The FEBS Journal
FEBS PRESS
science publishing by scientists

Structural insight into a partially unfolded state preceding aggregation in an intracellular lipid-binding protein

Gergő Horváth[1], László Biczók[2], Zsuzsa Majer[3], Mihály Kovács[4], András Micsonai[5], József Kardos[5] and Orsolya Toke[1]

1 Laboratory for NMR Spectroscopy, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
2 Institute of Materials and Environmental Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
3 Institute of Chemistry, Eötvös Loránd University, Budapest, Hungary
4 Department of Biochemistry, ELTE-MTA 'Momentum' Motor Enzymology Research Group, Eötvös Loránd University, Budapest, Hungary
5 Department of Biochemistry, MTA-ELTE NAP B Neuroimmunology Research Group, Institute of Biology, Eötvös Loránd University, Budapest, Hungary

Keywords
β-sheet proteins; disorder–order transitions; NMR spectroscopy; partial unfolding; protein aggregation

Correspondence
O. Toke, Laboratory for NMR Spectroscopy, Research Centre for Natural Sciences, Hungarian Academy of Sciences, 2 Magyar tudósok körútja, H-1117 Budapest, Hungary
Fax: +36 1 382 6295
Tel: +36 1 382 6575
E-mail: toke.orsolya@ttk.mta.hu

(Received 9 February 2017, revised 15 August 2017, accepted 5 September 2017)

doi:10.1111/febs.14264

Human ileal bile acid-binding protein (I-BABP) has a key role in the intracellular transport and metabolic targeting of bile salts. Similar to other members of the family of intracellular lipid-binding proteins (iLBPs), disorder–order transitions and local unfolding processes are thought to mediate ligand entry and release in human I-BABP. To gain insight into the stability of various protein regions, the temperature response of human I-BABP was investigated using NMR, CD and fluorescence spectroscopy, as well as molecular dynamics (MD) simulations. A joint analysis of NMR thermal melting and relaxation dispersion data indicates a complex pattern of internal dynamics with a dominating single barrier and a likely presence of rapidly exchanging conformational substates on both sides of the barrier. Moreover, our residue-specific analysis uncovers a partially unfolded U^* state in which part of the helical region with three proximate β-strands contains a substantial amount of residual structure, whereas several segments of the C-terminal half exhibit a high susceptibility to temperature elevation. Cluster analysis of atomic temperature responses indicates a thermodynamic coupling between distant protein sites including the bottom of the β-barrel, the E–F region and part of the helical cap. MD simulations up to 1 μs show correlated motions in the same protein regions and together with the NMR data suggest a role for the highly dynamic D–E turn and E–F region in the initiation of unfolding. The response of human I-BABP to temperature elevation is discussed in the context of the folding/unfolding behaviour of different members of the iLBP family.

Introduction

Human ileal bile acid-binding protein (I-BABP), a 14.2-kDa member of the family of intracellular lipid-binding proteins (iLBP) [1,2] is expressed in the absorptive enterocytes of the distal small intestine and has a key role in the transcellular trafficking of bile salts [3]. Additionally, along with other members of the iLBP family, BABPs are thought to participate in the regulation of the transcriptional activity of nuclear receptors. Specifically, stimulating the activity of the farnesoid X receptor [4], BABPs contribute to the

Abbreviations
cL-BABP, chicken liver bile acid-binding protein; CPMG, Carr–Purcell–Meiboom–Gill; CRBP, cellular retinol-binding protein; FABP, fatty acid-binding protein; hI-BABP, human ileal bile acid-binding protein; iLBP, intracellular lipid-binding proteins; MD, molecular dynamics.

regulation of bile acid, lipid and glucose homoeostasis. Accordingly, BABP-misfunctioning can lead to a wide range of metabolic disorders and diseases of the gastrointestinal tract including type 2 diabetes [5] and colorectal cancer [6,7].

The NMR structural studies of *apo* [8] and *holo* [9] human I-BABP show a topology characteristic of iLBPs. Specifically, a β-barrel comprised of 2 five-stranded antiparallel β-sheets encloses a binding cavity of ~ 1000 Å^3 covered by a helix-loop-helix motif at the top (Fig. 1). The cavity accommodates two binding sites exhibiting positive cooperativity [10] and a bile salt preference when complexed with a mixture of di- and trihydroxy bile salts [11]. Internal motions appear to have a major role in bile salt-BABP recognition [12,13]. Specifically, by providing access to the otherwise enclosed binding site, disorder–order transitions are thought to have a key role in regulating ligand entry in BABPs [9,14] and other members of the iLBP family [15]. Additionally, interaction of BABPs with lipid membranes, a process thought to be involved in transferring bile salts between cell compartments, has been reported to be accompanied by a partial unfolding of chicken L-BABP [16,17]. Conformational plasticity and local unfolding of iLBPs may also have a role in stimulation the transcriptional activity of nuclear receptors accomplished via nuclear translocation mechanisms triggered by ligand binding [18].

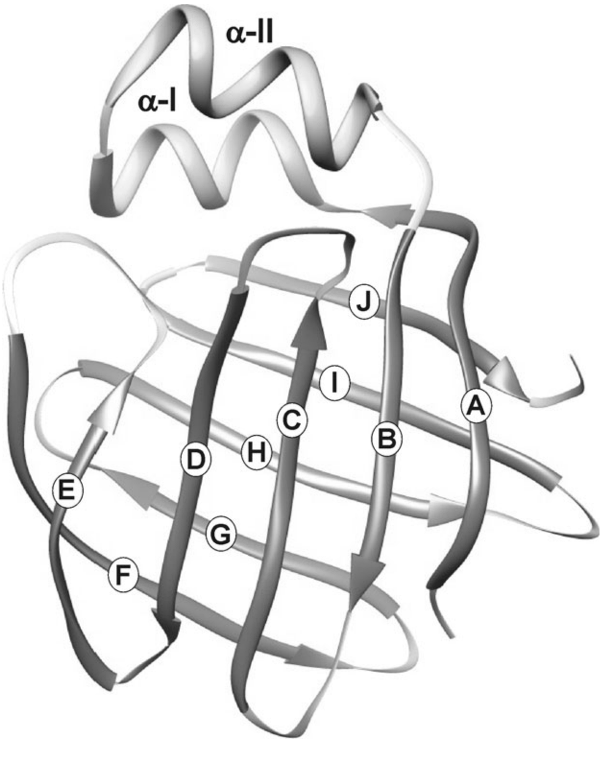

Fig. 1. Ribbon diagram of the solution NMR structure of human ileal bile acid-binding protein [8]. β-strands are labelled A–J for identification when discussed in the text.

Being a single-domain, predominantly β-sheet protein, the folding mechanism of iLBPs has been a subject of interest [19–26]. While folding studies of different members of the protein family uncover considerable differences in both the rates of folding steps and the number and type of intermediate states, most of the studies agree with the existence of a fast phase followed by one or two slower kinetic steps. For both rat and human I-BABP, stopped-flow kinetic study of refolding shows a fast phase resulting in an intermediate with native-like amounts of secondary structure, and a second, slower phase corresponding to the formation and stabilization of the native, tertiary fold [27]. During unfolding, three phases are observed for rat I-BABP, whereas for the human homologue, both fluorescence and CD detects a single phase [27]. Differences in the stability of kinetic intermediates have been suggested as a possible explanation for the different behaviour of the two highly homologous proteins. Comparative studies of various other members of the iLBP family indicate that the number of buried hydrophobic atoms in folding initiating sites might be a key determinant in the formation of different intermediate states during folding [20,24].

While most of these works have been concerned with chemically induced denaturation, in the current study, we are primarily focusing on thermally induced processes. Specifically, NMR spectroscopy enabling a residue-level monitoring of temperature response and internal protein dynamics was used to obtain insight into the stability of short- and long-range interactions and folding cooperativity in human I-BABP, a member of the iLBP family whose binding properties are currently under investigation in our laboratory. NMR experiments are complemented with CD, fluorescence measurements and molecular dynamics (MD) simulations. In the absence of denaturants, our studies indicate a complex dynamic behaviour with a dominating barrier separating the native state conformational ensemble (*F*) from a partially unfolded protein state (*U**) exhibiting a high propensity to dimerization and with a possible role in the formation of higher order aggregates of the predominantly β-sheet protein. Using a deconvolution technique suitable for accessing physicochemical parameters of protein states outside the experimental temperature range, we have characterized the $F \leftrightarrow U^*$ transition and identified some of the key determinants of protein stability. In agreement with the NMR data, MD simulations indicate a link between differences in vulnerability to temperature elevation and internal dynamics on the μs–ms time scale and suggest a role for the highly dynamic D–E turn and E–F region in the initiation of unfolding. Protein sites differing in their vulnerability to temperature elevation are discussed in the context of the folding/unfolding behaviour of different members of the iLBP family. In addition to providing a residue-level insight into the thermal behaviour of human I-BABP, our results have implications for aggregation mechanisms and protein stabilization of iLBPs, a family of proteins with a potential scaffold for drug carriers [28] and artificial enzymes [29].

Results

Temperature-induced unfolding monitored by fluorescence and CD spectroscopy

Representative fluorescence emission spectra of human I-BABP recorded between 5 and 85 °C are shown in Fig. 2A. Most of the fluorescence signal arises from a single buried tryptophan (Trp_{49}) exhibiting an emission maximum at 330 nm in the folded state. Thermal unfolding is accompanied by a significant loss of fluorescence intensity and a red shift in the emission

