systems, with TC5b(H+) behaving very similarly to the cold denatured state of TC5b, while TC5b_N1R corresponding to the heat denatured ensemble of TC5b (Fig. 5S). In case of TC5b(H+), beside the conformers where Ser¹⁴ points toward Trp⁶ and thus the inner region of the cage, other arrangements with similar backbone trace appear. Ser¹⁴ being immersed in the solvent allows for water to penetrate the hydrophobic core

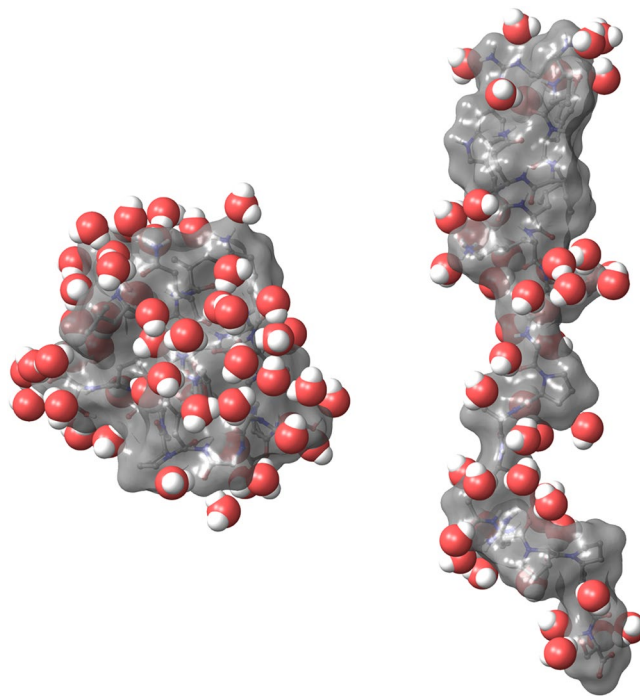


Figure 4. The closest 50 water molecules solvating a folded conformation of TC5b (left) and an unfolded conformation of TC5b_N1R (right) (snapshots from MDS).

of the folded protein. The typical number of H-bonds between Asp⁹ and Arg¹⁶ is reduced from 2 (seen in case of TC5b) to 0. Since the R_g does not increase appreciably (the fold remains intact), the slight increase in SASA can be attributed to the moderate increase in the hydration of the inner domains of the folded form. In contrast, in the simulation of TC5b_N1R, the small Trp⁶ ↔ Ser¹⁴ distances belonging to the folded scaffolds were sparsely sampled. Instead an almost uniform distribution of longer distances resulted, reflecting the presence of a multitude of differently unfolded conformations. Interestingly, while the typical number of H-bonds between Asp⁹ and Arg¹⁶ is 0 in this case too, in ~20% of all the structures, well-oriented interaction and the formation of 2 H-bonds is seen. This, however, isn't sufficient for the re-instatement of a folded scaffold. Broad distributions and a characteristic increase in the average values of both SASA and the R_g were seen, indicative of the presence of only a few really compact conformers in the equilibrium ensemble.

In the cold denaturing study of TC5b¹⁸ local solvent density fluctuation (commonly used to characterize the effective hydrophobicity of each amino acid residue) was also analyzed, and it was shown that at colder temperatures hydrophobic residues of the Trp-cage experience a pronounced decrease in hydrophobicity, a phenomenon measured by the decrease of the number of fluctuating H₂O of close proximity of a selected residue. Although the hydrophobic nature of Trp⁶, Gly¹¹ and Pro¹⁸ is confirmed by the low number of waters associated with these residues during the present MDSs of TC5b, TC5b(H⁺) and TC5b_N1R too, fluctuation of H₂O near the hydrophobic core did not decrease in case of TC5b(H⁺). A slight increase was seen, which however is in agreement with the fact that backbone fold conformational heterogeneity also increases in this region – an effect that is not expected in cold temperature simulations (Fig. 6S).

The similarities of the two sets of MDSs indicate that hydrophilic residues and changes in their interactions must contribute greatly to the processes of cold denaturation and heat induced unfolding. It was clearly demonstrated that turning off the Asp⁹ ↔ Arg¹⁶ interaction by a pH shift initiates changes that resemble those of the cold denaturation of TC5b, while disturbing the electrostatics of the N⁻termini leads to an unfolding similar to that experienced upon heating.

Conclusion/Outlook

Our results indicate that basic features of a protein fold are not necessarily altered by freezing its solution. We showed that monitoring the melting process of frozen protein solutions by wide-line NMR spectroscopy affords an adequate description and a quantitative measure (*HeR*) for the heterogeneity of F-, I- and U-state proteins, and presents a clear-cut method for differentiating IDPs from well folded proteins and also from those systems that simply possess a native folded state that is degenerate. The comparison of the present experimental and MDS data with the results of cold denaturing studies¹⁸ shows that the loosening of tightly packed structures of the native fold and water penetration into the hydrophobic core is prompted most likely by the loss of hydrophilic intramolecular interactions. By cooling and heating the system or by the chemical inactivation of certain crucial charge-mediated interactions, very similar structural changes can be induced. Results also indicate that the transition of proteins from “frozen” to functional form should be expected above ~–55 °C, where mobile waters of the first hydration shell allow for the start of functionally relevant protein movements too.

Methods

Protein expression and purification. The measured miniproteins were produced by bacterial expression using previous published protocol⁴⁸.

Wide-line NMR spectroscopy. ¹H-NMR measurements and data acquisition were accomplished by a Bruker AVANCE III spectrometer at $\omega_0/2\pi = 82.4$ MHz with a stability of better than $\pm 10^{-6}$, and with 2 ppm inhomogeneity of the magnet. The value represents the stability of the electromagnet, without any drift during measurements; the stability of the frequency is by two orders of magnitude better for our system. The recycle delay was $5 * T_1$ and T_1 was measured at various temperatures. The temperature was controlled by an open-cycle Janis cryostat, stable to ± 0.1 °C and the uncertainty of the scale was ± 1 °C. Data points are based on spectra recorded by averaging signals to reach a signal/noise ratio better than 50. The portion of mobile proton (water) fraction, n , is directly determined by the FID signal based on the comparison of the signal intensity extrapolated to zero time with the corresponding values measured at a temperature where the whole sample is in liquid state^{49,50}. This procedure was shown to give identical results to those derived from CPMG-echo experiments for a variety of different proteins¹⁶. The number of averaged NMR signals was varied to achieve the desired signal quantity for each sample and for unfrozen water quantities. The typical length of the $\pi/2$ pulse was 2.2 μ s. The sample volume was 70 μ l. The sample tube is a cylindrical, capped Teflon capsule of 5 mm diameter. We use a probe designed for wide-line NMR with a simple coil.

Data extraction, scaling and analysis. The chemical shifts or their evolution cannot be seen, they are buried by the width of the signal detected by wide-line NMR. All of the spins are on resonance as wide-line NMR can see them. The amplitude of the NMR response is directly proportional to the equilibrium state nuclear magnetization M_{i0} and so, to the number of contributing nuclei, n_i , in the i^{th} nuclear spin fraction in the sample. The free induction decay (FID) signal can be separated into distinct fractions for a multicomponent system. The slowly decaying component is generated by the mobile ¹H nuclei (mostly in hydration water molecules in our case). The extrapolated amplitude of a component corresponds to the magnetization induced by the involved nuclei as $M_{i0} = n_i B_0 / T$. This expression gives the direct measure of each ¹H fraction. In aqueous solution samples there are various water phases⁵¹ which freeze at lower temperature *i.e.* well below 0 °C. The on-resonance measured FID (¹H-NMR signal) at each temperature (-80 °C $< T < 0$ °C) comprises proton resonances decaying at different spin-spin relaxation rates, T_2 s, determined by their physico-chemical nature⁵². Protons of the ice and the protein relax faster ($T_2 < 50$ – 100 μ s) compared to those of freely moving liquid water ($T_2 > 100$ μ s) (Fig. 1S). Thus, the amount of ¹Hs forming the FID could be clustered and separated based on their relaxation properties. So the signal detected in the time interval from zero to ~ 100 μ s is the mixture of transient signals over the dead-time period (6–8 μ s) of the spectrometer and parts of ice and protein FID signals and in the analysis we used the signal of the moving water which is a protractedly decaying signal (between 200 μ s and 1 ms). In the previous study²⁷ we could demonstrate the absence of this signal, where we measured the protein sample after vacuum drying. In this case only the quickly decaying protein signal was detected without water protons.