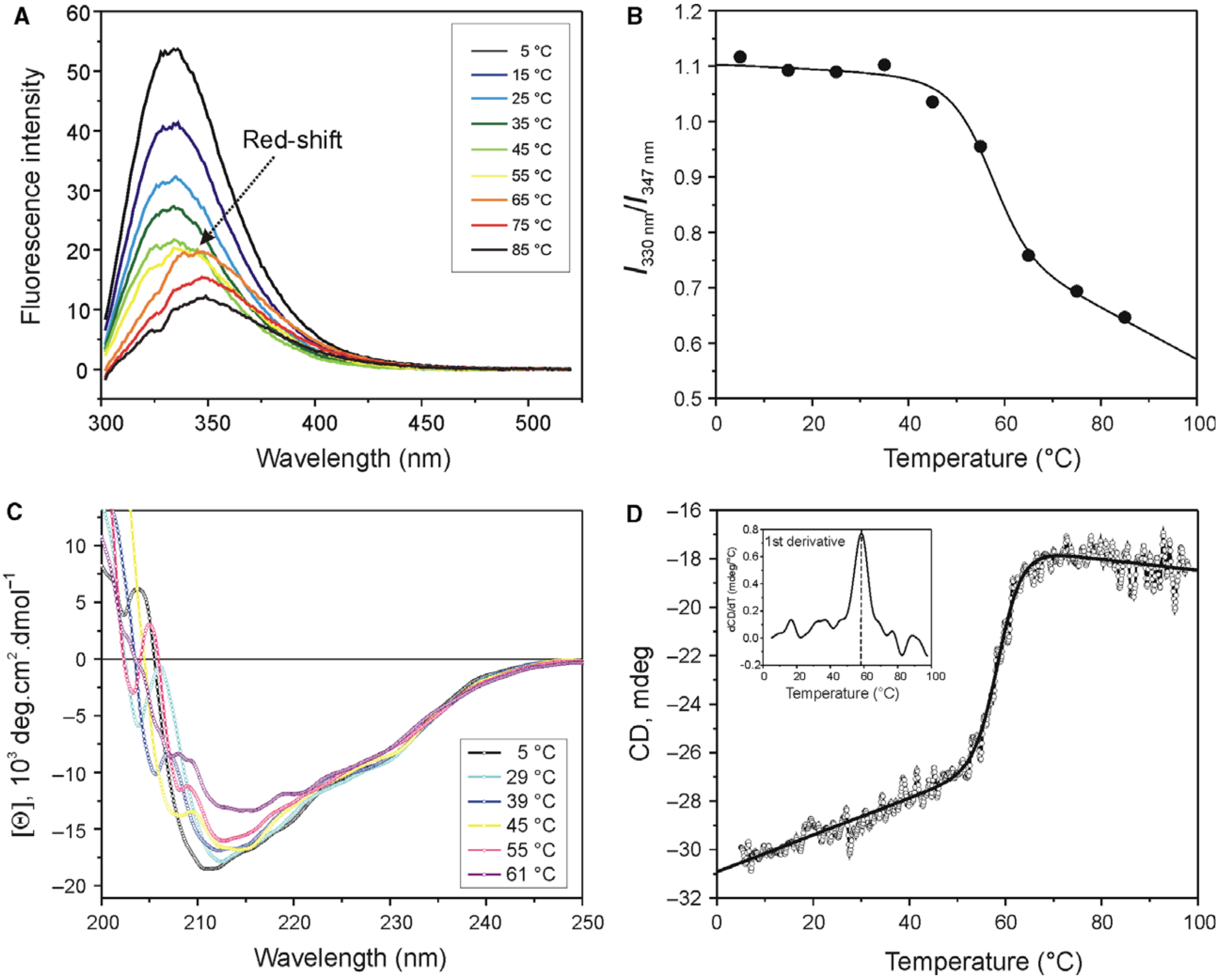

Fig. 2. Fluorescence and CD spectroscopic analysis of thermal unfolding. (A) Fluorescence emission spectra of human I-BABP at different temperatures. Experiments were performed in 20 mM potassium phosphate, 50 mM KCl, 0.05% NaN_3, pH 6.3. Protein concentration was 2 μM. The observed red shift in the wavelength of the emission maximum is indicated by an arrow. (B) Ratio of fluorescence intensities detected at 330 and 347 nm as a function of temperature. The solid line represents a sigmoidal fit assuming a two-state process with parameters $\Delta H = 212\ \mathrm{kJ{\cdot}mol^{-1}}$, $\Delta C_p = 5.6\ \mathrm{kJ{\cdot}°C^{-1}{\cdot}mol^{-1}}$, $T_m = 57\ °C$. (C) Far-UV CD spectra of human I-BABP as a function of temperature. Experiments were performed in the same buffer as in (A) with a protein concentration of 15 μM. (D) Heat denaturation of human I-BABP followed by the CD signal at 215 nm using a scanning rate of 1 °C·min^{-1}. The solid line represents a sigmoidal fit assuming a two-state process yielding a melting temperature of 59 °C. The first derivative of the heat-up profile is shown in the inset.

maximum to 347 nm indicating the exposure of Trp_{49} to the aqueous solvent. The ratio of fluorescence intensities at 330 and 347 nm was fit by a cooperative two-state model yielding values of $\Delta H = 212\ \mathrm{kJ \cdot mol^{-1}}$, $\Delta C_p = 5.6\ \mathrm{kJ \cdot {}^{\circ}C^{-1} \cdot mol^{-1}}$ and $T_m = 57\ ^{\circ}\mathrm{C}$ for the enthalpy and heat capacity change of unfolding and the melting temperature respectively (Fig. 2B). The value of λ_{max} between 65 and 85 °C shifts up to 347 nm, slightly below the wavelength of the emission maximum observed for urea-denatured I-BABP (data not shown) indicating that Trp_{49} may not become completely exposed.

Thermal unfolding of human I-BABP was also monitored by CD spectroscopy. Around room temperature, the far-UV CD spectrum (Fig. 2C) is characteristic of a predominantly β-sheet protein with a helical contribution. Both below and above room temperature substantial changes are observed in the spectral shape, indicating gradual changes in the secondary structure in a broad temperature range. According to the far-UV CD spectra, the most dramatic change in the secondary structure content takes place between 55 and 61 °C. Conformational changes were also monitored at a fixed wavelength of 215 nm at a heating rate of $60\ ^{\circ}\mathrm{C \cdot h^{-1}}$ (Fig. 2D). Fitting the heat-up profile by a cooperative two-state model indicates a main transition at 59 °C. However, based on the derivative of the heat-up curve, the existence of additional smaller transitions is possible both below and above the main transition (Fig. 2D inset). According to the BeStSel algorithm of secondary structure prediction [30], the presence of a substantial amount of residual structure (~ 10% α-helix, ~ 30% β-sheet, ~ 10% turn) is indicated above the melting temperature. Unfolding appears to be nearly completely reversible up to concentrations of ~ 15 μM. At higher concentrations (200 μM), attempted to be used in near-UV CD experiments, signs of protein aggregation start to occur at 53 °C, obscuring a full analysis.

Residue-specific temperature response by NMR and fitting of thermal melting curves

Superimposed $^{1}\mathrm{H}$-$^{15}\mathrm{N}$ HSQC spectra of human I-BABP in the temperature range of 5–54 °C are shown in Fig. 3A. Further increase in the temperature results in a sudden drop in signal intensity accompanied by the appearance of a white gel-like precipitate at protein concentrations of ~ 100 μM required to achieve appropriate signal-to-noise ratio for the temperature-dependent measurements. Examples of chemical shift trends of residues located in various parts and secondary structure elements of the protein are depicted in Fig. 3B. The position of most resonances changes substantially over the investigated temperature range indicating fast exchange of backbone amides between different local environments. This is in agreement with $^{19}\mathrm{F}$ NMR studies of unfolding in 6FTrp-hI-BABP showing no presence of intermediates displaying conformational exchange on the slow-to-intermediate time scale [27] as well as the investigation of 4FPhe-hI-BABP, indicating that several phenylalanine residues are involved in an intermediate structure that is in fast exchange with the unfolded state [31]. We note that this does not mean that slow exchange processes are absent. Merely, due to severe line broadening and low concentration of the corresponding higher energy states they may not become directly observable.

During our investigation, we restricted the analysis to residues whose backbone amide chemical shifts could reliably be determined at least at 10 of the 16 investigated temperatures. We also excluded those, which, due to ring current effects of their own or a proximate aromatic group or the temperature dependence of the $\mathrm{p}K_a$ of their side chain, gave rise to exotic-shaped plots. Inspection of the chemical shift trend of the remaining 106 residues reveals a nonuniform response to thermal stress throughout the protein. Notably, while some of the residues show linear temperature dependence, most of them exhibit a curved temperature response indicating a multistate process [34]. Specifically, under fast exchange conditions, the observed chemical shift (Ω_{obs}) at each temperature corresponds to a population-weighted average of the chemical shifts associated with the individual states. In the case of two exchanging states, this results in a linear shift of cross-peaks as a function of temperature according to $\Omega_{obs} = x_A\Omega_A + x_B\Omega_B$, where Ω_A and Ω_B are the chemical shift of the individual states and x_A and x_B are their mole fraction. If an additional C state is involved in the process with a chemical shift Ω_C differing from both Ω_A and Ω_B and C is in fast exchange with both A and B, the position of the cross-peak will change in a nonlinear way, typically resulting in a curved plot. The deviation from linearity depends on the temperature dependence of the mole fraction of the third state and the relation of Ω_A, Ω_B, Ω_C chemical shifts in the two spectral dimensions.

Residues exhibiting a linear chemical shift trend upon thermal melting are primarily located in the N-terminal βA, helix-I and some of the turn regions (B–C, C–D, F–G), whereas the presence of an additional state is apparent throughout the rest of the protein. Contribution of the latter to the observed chemical shifts was estimated by determination of the maximum curvature of the fitted chemical shift trends using a

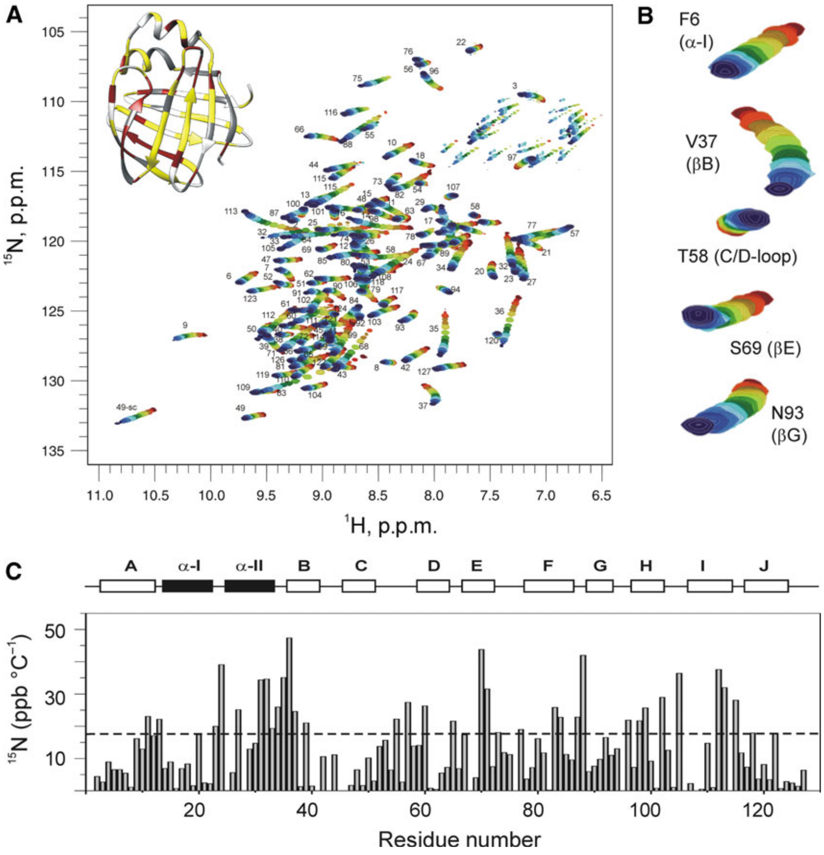

Fig. 3. (A) Series of ^{1}H-^{15}N HSQC NMR spectra of human I-BABP as a function of temperature. Experiments were performed in 20 mM potassium phosphate, 50 mM KCl, 0.05% NaN_3, pH 6.3 at a protein concentration of 100 µM. Temperature increase is indicated by a gradual change in the colour of resonances from blue (5 °C) to red (54 °C). Inset: Backbone amides showing linear (grey) and nonlinear (yellow, red) chemical shift trends in their temperature response mapped onto the ribbon representation of human I-BABP. Residues with above average curvature in their amide NH chemical shift trends are indicated in red. Residues with < 10 nonoverlapping data points were not included in the analysis and are shown in white. (B) Chemical shift trend of selected residues located in various secondary structure elements of the protein. (C) Temperature dependence of ^{15}N shifts in the native state obtained by linear regression of data at 5, 8, 10, 11 and 14 °C. Secondary structure elements are indicated at the top. The dashed line at 17.6 ppb·K^{-1} indicates the maximum slope reported for ubiquitin [32], whose thermal unfolding conforms to a two-state model [33].

Bayesian curve fit. Most prevalent contribution of a third state is observed in the EFGH region of the protein (Fig. 3A inset), with residues Gly_{66}, Ser_{69}, Gln_{72}, Met_{74}, Thr_{78}, Thr_{82}, Leu_{90}, Val_{91}, Val_{92}, Asn_{93} exhibiting the largest curvature.

For fitting the NMR thermal melting data, depending on the linear or nonlinear characteristics of the ^{1}H-^{15}N chemical shift trend of a given residue, two- or three-state transitions were considered assuming fast exchange. The molar fraction of a protein state at a specific temperature was calculated in terms of the Gibbs free energy change in the conformational transition for each different model (see below) as described in Materials and methods. Given the diversity of individual thermal melts, individual fits were performed. Moreover, to ensure that the considered models remain consistent with the physicochemical characteristics of the system, the melting point obtained from fluorescence measurements was used as an initial parameter. For the three-state transitions, two different models were considered. In model 1, an endothermic pretransition is assumed before the main melting point. In model 2, the observed protein aggregation prompted us to explore the possibility of an exothermic process following the melting point. Although the presence of larger aggregates in the solution cannot be excluded, due to their slow rotational correlation time (τ_c), low concentration and conformational heterogeneity, they are not expected to give an appreciable contribution to chemical shifts. Therefore, the observed NMR thermal melting curves must arise from monomers and if exist, small oligomeric species the most. As dimerization has been reported for several members of the iLBP family [35–38], it seemed reasonable to assume dimerization as a primary possible source of an exothermic contribution. We note that in the 10–40 °C temperature range, the value of τ_c (9.6, 8.1, 6.8 and 5.0 ns at 10, 18, 25 and 40 °C respectively), as obtained from ^{15}N T_1, T_2 relaxation and heteronuclear $\{^1\text{H}\}$-^{15}N NOE experiments [14], is in full agreement with a monomeric human I-BABP, suggesting that dimers or higher aggregates, if present, must be transient in nature.

Residues displaying a linear thermal melting fit well to a two-state transition with thermodynamic parameters averaging around $\Delta H^{\text{avg}} = 31 \pm 9$ kJ·mol^{-1}, $T_{\text{m}}^{\text{avg}} = 55 \pm 3$ °C, $\Delta C_{\text{p}}^{\text{avg}} = 50 \pm 70$ J·°C^{-1}·mol^{-1} (not shown). Fits of chemical shifts and populations of the deconvoluted states for two representative residues exhibiting a nonlinear thermal melting are depicted in Fig. 4. As most residues could adequately be fitted assuming a three-state transition using either model 1 (endothermic pretransition with $\Delta H_1^{\text{avg}} = 22 \pm 8$ kJ·mol^{-1}, $T_{\text{m1}}^{\text{avg}} = 29 \pm 1$ °C, $\Delta C_{\text{p1}}^{\text{avg}} = 50 \pm 100$ J·°C^{-1}·mol^{-1} and $\Delta H_2^{\text{avg}} = 29 \pm 8$ kJ·mol^{-1}, $T_{\text{m2}}^{\text{avg}} = 57 \pm 2$ °C, $\Delta C_{\text{p2}}^{\text{avg}} = 130 \pm 130$ J·°C^{-1}·mol^{-1}) or model 2 (exothermic post-transition assuming dimerization with $\Delta H^{\text{avg}} = 66 \pm 7$ kJ·mol^{-1}, $T_{\text{m}}^{\text{avg}} = 58 \pm 3$ °C, $\Delta C_{\text{p}}^{\text{avg}} = 100 \pm 50$ J·°C^{-1}·mol^{-1} and $\Delta H_{\text{dim}}^{\text{avg}} = -93 \pm 10$ kJ·mol^{-1}, $T_{\text{dim}}^{\text{avg}} = 59 \pm 1$ °C, $\Delta C_{\text{pdim}}^{\text{avg}} = 50 \pm 30$ J·°C^{-1}·mol^{-1}, Table S1), solely on the basis of thermal melting curves, it was not possible to differentiate between the validity of the two models. To evaluate the physical relevance of model 1 and 2, the relationship between the parameters inferred

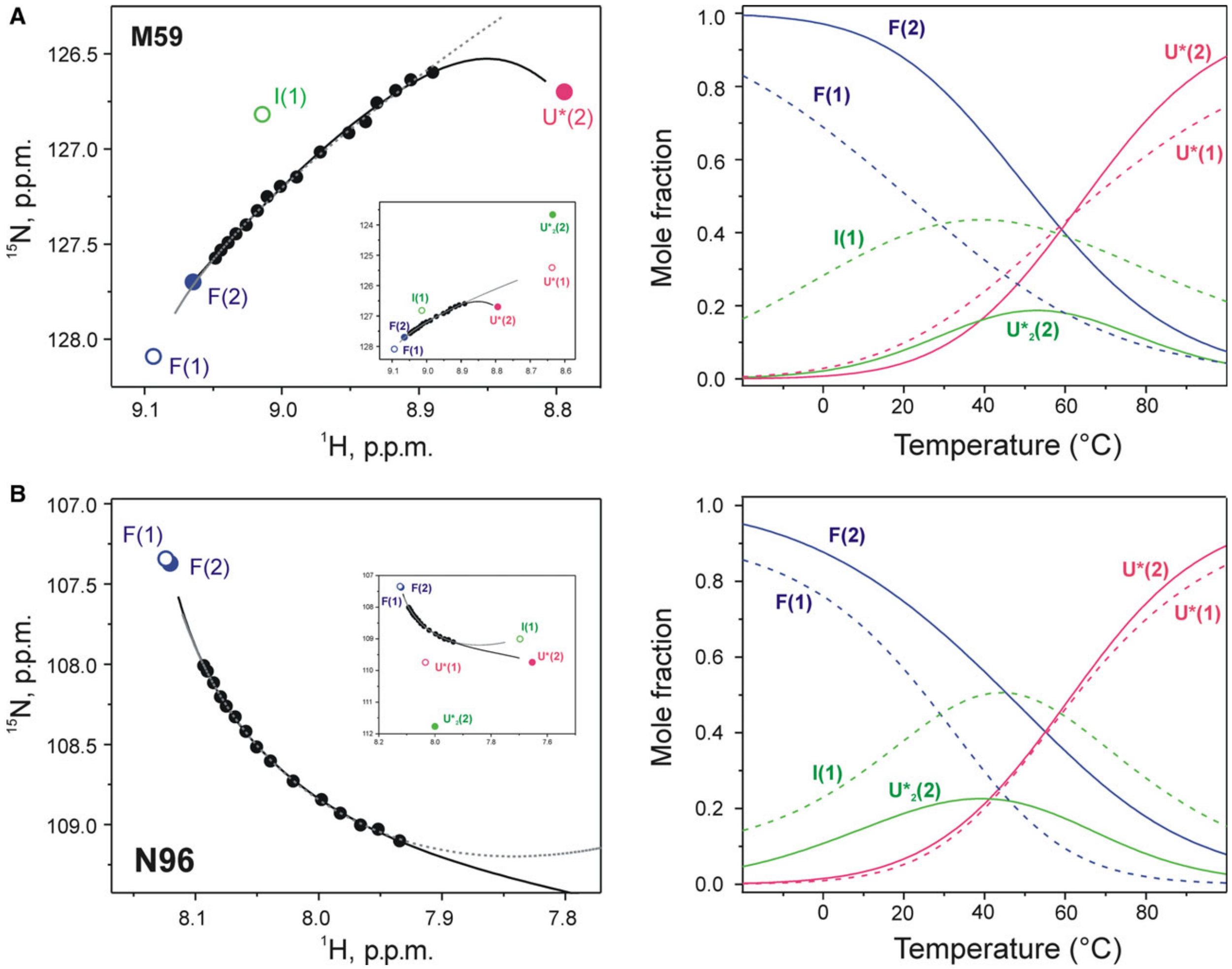

Fig. 4. Examples of nonlinear temperature response of backbone amide chemical shifts as detected in ^{1}H-^{15}N HSQC spectra of human I-BABP. Dashed and solid lines are fits of the experimental data using model 1 ($F \leftrightarrow I \leftrightarrow U^*$) and model 2 ($F \leftrightarrow U^*$, $2U^* \leftrightarrow U^*_2$) respectively. Chemical shifts and populations (right) of the deconvoluted *F* and *U** states are shown in blue and red, whereas the chemical shifts of the deconvoluted *I* and U^*_2 states are indicated in green.

from NMR thermal melting curves and backbone ^{15}N relaxation dispersion measurements was analysed.

Relationship between thermal unfolding and relaxation dynamics

The EFGH protein region exhibiting the most prevalent multistate unfolding behaviour in NMR thermal melting curves has previously been shown to be involved in a millisecond time scale motion in the *apo* form [12,14], and it is part of a hypothetic portal region in BABPs [15]. This prompted us to seek a correlation between the exchanging states observed by NMR relaxation measurements and the conformational transitions reported by the NMR thermal melting curves. Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion (R_{ex}) experiments [39,40] are suitable for the characterization of conformational exchange processes on the microsecond-to-millisecond time scale (τ_{ex} ~ 0.3–10 ms). Dependence of the transverse relaxation (R_2) on an effective field strength (established by variable trains of refocusing spin echoes) is analysed enabling the determination of the exchange rate, $k_{ex} = k_{forward} + k_{reverse}$, and the parameter $\Phi_{ex} = p_A p_B \Delta\omega^2_{AB}$, where p_A and p_B are the populations of the two exchanging states and $\Delta\omega_{AB}$ is the chemical shift difference between the two states. As it has been shown previously, CPMG R_{ex} analysis of *apo* human I-BABP in the 10–18 °C temperature range reports two clusters of residues with slightly different exchange rates (Fig. 5A,B). Values of k_{ex} as determined from a global fit analysis vary from 840 ± 60 s^{-1} at 10 °C to 1540 ± 120 s^{-1} at 18 °C and from 300 ± 40 s^{-1} at 10 °C to 880 ± 100 s^{-1} at 18 °C for the fast and slow clusters respectively [14]. The 'faster' R_{ex}-cluster includes residues (Thr$_{73}$, Val$_{91}$, Asn$_{96}$, His$_{98}$) with the largest contribution to transverse relaxation ($R_{ex} > 10\ s^{-1}$) and comprises primarily the E–F and G–H regions, overlapping with the segments exhibiting nonlinear ^{1}H-^{15}N thermal melting. The 'slower' R_{ex}-cluster involves part of the helical region, the proximate C–D turn, and β-strands B and D. Approaching room temperature, the two clusters merge into a single network of fluctuation with a k_{ex} ~ 2000–3000 s^{-1}, matching the time scale of an initial rate-limiting unimolecular step in ligand binding [41]. As it was noted, the strong entropy–enthalpy compensation observed for the two clusters suggests a disorder–order transition between the ground and the higher energy states of the protein [14].

Intriguingly, a substantial number of residues displaying large-amplitude relaxation dispersions exhibit large ^{15}N chemical shift differences between the deconvoluted states suggesting a relation between millisecond time scale global fluctuations and unfolding transitions in human I-BABP. To explore this possibility, the correlation between Φ_{ex} obtained from R_{ex} measurements versus the chemical shift difference ($\Delta\delta$) of the deconvoluted states obtained from thermal melting curves has been analysed in detail. (We used the parameter Φ_{ex} instead of the R_{ex}-derived chemical shift differences between the exchanging states given the ambiguities in the determination of the latter when k_{ex} ~ $\Delta\omega$.) If the two processes are related, a linear correlation must exist between $\mathit{\Phi}_{ex}$ and $\Delta\delta^2$. As it is depicted in Fig. 5C, a fairly good correlation exists between the parameters inferred from the relaxation and the thermal melting data using the assumptions of model 2. Moreover, the slope of $\mathit{\Phi}_{ex}$ versus $\Delta\delta^2_{F-U^*}$ is in good agreement with the product of the populations of the F and U^* states (blue lines), indicating a direct connection between the global conformational exchange processes and thermal unfolding. Remarkably, in the 10–18 °C range, residues of the two R_{ex}-clusters exhibit two different slopes corresponding to slightly different populations. Accordingly, below room temperature, there is a slight heterogeneity in the protein in the way the unfolding transition is manifested. For model 1, the correlation between the relaxation and thermal melting data is much worse (Fig. 5D). Some correlation between $\mathit{\Phi}_{ex}$ and $\Delta\delta^2_{F-I}$ appears to exist for a subset of 5–10 residues, however, the population of state I is much higher (15–30%, red line) than it would be indicated by the Φ_{ex} versus $\Delta\delta^2_{F-I}$ slope. Accordingly, a joint analysis of the thermal melting and relaxation dispersion data argues for the physical relevance of model 2 and strongly suggests that the deconvoluted U^* state, which appears to have a high propensity to dimerization, corresponds to a conformational state associated with global μs–ms fluctuations in the protein. Amino acid positions exhibiting the most apparent correlation between the two independent experiments appear throughout the protein but primarily in β-strands B, D, and the E–F and G–H regions. We add that from the temperature dependence of relaxation dynamics between 10 and 18 °C, the enthalpy difference between the two exchanging states reported by the R_{ex} measurements is ~ 20–40 kJ·mol^{-1} [14], which in overall is in agreement with the enthalpy change (~ 30–60 kJ·mol^{-1}) associated with the $F \leftrightarrow U^*$ transition obtained from the thermal melting curves using model 2.