In this article the measured FID signals are converted into the portion of mobile proton (water) fraction, n . The decays of the mobile elements are not exponential due to the inhomogeneity of the external magnetic field. We fitted Gaussian functions to the FID data points at each temperature and used the coefficient of slowly relaxing protons (200 μ s – 1 ms). After normalization of n at every temperature with all mobile water (the value is 100% above 0 °C) first we determined the normalized fundamental temperature ($T_{\text{fn}} = T/273.15$ K) and then the potential barrier ($E_a(0$ °C) = 6.01 kJ/mol) which energy is necessary for excitation of movement of water molecules. (The normalization of the energy scale is made by the melting heat of ice⁵³.)

T_{fno} and T_{fne} are the beginning and end point of the $dn/dT_{\text{fn}} \sim 0$ plateau region. $T_{\text{fno}} \ll T_{\text{fne}}$ is indicative of a folded protein (a protein in its F-state), while $T_{\text{fno}} \approx T_{\text{fne}}$ refers to a disordered one (a protein in its U-state). Based on this, the ratio of heterogeneous binding interface (*HeR*) was defined as $HeR = (1 - T_{\text{fne}})/(1 - T_{\text{fno}})$. For details see *Supporting information*. Using n and the concentration of the samples (corresponding roughly to the number of protein molecules/water molecules), the number of mobile waters at each temperature can be estimated. We converted the measured signal intensities $n = N_i/N_0$ into hydration as $h = n \cdot m_{\text{water}}/m_{\text{protein}}$, where N_0 is the signal intensity measured above 0 °C, m_{water} is the mass of water, and m_{protein} is the mass of protein.

Molecular dynamics simulations. MDSs concerning both TC5b under neutral and acidic conditions of and TC5b_N1R were carried out as implemented in GROMACS⁵⁴, using the AMBER99SB-ILDNP forcefield⁵⁵. Starting conformers of all three models were derived from that of TC5b as determined by NMR (PDB 1L2y) with identical backbone structure. Systems were solvated by TIP3P water molecules in dodecahedral boxes with 10 Å buffer. The total charge of the system was neutralized and physiological salt concentration set using Na⁺ and Cl⁻ ions. Energy minimization of starting structures was followed by sequential relaxation of constraints on protein atoms in three steps (all of 100 ps). Trajectories of 2000–3000 ns NPT simulations at 325 K and 1 bar were recorded for further analysis (collecting snapshots at every 4 ps). Clustering of conformations⁵⁶ was carried out based on the main-chain conformation of snapshots of the last 800 ns of the simulations using a cutoff of 1.0 Å. Waters near the surface of the protein were counted using the select command of GROMACS⁵⁴.

References

- Roy, I. & Gupta, M. N. Freeze-drying of proteins: Some emerging concerns. *Biotechnol Appl Biochem.* **39**, 165–77 (2004).
- Kardos, J. *et al.* Phosphorylation as Conformational Switch from the Native to Amyloid State: Trp-Cage as a Protein Aggregation Model. *J. Phys. Chem. B.* **119**, 2946–2955 (2015).
- Szyperski, T., Mills, J. L., Perl, D. & Balbach, J. Combined NMR-observation of cold denaturation in supercooled water and heat denaturation enables accurate measurement of $\Delta C(p)$ of protein unfolding. *Eur. Biophys. J.* **35**, 363–366 (2006).