We note that there are a few residues (e.g. Tyr$_{53}$, Thr$_{58}$, Gly$_{88}$, Val$_{109}$, Thr$_{113}$) for which there is a significant deviation from the overall good correlation between $\mathit{\Phi}_{ex}$ and $\Delta\delta^2_{F-U^*}$ indicating that the protein likely undergoes a complex array of internal motions. In fact, while in light of the joint analysis of relaxation

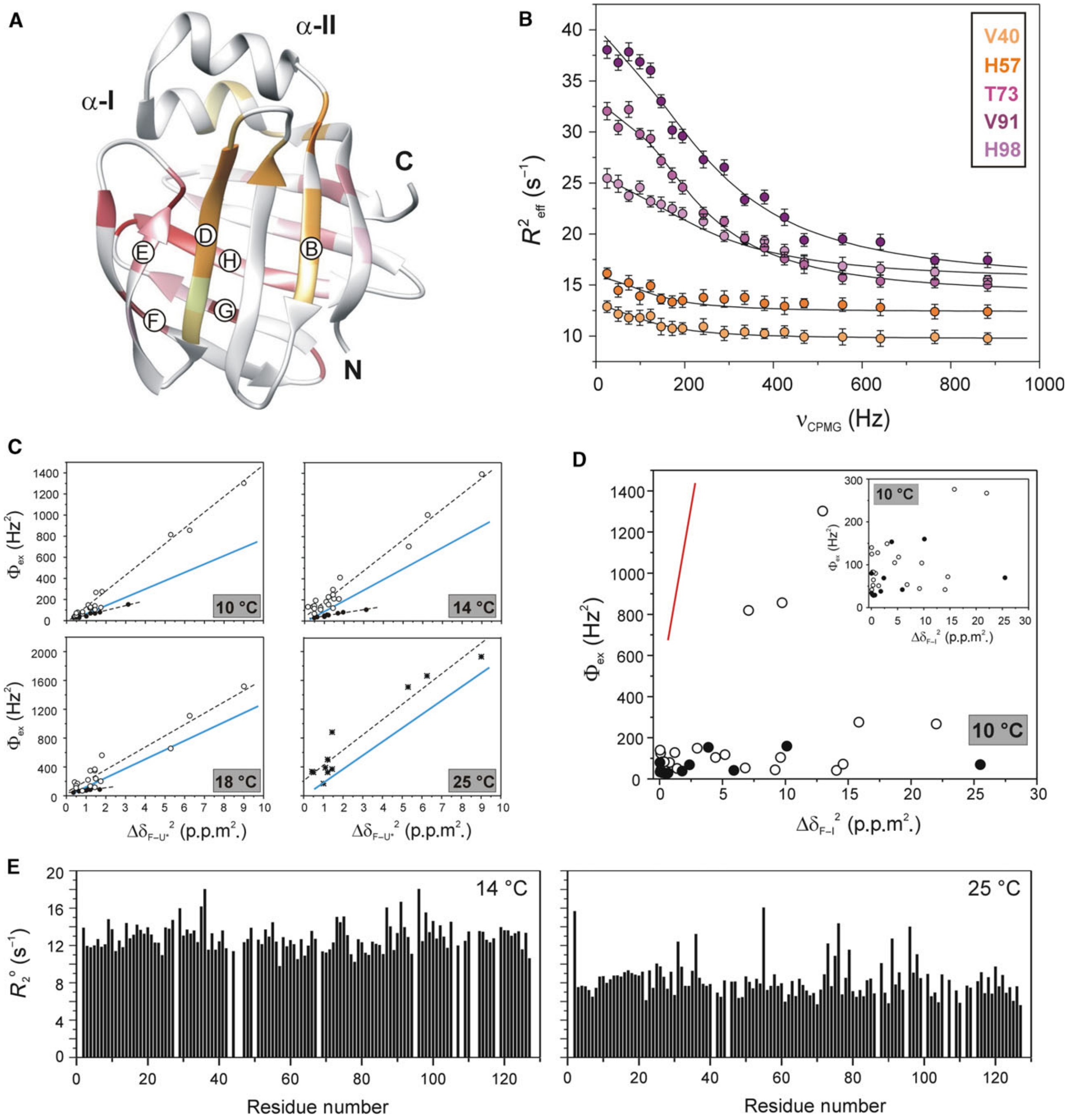

Fig. 5. (A) Residues exhibiting a conformational exchange on the fast end of ms timescale as indicated by CPMG relaxation dispersion experiments at 14 °C (600 MHz for ^{1}H) are mapped on the ribbon representation of the most representative element of the lowest energy structural ensemble of human I-BABP (PDB entry 1O1U [8]). The two clusters with slightly different exchange rates are indicated in orange (slower) and red (faster). Increasingly darker colour indicates larger ^{15}N chemical shift difference between the ground and the higher energy state for each R_{ex} cluster. Residues exhibiting a flat dispersion profile or with no available data are coloured in grey. (B) Transverse relaxation dispersions of the backbone ^{15}N nuclei of representative residues from the slow (orange) and the fast (red) cluster as a function of CPMG B_1 field strength at 14 °C. Solid lines correspond to global two-state exchange models with parameters listed in [14]. Error bars correspond to standard deviations obtained from three parallel measurements (n = 3). (C, D) Φ_{ex} parameters derived from CPMG relaxation dispersion experiments as a function of the squared chemical shift difference between the deconvoluted (C) F and U^* states (model 2) and (D) F and I states (model 1) at different temperatures. Data points for the fast and the slow cluster in the 10–18 °C temperature range are indicated by open and closed symbols respectively. Dashed lines in (C) are obtained by linear regression. Slopes corresponding to the product of the populations of the deconvoluted states are shown in blue ($p_F p_{U^*}$) and red ($p_F p_I$), for model 2 and 1 respectively. (E) Asymptotic values of transverse relaxation rates at high ν_{CPMG} field strength obtained at 14 and 25 °C (600 MHz for ^{1}H).

and thermal NMR data, model 2 appears to be the most adequate, the existence of preunfolding transition(s) cannot be excluded. This is suggested by the strong temperature dependence of chemical shifts of several residues on the native side of the barrier indicative of large changes in the local environment (Fig. 3C). Amino acid positions exhibiting higher than usual temperature dependence in their ^{15}N chemical shifts in the native state are distributed quite evenly throughout the protein, involving segments both in the N-terminal half (α-I, α-II, βB, C–D turn) as well as in the more plastic C-terminal half (E–F, G–H regions, βI). Additionally, the complexity of internal dynamics in the protein is indicated by the asymptotic values of transverse relaxation rates (R_2^0) at high CPMG field strengths at specific amino acid positions (Fig. 5E). Elevated R_2^0 values are manifestations of motions that are slow enough to cause line broadening but too fast to be suppressed by the CPMG pulse train even at high repetition rates [40]. Residues with above average values of R_2^0 are primarily located in β-strand F and the G–H region as well as near the linker connecting the helical cap to the β-barrel highlighting the flexibility of these segments on the μs time scale which becomes more intense with increasing temperature.

Structural characteristics of the *U** state

Based on the substantially smaller enthalpy difference between the *F* and *U** states than the enthalpy change of unfolding indicated by the fluorescence and CD experiments or temperature-dependent equilibrium urea denaturation (data not shown), it is clear that state *U** is a partially unfolded state containing a substantial amount of residual structure. The latter is also suggested by the practically zero heat capacity difference between the *F* and *U** states, as well as the calculated chemical shifts. For a fully denatured protein, due to the loss of dispersion in the chemical environment, amide protons are detected in a narrow chemical shift range. This is shown for human I-BABP in 6 M urea in Fig. 6A. Calculated chemical shifts of the *F* (blue), *U** (orange) and U^*_2 (green) states are depicted in Fig. 6B. While the resonances of many residues move towards the expected ppm range upon going from the pure *F* state to *U**, substantial dispersion remains in the chemical shifts. This is most clearly seen for βA and part of the helical region (Fig. 6C), whereas several segments of the C-terminal half show a significant loss of dispersion in the *U** state (Fig. 6D). Combined (^{1}H, ^{15}N) chemical shift differences between the pure *F* and *U** states along the amino acid sequence are shown in Fig. 7A with secondary structure elements indicated at the top. The presence of a residual structure in the *U** state is most clearly indicated in the helical segment and the N-terminal ABC β-strands. In contrast, regions most affected by temperature increase include the βD–βE and βF–βG regions, in particular the bottom of the β-barrel, as well as the vicinity of the I–J turn.

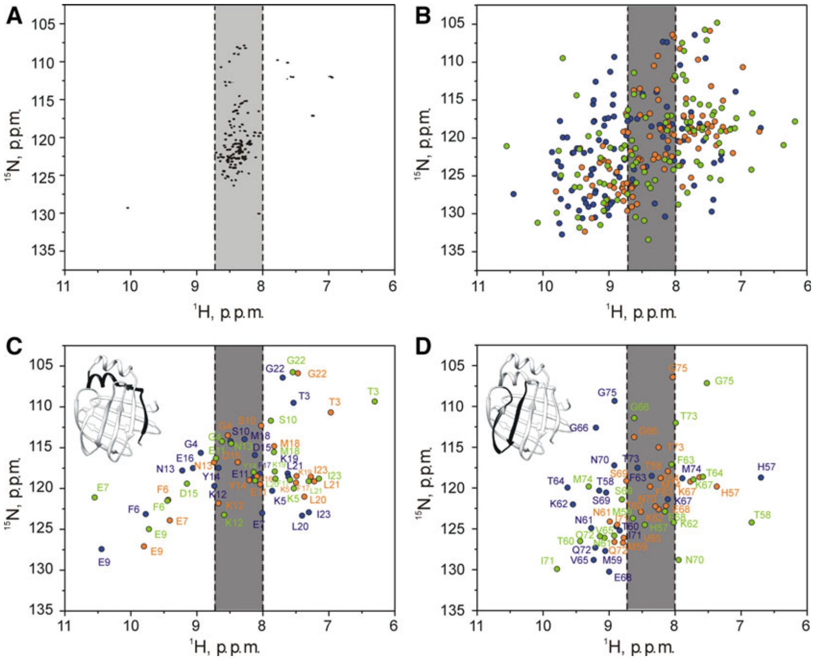

Fig. 6. (A) ^{1}H-^{15}N HSQC spectrum of human I-BABP in 6 M urea at 10 °C. (B–D) Deconvoluted ^{1}H-^{15}N spectra of human I-BABP in the *F* (blue), *U** (orange) and U^*_2 (green) states obtained by fitting the experimental temperature-dependent ^{1}H-^{15}N HSQC spectra of human I-BABP using the three-state model ($F \leftrightarrow U^*$, $2U^* \leftrightarrow U^*_2$) with thermodynamic parameters listed in Table S1. Full spectra are depicted in (B). Spectra of protein segments (C) T3-I23 and (D) H57-G75 are shown with assignment.

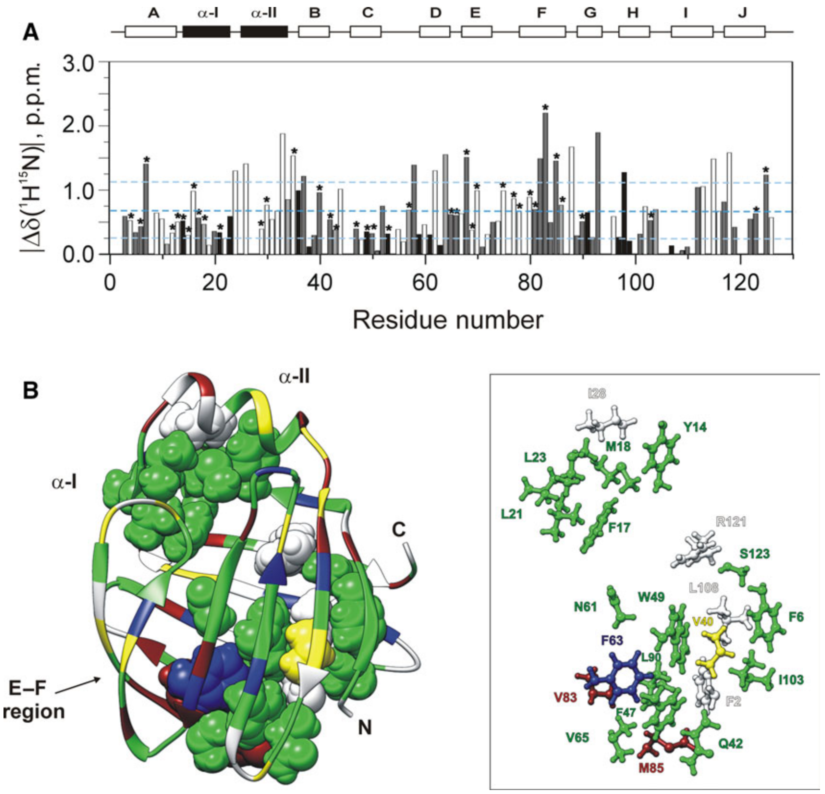

Fig. 7. (A) Chemical shift differences between the deconvoluted *F* and *U** states as a function of amino acid sequence using the three-state model ($F \leftrightarrow U^*$, $2U^* \leftrightarrow U^*_2$) with thermodynamic parameters listed in Table S1. Combined (^{1}H, ^{15}N) chemical shift differences are calculated using the equation of $\Delta\delta_{1HN,15N} = \sqrt{[(\Delta\delta_{HN})^2 + (w_1 * \Delta\delta_N)^2]}$, where w_1 (=0.154) is a weight factor determined using the BioMagResBank chemical shift database [42]. Secondary structure elements are depicted at the top. Dashed lines correspond to the average value of $\Delta\delta_{1HN,15N}$ and $\Delta\delta_{1HN,15N} \pm$ std. Conserved residues on the basis of similarity or identity are marked with an asterisk. Amino acid positions involved in at least two pairwise interactions are indicated by colour coding: light grey = short-range only, grey = long-range only, black = both short- and long-range interactions. (B) Values of $\Delta\delta_{1HN,15N}$ are mapped on the ribbon diagram of human I-BABP [8] with a colour coding where blue to green to yellow to red corresponds to an increasing value of $\Delta\delta_{1HN,15N}$. (Specifically, showing $\Delta\delta_{1HN,15N}{}^{i} < \Delta\delta_{1HN,15N}{}^{mean}$-std in blue, $\Delta\delta_{1HN,15N}{}^{mean}$-std $< \Delta\delta_{1HN,15N}{}^{i} < \Delta\delta_{1HN,15N}{}^{mean}$ in green, $\Delta\delta_{1HN,15N}{}^{mean} < \Delta\delta_{1HN,15N}{}^{i} < \Delta\delta_{1HN,15N}{}^{mean}$ + std in yellow, and $\Delta\delta_{1HN,15N}{}^{mean}$ + std $< \Delta\delta_{1HN,15N}{}^{i}$ in red.) Side-chain atoms of residues involved in multiple conserved pairwise interactions are highlighted. Residues not included in the analysis are shown in white. For clarity, side chains are labelled separately on the right in the same perspective.

Additionally, the two short linkers connecting the two α-helices as well as α-II to the β-barrel display high susceptibility to temperature elevation.

Clusters of conserved pairwise interactions in human I-BABP

To relate the NMR-derived site-specific temperature response to structural determinants of protein stability, conserved short- and long-range pairwise interactions in human I-BABP were analysed. At first, amino acid conservation on the basis of identity and similarity (Materials and methods) has been assessed by multiple sequence alignment of 94 members of the iLBP family with different organism-, tissue- and ligand specificity (Fig. S1). Eleven positions have been identified displaying identity in over 80% of the analysed sequences. This includes five positions where in human I-BABP itself and in some other members of the iLBP family (mostly BABPs) the conserved residue is substituted by an amino acid similar (W6F/Y, I40V) or different (R28I, F69S, D75G) in characteristics. Additionally, 37 positions have been found to show conservation on the basis of similarity. Positions exhibiting sequential identity within the iLBP family are primarily located in two main segments of the protein. The longer segment comprises the N-terminal βA-αI-αII-βB region, whereas the shorter one includes residues in the βD–βE region (Fig. S1). Intriguingly, 5 of the 11 positions are located in linker (Asn$_{13}$, Lys$_{35}$) or turn (Gly$_{66}$, Phe/Ser$_{69}$, Asp/Gly$_{75}$) regions. Conservation on the basis of similarity in the two regions together with the connecting βC and the subsequent βF elements is even more apparent. Despite the structural conservation, substantially less conserved positions can be found in the rest of the β-barrel, specifically, in the GHIJ β-strands. Qualitatively, our analysis agrees with two earlier reports conducted on two smaller data sets [1,43]. In comparison to the analysis by Gunasekaran *et al.* [43] based on the alignment of 52 CR(A)BP and fatty acid-binding protein (FABP) sequences, incorporation of BABPs results in a smaller number of fully conserved residues and in general less conserved positions in the C-terminal half of the protein.

Many of the identified conserved residues are involved in pairwise interactions. Short- and long-range pairwise interactions among the conserved residues in human I-BABP are listed in Table 1. Conserved long-range interactions form two distinct clusters (Fig. 8A). In the more extensive cluster I, a number of residues have multiple contacts tying together several β-strands. Among these networking residues, Phe_2 of βA exhibits the most long-range contacts interacting with residues in six different β-strands (βB (Val_{40}, Gln_{42}), βC (Phe_{47}, Trp_{49}), βF (Met_{85}), βG (Leu_{90}), βH (Ile_{103}), βI (Leu_{108}), followed by Trp_{49} and Leu_{90} interacting with five different strands each. Many residues of the cluster are involved in stabilizing short-range interactions, as well (Phe_{47}-Trp_{49}, Phe_{63}-Val_{65}, Met_{85}-Leu_{90}, Ile_{103}-Leu_{108}). The substantially smaller cluster II is limited to the helical region, in particular, tying together the two helices via multiple interactions of Ile_{28} (α-II) with residues in α-I, as well as with Ile_{23}. The latter is additionally engaged in a contact with Gly_{75}, forming a single conserved tie between the helical cap and the rather loose E–F region. Among the long-range interactions, hydrophobic–hydrophobic contacts are the most prevalent, whereas interactions between two polar side chains or between a hydrophobic and a polar side chain are limited to a few contacts (Table 1). Most importantly, they contribute to the stabilization of βA (Trp/Phe_2-Gln_{42}, Phe_6-Arg_{125}, Glu_7-Lys_{35}, Met_8-Arg_{121}), the helical region (Asp_{15}-Ile_{28}, Lys_{19}-Ile_{28}), the B–C- and D–E turns at the bottom of the β-barrel (Gln_{45}-Val_{65}), and the E–F region (Glu_{68}-Lys_{80}).

Many of the residues involved in conserved pairwise interactions are characterized with below-average chemical shift difference between the NMR-derived *F* and *U** states underlying the importance of these interactions in the thermal stability of the protein (Fig. 7B). Specifically, in cluster I, 13 (Gly_4, Phe_6, Gln_{42}, Phe_{47}, Trp_{49}, Asn_{61}, Phe_{63}, Val_{65}, Leu_{90}, Ile_{103}, Lys_{107}, Ser_{123}, Lys_{124}) of the 19, in cluster II, 8 (Tyr_{14}, Asp_{15}, Phe_{17}, Met_{18}, Lys_{19}, Leu_{21}, Leu_{23}, Tyr_{53}) of the 11 analysed residues display below-average value of $\Delta\delta_{F-U^*}$. On the other hand, both clusters contain a small number of residues (Lys_{35}, Glu_{68}, Gly_{75}, Val_{83}, Met_{85}), which possess a high susceptibility to temperature elevation. Each of these residues is located in peripheral segments of the protein. Lys_{35} is part of a short link covalently connecting the helical segment to the β-barrel, Gly_{75} is within the E–F turn, whereas $Glue_{68}$ (βE), Val_{83} (βF) and Met_{85} (βF) are at the bottom of the β-barrel.

Table 1. Conserved long- and short-range interactions in human I-BABP classified as hydrophobic–hydrophobic (H–H), hydrophobic–polar (H–P), polar–polar (P–P) or glycine–hydrophobic/polar (G–H/P). Interactions with a conservation exceeding 95% are shown in bold. Residues present in BABPs but in > 80% of the considered iLBP sequences are substituted by a different conserved residue are marked with an asterisk.

	H–H	H–P	P–P	G–H/P
Long-range interactions	F2–V40	F2–Q42	E7–K35	G4–R125
	F2–F47	**F6–R125**	E68–K80	I23–G75*
	F2–W49	M8–R121	K107–K124	
	F2–M85	D15–I28*		
	F2–90	K19–I28*		
	F2–I103	Q45–V65		
	F2–L108	W49–N61*		
	F6–V40	L108–S123*		
	F6–I103			
	F6–L108			
	Y14–I28*			
	M18–I28*			
	V40–F47			
	V40–W49			
	V40–L108			
	F47–F63			
	F47–V65			
	F47–M85			
	W49–F63			
	W49–L90			
	W49–L108			
	F63–V83			
	F63–L90			
	V65–V83			
	V65–M85			
	V65–L90			
	V83–L90			
Short-range interactions	**Y14–M18**	Q42–F47	R121-S123*	
	F17–L21	Y53*–H57		
	M18–I23	N61*–F63		
	L21–I23			
	I23–I28*			
	F47–W49			
	F63–V65			
	A81–V83			
	M85–L90			
	I103–L108			

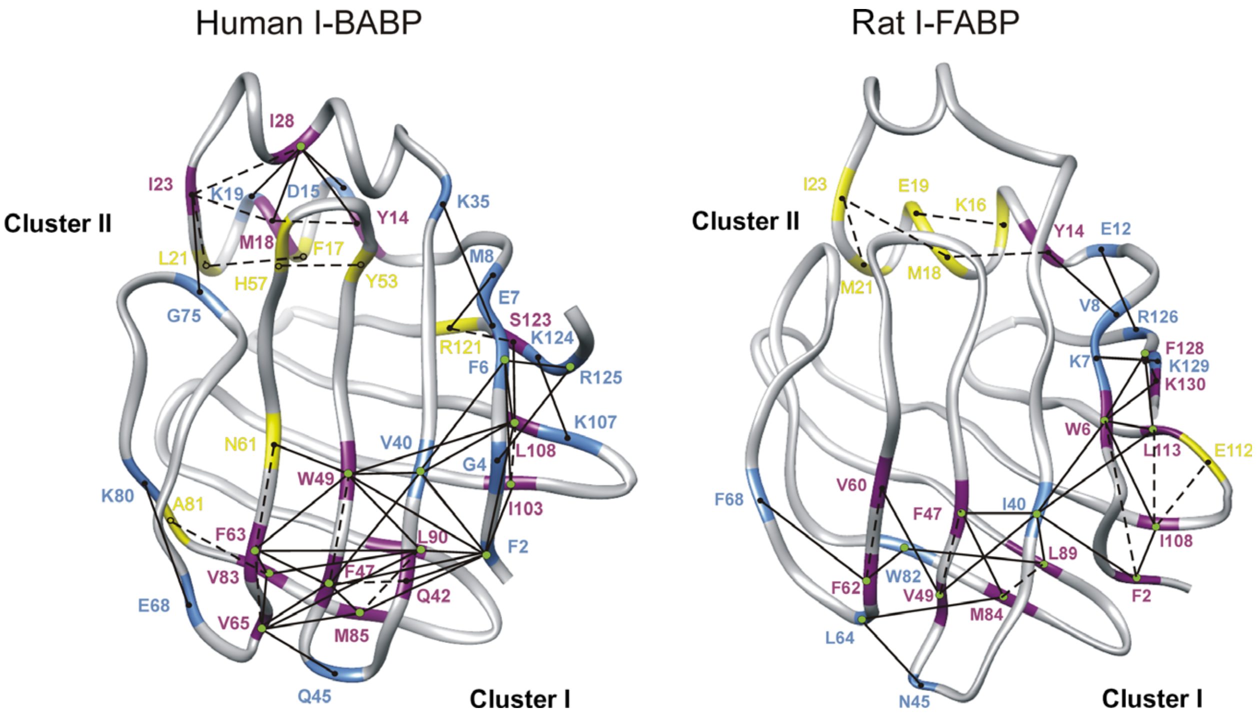

Fig. 8. Ribbon representation of human I-BABP (PDB: 1O1U [8], left) and rat I-FABP (PDB: 1AEL [44], right) showing the conserved short- (dashed line) and long-range (solid line) interactions. Residues involved in short-range, long-range, or both short- and long-range interactions are depicted in yellow, blue and magenta respectively. Residues involved in multiple long-range interactions are marked with a green dot.