4. Pometun, M. S., Peterson, R. W., Babu, C. R. & Wand, A. J. Cold denaturation of encapsulated ubiquitin. *J. Am. Chem. Soc.* **128**, 10652–10653 (2006).
5. Adrover, M., Esposito, V., Martorell, G., Pastore, A. & Temussi, A. P. Understanding Cold Denaturation: The Case Study of Yfh1. *J. Am. Chem. Soc.* **132**, 16240–16246 (2010).
6. Adrover, M. *et al.* The role of hydration in protein stability: comparison of the cold and heat unfolded states of Yfh1. *J. Mol. Biol.* **417**, 413–24 (2012).
7. Jaremko, M. *et al.* Cold denaturation of a protein dimer monitored at atomic resolution. *Nat. Chem. Biol.* **9**, 264–70 (2013).
8. Pastore, A. *et al.* Unbiased cold denaturation: Low- and high-temperature unfolding of yeast frataxin under physiological conditions. *J. Am. Chem. Soc.* **129**, 5374–5375 (2007).
9. Nakao, Y., Yaka, A., Ebata, S. & Hiayama, T. A Dramatic Effect of Lewis-Acid Catalysts on Nickel-Catalyzed Carbocyanation of Alkynes. *J. Am. Chem. Soc.* **129**, 2428–2429 (2007).
10. Sanfelice, D., Morandi, E., Pastore, A., Niccolo, N. & Temussi, A. P. Cold Denaturation Unveiled: Molecular Mechanism of the Asymmetric Unfolding of Yeast Frataxin. *ChemPhysChem* **16**, 3599–3602 (2015).
11. Schirò, G. *et al.* Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins. *Nature Communications* **6**, 6490, <https://doi.org/10.1038/ncomms7490> (2015).
12. Rupley, J. A., Gratton, E. & Careri, G. Water and globular proteins. *Trends Biochem. Sci.* **8**, 18–22 (1983).
13. Laage, D., Elsaesser, T. & Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **117**, 10694–10725 (2017).
14. Kuntz, I. D. & Kauzmann, W. Hydration of Proteins and Polypeptides. *Adv. Prot. Chem.* **2**, 239–345 (1974).
15. Tompa, P. *et al.* 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).
16. Tompa, K. *et al.* Interfacial water at protein surfaces: wide-line NMR and DSC characterization of hydration in ubiquitin solutions. *Biophys. J.* **96**, 2789–2798 (2009).
17. Schay, G. *et al.* Millisecond Time-Scale Protein Dynamics Exists Prior to the Activation of the Bulk Solvent Matrix. *J. Phys. Chem. B.* **115**, 5707–5715 (2011).
18. Kim, B. S., Palmer, C. J. & DeBenedetti, G. P. Computational investigation of cold denaturation in the Trp-cage miniprotein. *Proc. Nat. Ac. Sci.* **113**, 8991–8996 (2016).
19. Dias, L. C., Ala-Nissila, T., Karttunen, M., Vattulainen, I. & Grant, M. Microscopic Mechanism for Cold Denaturation. *Phys. Rev. Lett.* **100**, 118101, <https://doi.org/10.1103/PhysRevLett.100.118101> (2008).
20. Yang, C., Jang, S. & Pak, Y. A fully atomistic computer simulation study of cold denaturation of a β -hairpin. *Nat. Commun.* **5**, 5773, <https://doi.org/10.1038/ncomms6773> (2014).
21. Vajpai, N., Nisius, L., Wiktor, M. & Grzesiek, S. High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin. *Proc Natl Acad Sci USA* **112**, E368–376, <https://doi.org/10.1073/pnas.1212222110> (2013).
22. Davidovic, M., Mattea, C., Qvist, J. & Halle, B. Protein cold denaturation as seen from the solvent. *J. Am. Chem. Soc.* **131**, 1025–1036 (2009).
23. Buchner, G. S., Shih, N., Niebling, S., Reece, A. E. & Kubelka, J. Unusual cold denaturation of a small protein domain. *Biochemistry* **51**, 6496–6498 (2012).
24. Dyer, B. R. *et al.* Hairpin folding dynamics: the cold denatured state is predisposed for rapid refolding. *Biochemistry* **44**, 10406–10415 (2005).
25. Ben-Naim, A. Theory of cold denaturation of proteins. *Adv. Biol. Chem.* **3**, 29–39 (2013).
26. Tompa, K., Bokor, M., Verebélyi, T. & Tompa, P. Water rotation barriers on protein molecular surfaces. *Chemical Physics* **448**, 15–25 (2015).
27. Tompa, K. *et al.* Hydrogen Mobility and Protein–Water Interactions in Proteins in the Solid State. *ChemPhysChem* **18**, 677–682 (2017).