Cluster analysis of individual unfolding events

Amide proton chemical shifts sensing to a large extent the H-bonding status of the peptide bond [45] have extensively been used to monitor the thermostability of protein regions [32] and gain insight into the source of cooperativity in folding processes [46]. To quantify the degree of similarity between unfolding events at the residue level, amide proton chemical shifts were normalized as described in Materials and methods and the resulting thermal melting profiles were subjected to cluster analysis. We note that as NMR data could be collected up to slightly below the melting temperature, our analysis primarily provides information on the early stages of unfolding. Accordingly, cluster analysis of thermal melting in the 5–54 °C temperature range indicates three main groups of residues with distinct temperature dependences (Fig. 9). The major cluster (blue) includes most segments of the N-terminal β-sheet, some regions of α-I, as well as βG, βH and βJ of the C-terminal half. A set of additional 12 residues (orange) located primarily in turn regions and at the termini of secondary structure elements comprises the smallest among the three main groups. Finally, a fairly large cluster of residues (yellow) displays similar thermal melting profiles for residues in βD continuing in the E–F region together with specific segments in the helical cap. Chain of pairwise interactions appears to have a key role in transmitting information between distant sites. Specifically, interactions of α-I with residues in the I–J β-strands (Ser_{112}, Ile_{114}, Val_{117}, Tyr_{119}) followed by subsequent strand–strand interactions (βI–βB, βB–βC–βG) appear to be a key element of structural integrity. Likewise, by forming contacts with α-II and the loop connecting the two helices, the E–F turn (Thr_{73}, Gly_{75}) may have an important role in the propagation of order/disorder between the helical cap and the bottom of the barrel.

Local and global unfolding of human I-BABP addressed by molecular dynamics simulations

To study further the internal motions and local and global unfolding of human I-BABP, the solution NMR structure (PDB: 1O1U, [8]) was subjected to MD simulations in explicit water at various temperatures of 300, 350, 370 and 400 K. Although the 1-μs simulations (400 ns at 400 K) are at the short end of the time scale of the NMR studies, the results complement the NMR experiments revealing important dynamic features of the molecule. Moreover, carrying out the simulations at elevated temperatures, the 1 μs

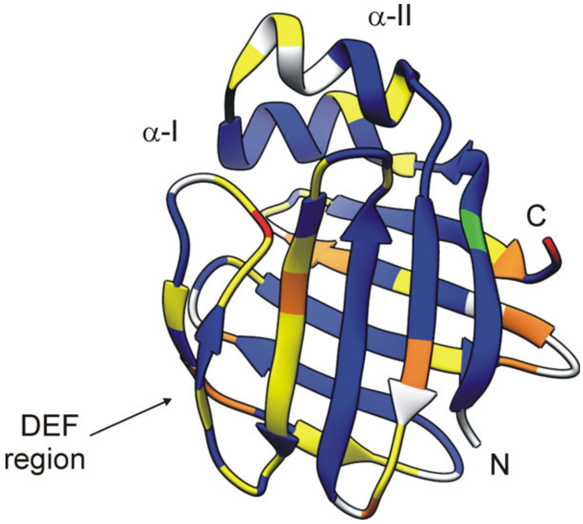

Fig. 9. Clustering of regions in human I-BABP based on the assessment of similarity between individual amide ^{1}H thermal melting profiles. Chemical shift changes as a function of temperature were normalized to the difference between the experimentally detected chemical shift and the calculated random coil value at 5 °C. The analysis was based on calculating the pairwise root-mean-square deviations between normalized thermal unfolding profiles. Clustering the data into six clusters yields three main groups of residues (blue, yellow, orange) reflecting similarities and differences in the temperature response of specific protein regions. Four additional residues separating from the rest are depicted in red (Met$_{74}$, Ala$_{127}$), green (Glu$_7$), and black (Ile$_{23}$). Residues with less than ten temperature points were not included in the analysis and are shown in white.

time was sufficient to observe local and global unfolding events of the protein. Results of MD simulations are presented in Figs 10 and 11 and in Movies S1–S6 of the Supplementary Information. The MD trajectories were subjected to the following analyses: (a) secondary structure changes upon the MD trajectories were calculated by DSSP algorithm [47]. (b) The trajectories were analysed to find out which regions of the protein exhibit correlated motions by the Bio3D R-package cross-correlation analysis module [48]. (c) By clustering the changes of the pairwise distances between Cα atoms, we determined the regions of the molecule that move together as rigid parts. To focus on the slow motions, the fast fluctuations were averaged out by 5 and 10 ns smoothing (averaging) on the Cα positions. (d) 3D structures corresponding to characteristic points of the simulations were analysed and the polar and apolar solvent-accessible surface areas were calculated.

At 300 K, the structure of human I-BABP is fairly stable with fluctuations in the helical region (α-I, α-II) and at the region of the E–F strands (Fig. 10A). Correlated motion analysis with principal component analysis resulted in cross-correlation matrices showing regions with positive and negative correlations as well, revealing the 'breathing' motion of the helical region (α-I, α-II) and the E–F strands (Fig. 11A,B). This is concordant with earlier results on conformational changes occurring upon ligand binding [9]. Movies on the conformational changes corresponding to the first two principal components are presented as supplementary material (Movies S1 and S2). When decomposing the together-moving residues into two clusters, the E–F region together with the bottom of the β-barrel is separated from the rest of the molecule. Decomposing into three clusters shows the separation of the helical cap (not shown). These results are consistent with the NMR experiments.

At 350 K, fluctuations are observable in the helical cap, and they are more pronounced in the DEF region. The analysis for correlated motions points out an increased breathing of the helical cap and the E–F region (Fig. 11C, Movies S3 and S4). Noteworthy, in the open conformation, a significantly increased exposure of the inner hydrophobic surface is observed. Clustering of the together-moving regions shows the separation of the E–F region together with the bottom of the β-barrel from the rest of the molecule (Fig. 11D). Further splitting brings out the helical cap and clustering into four groups separates the loose turn between the E and F strands (Fig. 11E,F).

At an even higher temperature of 370 K, after 400 ns simulation a locally unfolded intermediate is observed with fully opened and unfolded conformation at the region of the E–F strands (Figs 10B and 11G, H). This conformation is preserved for another 400 ns followed by the uptake of a β-structure in the corresponding region, however, with an altered β-sheet architecture. Specifically, βF separated from βE, rejoins the C-terminal β-sheet by forming stabilizing contacts with βG. The rearrangement of the E–F region is accompanied by a shortening of βA. This 'unfolding intermediate' exhibits largely increased hydrophobic surface which can provide a propensity for intermolecular interactions and aggregation (Fig. 11H). Correlated motion analysis points out the local unfolding event of the E–F region as a main component (data not shown).

Finally, at 400 K, we observed large fluctuations in the helical cap, and after 160 ns, a local unfolding of the E–F strands. After 200 ns, starting at the E–F region and spreading onto the entire molecule, a complete unfolding of the molecule takes place within the next 100 ns (Fig. 10C).

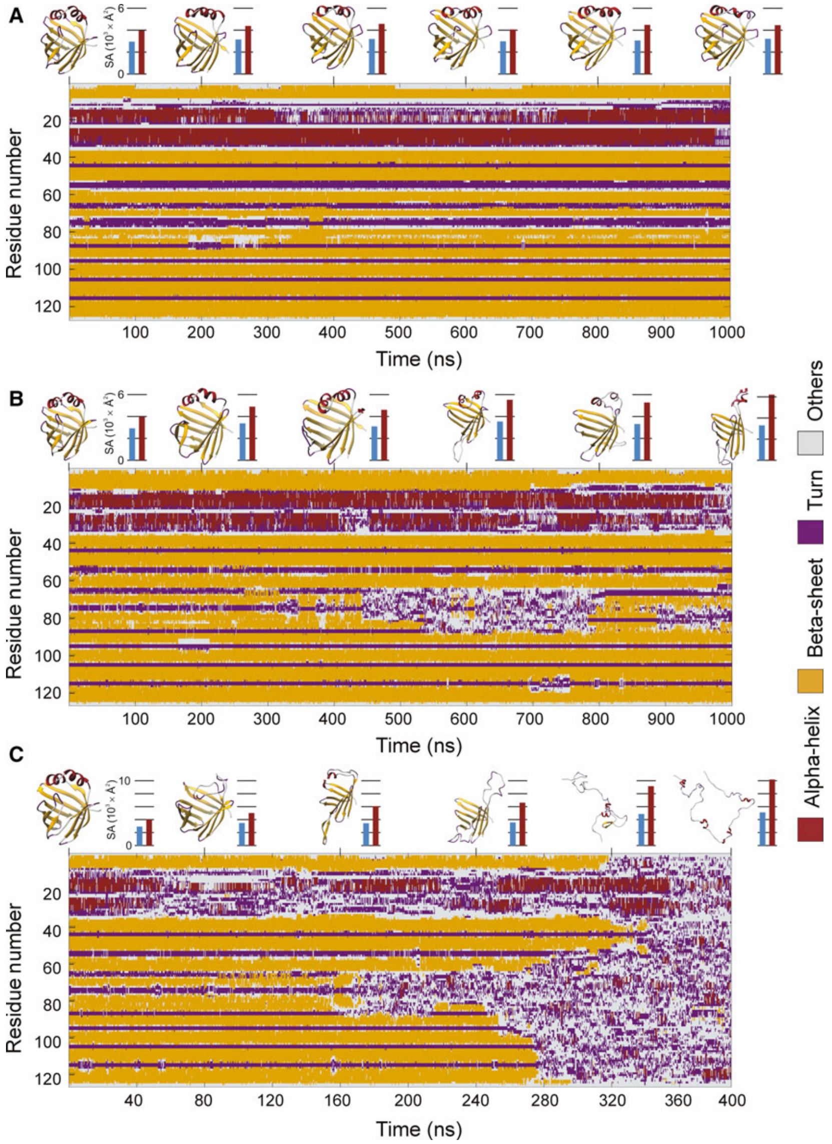

Fig. 10. Fluctuation of the secondary structure of human I-BABP as revealed by MD simulations. Secondary structure composition of human I-BABP upon 1 μs MD simulation at 300 K (A), 370 K (B) and 400 ns at 400 K (C) with 3D structures and solvent accessible polar (blue) and apolar (red) surface areas corresponding to every 200 ns (panel A, B) and 80 ns (panel C) time-point shown at the top. The structure of the protein is fairly stable with fluctuations in the helical region (residues Asn_{13}-Asn_{33}) and at the region of the E–F strands (residues Glu_{68}-Met_{85}). At 370 K, after 400 ns, a locally unfolded intermediate is formed with fully opened conformation at the region of the E–F strands (panel B). This conformation is preserved for another 400 ns and then the corresponding region uptakes a β-structure, however, in an altered β-sheet architecture, where βF separated from βE rejoins the C-terminal β-sheet by forming stabilizing contacts with βG (see the structure at 800 ns). The rearrangement of the E–F region is accompanied by a shortening of βA. In the locally unfolded conformations, a significantly increased exposed hydrophobic surface is observed. At 400 K (panel C) increased fluctuations are observed in the helical regions and at the E–F strands. After 200 ns, starting in the E–F region, a global unfolding process takes place which is completed around 300 ns.

Discussion

Cooperative unfolding and characteristics of the partially unfolded *U** state

Proteins trapped in partially unfolded states can undergo uncontrolled self-association, a well-known source of impaired functionality and disease states [49,50]. In other cases, partial unfolding of specific protein regions has functional implications in processes such as ligand binding [51], signal transduction [52], chaperone activity [53], channel gating [54]. iLBPs are predominantly β-sheet proteins in which order–disorder transitions are thought to have an important role in ligand uptake and release hence in intracellular lipid transport and targeting, as well as ligand-mediated regulation of the transcriptional activity of nuclear receptors. Our joint NMR thermal melting and relaxation analysis complemented by MD simulations indicates a complex internal dynamics for human I-BABP, an extensively studied member of the iLBP family. According to our data, thermal unfolding of the protein around 58–59 °C gives rise to global fluctuations at the fast end of the ms time scale below room temperature (10–18 °C) between the ground state and two higher energy states with slightly different populations and exchange kinetics, which at higher temperature (25 °C) merge. The higher energy state(s), in the 10–25 °C temperature range present at a few per cent, are related to a partially unfolded state with a substantial amount of residual structure in the N-terminal half

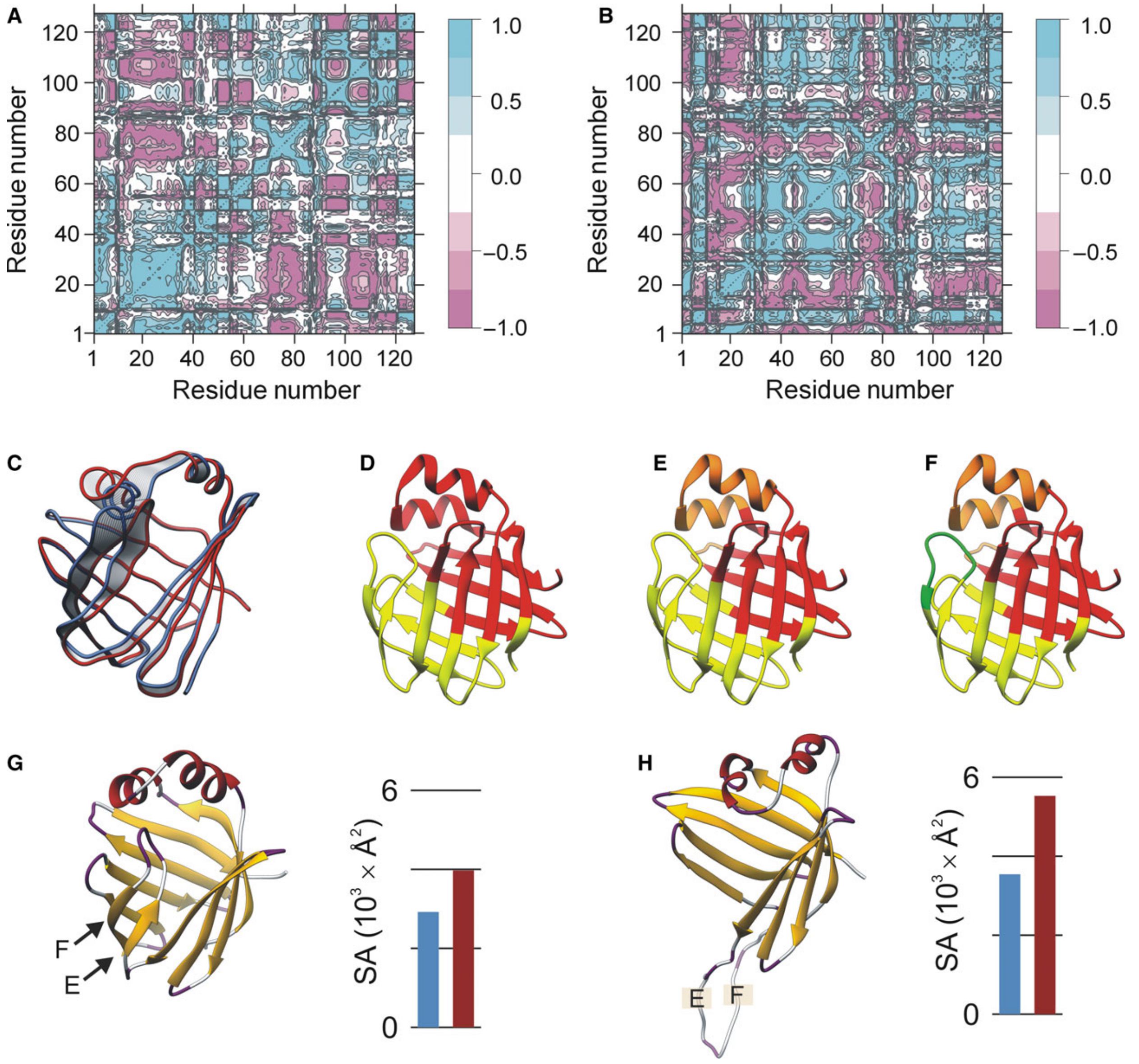

Fig. 11. Dynamics of human I-BABP as revealed by MD simulations. (A, B) Cross-correlation matrices corresponding to the first and second principal components of correlated movements in human I-BABP upon MD trajectory, respectively, at 300 K. Blue colour indicates positive and magenta indicates negative correlation. (C) Human I-BABP structures characterizing the first principal component of the correlated motions at 350 K. Grey ribbons show the transition between the two extremes (blue and red) reflecting a closed and an open conformation. (D–F) Clustering of regions for the 'slow' motions that move together at 350 K into two, three and four clusters respectively. (G) Energy minimized structure of the molecule at the beginning of the simulation at 370 K, and (H) locally unfolded intermediate observed around 600 ns. The corresponding polar (blue) and apolar (red) solvent-accessible surface areas are shown.

(primarily in the helical region and the ABC β-strands) and more disordered protein segments in the C-terminal half (primarily in the DEF region). The direct relation of thermal unfolding to a global exchange process in a wide temperature range rules against a downhill folding mechanism in which local structure would be expected to melt independently with no substantial correlation between the relaxation data and the spectral properties of the partially unfolded state (line broadening induced by conformational fluctuations versus chemical shift difference between the *F* and *U** states). Accordingly, our joint analysis of NMR thermal melting and relaxation data strongly suggests a cooperative folding/unfolding mechanism for human I-BABP, where the energy landscape is dominated by a single barrier with a relatively ordered ensemble on one side (native state) and less ordered, more heterogeneous conformations on the other side (denatured state). Cooperativity is also indicated by the correlated motions revealed by MD simulations involving the DEF region and the helical cap, as well as the overlap of these segments with a cluster of residues displaying a highly similar thermal melting profile (Fig. 9, yellow). We note that our findings do not rule out the possibility of intermediates along the unfolding pathway. In fact, the strong temperature dependence of chemical shifts in the native state (Fig. 3C) and the unfolding-related two distinct fluctuations with slightly different exchange rates below room temperature (Fig. 5A,C) indicate a complex pattern of internal dynamics for human I-BABP where the presence of intermediates is still possible.

According to our data, the partially unfolded U^* state has a high propensity to undergo dimerization. This can be most easily envisioned by a conformational rearrangement in the most vulnerable DEF protein region (characterized with the most prevalent $\Delta\delta_{F-U^*}$ chemical shift differences) giving rise to the exposure of hydrophobic patches in the U^* state. This is strongly suggested by MD simulations where the caught partially unfolded conformational state (Fig. 10C, top) exhibits a significantly increased hydrophobic surface. We note that dimerization via hydrophobic interactions has been indicated for iLBPs such as kidney and liver FABPs in the past [35,38]. It should be emphasized that as rotational correlation times correspond to a monomeric human I-BABP [14], the formation of dimers must be transient in nature. Transient association of the monomeric form has also been suggested for chicken liver (cl) BABP [26], where the formation of transient aggregates have been proposed to account for the singular behaviour of a subset of residues in urea denaturation experiments. The affected residues in chicken liver bile acid-binding protein (cL-BABP) are mainly located in the G–H β-strands as well as in the D–E, E–F and F–G turns that is overlapping with segments exhibiting a pronounced contribution of a third state in NMR thermal melting in the human analogue. We note that preliminary stopped-flow kinetic refolding experiments of human I-BABP in the presence of a decreasing amount of urea indicate conformational heterogeneity in the denatured state with a rate-limiting kinetic step accounting for ~ 15% of the observed fluorescence change, possibly corresponding to the dissociation of a dimerized state before refolding can take place (data not shown).

According to the chemical shift differences between the deconvoluted F and U^* states, the most resistant sites to temperature elevation are regions involved in multiple pairwise interactions tying together different secondary structure elements (Fig. 7B). However, both conserved clusters contain a small number of primarily peripheral residues (e.g. Lys_{35}, Gly_{75}, Val_{83}, M_{85}), which exhibit high susceptibility to temperature rise. It has been suggested that such protein sites of increased vulnerability to physicochemical conditions (denaturant, temperature) may act as unfolding initiation sites [55]. As indicated by our NMR and MD simulation results, differences in dynamic properties on the µs–ms time scale are translated into different susceptibility to temperature stress. Accordingly, we propose that the highly dynamic, loose D–E turn and E–F regions are likely the first to unfold in human I-BABP. Regarding the spreading of disorder, peripheral residues of conserved clusters in mobile regions exhibiting a high susceptibility to temperature elevation may have a key role in the propagation of unfolding. This has implications for both the mechanism of unfolding and the onset of self-association.

Comparison with the folding/unfolding behaviour of other iLBPs

Most kinetic studies in the presence of denaturant indicate a multistate folding/unfolding mechanism of iLBPs [21,22,25,27]. Intriguingly, for human I-BABP, fluorescence and CD kinetic analysis of urea-induced unfolding reports a single phase [27]. Our NMR thermal melting and relaxation data in the absence of denaturants indicate a complex pattern of internal dynamics for the protein with a dominating single barrier and a likely presence of rapidly exchanging conformational substates along the unfolding pathway. Moreover, our data show that before complete unfolding, human I-BABP is trapped in a partially unfolded state containing a significant amount of residual structure. The U^* state with retained secondary and tertiary interactions in response to thermal stress might share common features with urea-induced molten globule-like unfolding intermediates reported for CRABP-I [21] and rat I-BABP [27] by stopped-flow kinetic measurements.

Several of the most resistant sites to temperature elevation in human I-BABP (i.e. residues exhibiting low values of chemical shift difference between the F and U^* states) correspond to conserved amino acid positions, which have previously been indicated to be part of folding initiation sites in BABPs and related proteins. The most illustrative examples include Phe_{47} (βB) and Phe_{63} (βC). These two amino acid positions have been proposed to be part of a folding initiating site in human I-BABP [31] and belong to a cluster in I-FABP, which by stopped-flow fluorescence has been suggested to act as a nucleation centre for the propagation of β-strands [22].

One of the major differences between hI-BABP and I-FABP appears to be related to the local stability of the N-terminal β-strand. While in I-FABP, the middle of βA has been reported to be completely unfolded at denaturant concentrations where residual structure is still present elsewhere [56], in hI-BABP, the entire strand possesses low susceptibility to temperature elevation. The different folding behaviour of βA in I-BABP and I-FABP are likely related to differences in long-range interactions of βA between the two proteins. In I-FABP, βA is primarily engaged in interactions with the two C-terminal β-strands (Fig. 8B), whereas in I-BABP it forms additional contacts with six other strands (βB, βC, βF, βG, βH and βI) via hydrophobic interactions. Additionally, there is a nonconserved electrostatic interaction between Lys_5 and Glu_{39} (βB) in hI-BABP, with Glu_{39} being further engaged in a salt bridge with His_{52} (βC). We note that localized residual structure in the three N-terminal β-strands has also been suggested for the chicken liver BABP homologue in the presence of high concentrations of urea [26].

Another important difference between hI-BABP and FABPs involves the helical region. While in FABPs, the helical region has been shown to have no significant role in protein stability or the mechanism of the formation of the β-barrel [57,58], recent studies of rabbit I-BABP [59] and human I-BABP indicate a possible role of the helices in the folding process. Specifically, an ^{19}F NMR investigation of human I-BABP unfolding reports an exceptionally high chemical shift change for Phe_{17} between 4 and 8 M urea, raising the possibility of its involvement in an early folding intermediate [31]. Our residue-specific investigation of the thermal response supports the role of the helical segment in protein stability. Specifically, a substantially below-average $\Delta\delta_{F-U^*}$ throughout α-I indicates a residual structure in the region before global unfolding. While stabilization of α-I by short-range pairwise interactions in hI-BABP and I-FABP is rather similar, long-range interactions of Ile_{28} with multiple residues in α-I provides an additional source of stabilization in hI-BABP. Furthermore, as mentioned above, side-chain interactions between residues Tyr_{14}, Phe_{17}, Leu_{21}, Ile_{114}, Val_{117} and Tyr_{119} couple the helical cap to the C-terminal segment of the β-barrel, specifically to βI.