28. Qiu, L., Pabit, S. A., Roitberg, A. & Hagen, S. J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μ s. *J. Am. Chem. Soc.* **124**, 12952–12953 (2002).
29. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9**, 425–430 (2002).
30. Eng, J., Kleinman, W. A., Singh, L., Singh, G. & Raufman, J. P. Isolation and Characterization of Exendin-4, an Exendin-3 Analogue, from *Holoderma suspectum* Venom. *J. Biol. Chem.* **267**, 7402–7405 (1992).
31. Göke, R., Fehmann, H. C., Linn, T., Schmidt, H. & Krause, M. Exendin-4 is a high potency agonist and truncated exendin-(9–39)-amide an antagonist at the glucagon-like peptide 1-(7–36)-amide receptor of insulin-secreting β -cells. *J. Biol. Chem.* **268**, 19650–19655 (1993).
32. Montrose-Rafizadeh, C. *et al.* High potency antagonists of the pancreatic glucagon-like peptide-1 receptor. *J. Biol. Chem.* **272**, 21201–21206 (1997).
33. Young, A. A. *et al.* Glucose-lowering and insulin-sensitizing actions of exendin-4: Studies in obese diabetic (ob/ob,db/db) mice, diabetic fatty Zucker rats, and diabetic rhesus monkeys (*Macaca mulatta*). *Diabetes* **48**, 1026–1034 (1999).
34. Gallwitz, B. *et al.* GLP-1 analogues resistant to degradation by dipeptidyl-peptidase IV *in vitro*. *Regul. Pept.* **86**, 103–111 (2000).
35. Green, B. D. & Flatt, P. R. Incretin hormone mimetics and analogues in diabetes therapeutics. *Best Pract. Res. Clin. Endocrinol. Metab.* **21**, 497–516 (2007).
36. Bhavsar, S., Mudaliar, S. & Cherrington, A. Evolution of Exenatide as a Diabetes Therapeutic. *Curr. Diabetes Rev.* **9**, 161–193 (2013).
37. Rovó, P. *et al.* Structural insights into the Trp-cage folding intermediate formation. *Chem. Eur. J.* **19**, 2628–2640 (2013).
38. Juraszek, J. & Bolhuis, P. G. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA* **103**, 15859–15864 (2006).
39. Juraszek, J. & Bolhuis, P. G. Rate Constant and Reaction Coordinate of Trp-Cage Folding in Explicit Water. *Biophys. J.* **95**, 4246–4257 (2008).
40. Marinelli, F., Pietrucci, F., Laio, A. & Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **5**, e1000452, <https://doi.org/10.1371/journal.pcbi.1000452> (2009).
41. Farkas, V., Csordás, B., Hegyi, O., Tóth, K. G. & Perczel, A. Foldamer Stability Coupled to Aggregation Propensity of Elongated Trp-Cage Miniproteins. *Eur. J. Org. Chem.* **13**, 3513–3522 (2013).
42. Culik, R. M., Serrano, A. L., Bunagan, M. R. & Gai, F. Achieving Secondary Structural Resolution in Kinetic Measurements of Protein Folding: A Case Study of the Folding Mechanism of Trp-cage. *Angew. Chem.* **123**, 11076–11079 (2011).
43. Farkas, V., Jákl, I., Tóth, K. G. & Perczel, A. Aromatic cluster sensor of protein folding: near-UV ECD bands assigned to fold compactness. *Chemistry A. Eur. J.* **22**, 13871–13883 (2016).
44. Streicher, W. W. & Makhatadze, G. I. Unfolding Thermodynamics of Trp-Cage, a 20 Residue Miniprotein, Studied by Differential Scanning Calorimetry and Circular Dichroism Spectroscopy. *Biochemistry* **46**, 2876–2880 (2007).
45. Rovó, P. *et al.* Rational Design of α -Helix-Stabilized Exendin-4 Analogues. *Biochemistry* **53**, 3540–3552 (2014).
46. Rovó, P. *et al.* Cooperativity network of Trp-cage miniproteins: probing salt-bridges. *J. Pept. Sci.* **17**, 610–619 (2011).
47. Zanotti, J. M. *et al.* Competing coexisting phases in 2D water. *Sci Rep.* **6**, 25938, <https://doi.org/10.1038/srep25938> (2016).
48. Stráner, P., Taricska, N., Szabó, M., Tóth, K. G. & Perczel, A. Bacterial expression and/or solid phase peptide synthesis of 20–40 amino acid long polypeptides and miniproteins, the case study of Class B GPCR ligands. *Current Protein and Peptide Science* **17**, 147–155 (2016).

49. Tompa, K., Bokor, M. & Tompa, P. PART II. Chapter 12. Hydration of intrinsically disordered proteins from wide-line NMR in *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation* (ed. Uversky, V. N. & Longhi, S.) 345–368 (New Jersey, 2010).
50. Tompa, K., Bokor, M., Tompa, P. Wide-line NMR and Protein Hydration in *Intrinsically Disordered Protein Analysis*: (ed. Uversky, N. V. & Dunker, A. K.) 167–196 (New Jersey, 2012).
51. Cooke, R. & Kuntz, I. D. The Properties of Water in Biological Systems. *Annu. Rev. Biophys. Bioeng* **3**, 95–126 (1974).
52. Bokor, M. *et al.* NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophys J.* **88**, 2030–2037 (2005).
53. Haynes W. M. (Ed.-in-Chief), Section 6: Fluid properties (6–151) in *CRC Handbook of Chemistry and Physics*, 95th edition 2014–2015, CRC Press Taylor & Francis Group, Boca Raton, FL., USA.
54. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
55. Aliev, A. E. *et al.* Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. *Proteins* **82**, 195–215 (2014).
56. Daura, X., Antes, I., Gunsteren, W. F., Thiel, W. & Mark, A. E. The effect of motional averaging on the calculation of NMR-derived structural properties. *Proteins* **36**, 542–555 (1999).

Acknowledgements

The authors gratefully acknowledge Tamás Verebéli for his contribution to wide-line NMR measurements. This work was supported by Grants from the Hungarian Scientific Research Fund (OTKA K116305) and VEKOP-2.3.2-16-2017-00014. MDSs were carried out at the facilities of the Hungarian NIIF Institute.

Author Contributions

N.T. prepared the measured miniproteins. M.B., K.T. performed wide-line NMR spectroscopy measurements and N.T. performed data analysis. D.K.M. performed molecular dynamics simulations. N.T. and D.K.M. prepared the figures. N.T., D.K.M., M.B., K.T. and A.P. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-39121-5>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019