Unlike the helical segment and the three N-terminal β-strands, the DEF region of human I-BABP exhibits $\mathit{\Delta}\delta_{F-U^*}$ chemical shift differences substantially exceeding the average (Fig. 7). The exceptions include two conserved residues on βD (Phe_{63}, Val_{65}) and a short segment in the E–F turn (Ile_{71}-Met_{74}). The D–E region of iLBPs is unique in the sense that there are no hydrogen bonds between the backbone atoms of the adjacent strands and the source of stabilization is limited to van der Waals contacts between hydrophobic side chains (in hI-BABP: Phe_{63}, Val_{65} (βD), Ile_{71}(βE) and Val_{83}(βF)). Similar to hI-BABP, residues stabilizing the DE-region in I-FABP are part of a hydrophobic cluster associated with an initial hydrophobic collapse during the folding process [60]. Also, Leu_{64} of the D–E turn has been found to have a crucial role in the final stabilization of the β-barrel in rat I-FABP [19]. Intriguingly, while in I-FABP there are additional nonconserved hydrophobic residues in βD and βE stabilizing the region (**VVF**ELG**V**D F**AY**S**L**), in hI-BABP many of them are substituted by a polar side chain (NK**F**T**V**GKESN**I**QT) suggesting that in hI-BABP the D–E region is substantially less stabilized.

Functional implications of variations in local stability

The mechanism of ligand entry and exit in iLBPs has been a subject of interest for many years [1,15]. Disorder–order transitions are thought to play a key role in the regulation of ligand entry and release. According to the so-called dynamic portal hypothesis, a localized region of backbone disorder exists in the helical cap in the *apo* form of rat I-FABP, which together with the proximate C–D- and E–F turns acts as a flexible gate permitting ligand entry [44,61]. Weakened capping interactions of α-II has been proposed to be responsible for the disorder in the *apo* state, whereas the binding of palmitate induces a series of interactions between α-II and the neighbouring C–D turn stabilizing the last four residues of the helix and shifting the order-disorder equilibrium towards a more ordered, closed state [61]. In the related CRBP I and II proteins, while there is no sign of increased picosecond mobility in the absence of ligand, CPMG relaxation dispersion experiments have shown the presence of conformational fluctuations in the helical region [62]. Moreover, in *apo* CRBP II, the increased values of conformational exchange rates and ^{15}N chemical shift differences between the exchanging populations in α-II are parallel with a structural disorder in the region [62]. Unlike in FABPs and some of the reported CRBP structures, studies of chicken liver and human ileal BABPs show evidence of a conformational equilibrium between a closed and a more open protein state with the involvement of the E–F and G–H protein regions allowing access to the otherwise enclosed binding cavity [13,14]. In cl-BABP, protonation/deprotonation of a nearby histidine has been found to be closely related to the opening/closing transition [13], and in both proteins a conformational selection mechanism of ligand binding is suggested by NMR dynamics and structural data [9,63]. The increased vulnerability to temperature elevation of the DEF region in human I-BABP revealed by our NMR spectroscopic investigation indicates that under specific conditions (proton gradient, collision with lipid bilayers), segments in this protein region are likely the first to undergo a partial unfolding process. This can have implications for ligand entry and release modulated by interaction with the cell membrane of enterocytes and hepatocytes during enterohepatic circulation [3]. Importantly, the partial unfolding accompanied by membrane interaction is thought to be a reversible process modulated by the presence of ligands [17]. In line with this, the stabilizing role of bile salts is supported by our observation that when monitoring intrinsic tryptophan fluorescence, the *holo* form of human I-BABP melts at approximately one degree higher temperature than the unligated form (data not shown) and that millisecond fluctuations, which according to our results, are closely related to a partial unfolding process, cease upon ligand binding [12].

The protein sites exhibiting increased susceptibility to temperature elevation can also have implications for ligand-mediated nuclear translocation mechanisms and the enhancement of the transcription activities of nuclear receptors [4]. Specifically, the presence of nuclear localization signals and nuclear export signals have been reported to mediate nuclear translocation in FABP4 [64] and CRABP-II [18], and it has been suggested to be a general mechanism of signalling in iLBPs [64]. Nuclear import is in general thought to be mediated by three basic residues in the helix-loop-helix region, whereas appropriately positioned hydrophobic residues at the bottom of the β-barrel [64] have been found to be critical for nuclear export. According to our results, both the linker connecting the helical region to the β-barrel and segments of the DEF β-strands show an enhanced plasticity in human I-BABP and may have a role in exposing the corresponding basic (e.g. Lys_{19}, Lys_{30}, Arg_{32}) and hydrophobic (e.g. Val_{65}, Leu_{90}, Val_{91}) side chains in a geometrical arrangement appropriate for signalling.

In conclusion, NMR spectroscopy in conjunction with CD and fluorescence spectroscopies and MD simulations were used to gain insight into the relationship between internal dynamics and protein stability in human I-BABP, a predominantly β-sheet protein, where aggregation poses a serious challenge. One of the main conclusions of our work is that there is a direct structural and thermodynamic relation between the main unfolding transition and a global μs–ms conformational fluctuation observed below the melting point in a wide temperature range. This strongly indicates a cooperative mechanism of unfolding for human I-BABP noting that the protein displays a rather complex pattern of internal dynamics with a likely presence of conformational substates and even the possibility of intermediates along the unfolding pathway. Besides arguing for cooperativity, the joint analysis of NMR thermal melting and relaxation dispersion data reveals the existence of a partially unfolded U^* state trapped before protein aggregation with a high propensity for dimerization. According to NMR measurements and MD simulations, non-native conformation responsible for the onset of self-association is primarily located in segments of the DEF region. The identified regions of high plasticity can have a key role in regulating ligand binding and release, as well as the exposure of signalling motifs associated with nuclear translocation. Besides providing insights into the mechanism of unfolding, our residue-specific analysis of temperature response reveals several key amino acid positions and protein segments, which can serve as targets for the modulation of protein stability. This can have implications for the future development of pharmaceutical agents affecting bile salt metabolism and the engineering of BABP-based drug carrier systems.

Materials and methods

Protein biosynthesis and purification

For unenriched protein, the *Escherichia coli* strain MG1655, transformed with the pMON5840-hI-BABP construct, was incubated overnight in 50 mL LB media containing 100 $\mu g \cdot mL^{-1}$ ampicillin, then divided and transferred into two 2-L flasks, containing 700 mL LB media each. Protein expression, under control of the recA promoter, was induced by the addition of 7 mL 10 $mg \cdot mL^{-1}$ nalidixic acid to the growing culture at $OD_{600} \sim 1.5$ and cells were allowed to grow 3–4 h more, harvested and frozen at −70 °C. For uniformly ^{15}N-enriched protein, the transformed *E. coli* strain was incubated overnight (37 °C, 250 r.p.m.) in 50 mL nonisotope-enriched minimal media (6 $g \cdot L^{-1}$ Na_2HPO_4, 3 $g \cdot L^{-1}$ KH_2PO_4, 0.5 $g \cdot L^{-1}$ NaCl, 1 $g \cdot L^{-1}$ NH_4Cl, 1 mM $MgSO_4$, 5 $g \cdot L^{-1}$ D-glucose, 25 $mg \cdot L^{-1}$ thiamine.HCl, 0.1 mM $CaCl_2$ and trace metals according to Li *et al.* [65] containing 100 $\mu g \cdot mL^{-1}$ ampicillin, then transferred into a 2-L flask, containing 700 mL fresh isotope-enriched minimal media with 0.7 g $^{15}NH_4Cl$. Protein expression, under the control of the recA promoter, was induced by the addition of 7 mL 10 $mg \cdot mL^{-1}$ nalidixic acid to the growing culture at $OD_{600} \sim 2.0$ and cells were allowed to grow 3–4 h more, harvested and frozen at −70 °C. For unenriched and ^{15}N-enriched proteins, the cell pellet was thawed in ~ 10–20 mL of 20 mM Tris-Cl, pH 8.0, 5 mM EDTA, containing a broad-spectrum protease inhibitor mixture (Roche Molecular Biochemicals, Basel, Switzerland). The protein was released by processing the cell suspension through an Emulsiflex C5 homogenizator (Avestin, Mannheim, Germany) three to four subsequent times. The homogenized cell suspension was subjected to centrifugation at 4 °C, 23 000 ***g*** for 30 min. The clarified supernatant was chromatographed on a 25 × 5 cm column of Q-Sepharose Fast Flow. Dialysis was carried out against multiple changes of 10 mM Tris, pH = 8.2, 0.05% NaN_3 at 4 °C, until the A_{260}/A_{280} absorbance ratio decreased below 0.7 followed by gel filtration chromatography using a 110 × 2.5 cm column of Sephadex G-50. Delipidation was achieved by passing the protein solution over a column of lipophilic Sephadex type VI (Sigma product no. H-6258, Sigma-Aldrich, St. Louis, MO, USA) pre-equilibrated with 20 mM potassium phosphate, 50 mM KCl and 0.05% NaN_3 (pH 6.3) at 37 °C [66]. Protein purity, as assessed by overloaded Coomassie-stained SDS/PAGE gels, was > 98%. Protein concentration was determined by absorbance at 280 nm using an extinction coefficient of 12 930 $M^{-1}\cdot cm^{-1}$ obtained by composition analysis according to Pace *et al.* [67].

Fluorescence and CD spectroscopy

To monitor temperature-induced spectral changes, fluorescence emission spectra were recorded in the range of 290–510 nm on a Jobin-Yvon Fluoromax-4 (Horiba Scientific, Kyoto, Japan) photon counting spectrofluorometer between 5 and 85 °C in 10 °C steps. Protein solutions were excited at 297 nm and the fluorescence spectra were recorded.

Far-UV CD measurements were performed in sealable quartz cells of 0.1 cm pathlength using a Jasco J-810 (Jasco, Easton, MD, USA) spectrophotometer. Spectra were recorded in the range of 200–280 nm between 5 and 100 °C in 5 °C steps. Five scans were averaged at each temperature point. A continuous heat-up of the sample was measured at 215 nm from 5 to 98 °C with a scanning rate of 60 $°C\cdot h^{-1}$. Temperature was controlled using a PFD-425S Peltier-type heating system. CD spectra were analysed using the BeStSel algorithm [30]. CD band intensities are expressed in mean residue ellipticity ($[\Theta]_{MR}$, deg $cm^2\cdot dmol^{-1}$). Fluorescence and far-UV CD spectroscopy measurements were carried out in 20 mM potassium phosphate, 50 mM KCl and 0.05% NaN_3 at pH = 6.3 with a protein concentration of 2 and 15 μM respectively.

NMR data collection and analysis

Two-dimensional ^{1}H-^{15}N HSQC experiments [68] were carried out on a 600 Varian NMR SYSTEM™ (Varian, Palo Alto, CA, USA) spectrometer equipped with a 5-mm indirect detection triple ($^{1}H^{13}C^{15}N$) *z*-axis gradient probe at the following temperatures: 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 37, 42, 45, 48, 51, 54 °C. Pulse widths were adjusted at each temperature and were ranging from 7.5 μs (5 °C) to 8.7 μs (54 °C) for ^{1}H and 32–33 μs for ^{15}N. Spectra were acquired with spectral widths of 8012 and 2000 Hz in the ^{1}H and ^{15}N dimensions respectively. The number of complex points in the ^{1}H dimension was 2048, subsequently zero-filled to a total of 4096 points. The number of increments in the indirectly detected dimension was 256, subsequently linear predicted to 512. Gaussian and exponential weighting functions were applied in the F2 (^{1}H) dimension, whereas the F1 (^{15}N) dimension was Gaussian weighted only. Measurements were carried out in a buffer containing 20 mM potassium phosphate, 50 mM KCl and 0.05% NaN_3 at pH = 6.3 with a protein concentration of 100 μM. The ^{1}H chemical shifts were referenced externally to 2,2-dimethylsilapentane-5-sulphonic acid (DSS), whereas the ^{15}N chemical shifts were referenced indirectly to DSS [69]. Backbone amide (δ_{1H}, δ_{15N}) chemical shifts as a function of temperature were fit to

$$\delta_{1H/15N}(T) = x^{F}(T)\delta^{F}_{1H/15N} + x^{U^*}(T)\delta^{U^*}_{1H/15N} \qquad (1)$$

for two-state transitions, and

$$\delta_{1H/15N}(T) = x^{F}(T)\delta^{F}_{1H/15N} + x^{I}(T)\delta^{I}_{1H/15N} + x^{U^*}(T)\delta^{U^*}_{1H/15N} \qquad (2a)$$

or

$$\begin{aligned}\delta_{1H/15N}(T) = {} & x^{F}(T)\delta^{F}_{1H/15N} + c_1 x^{U^*}(T)\delta^{U^*}_{1H/15N} \\ & + 2c_2 x^{\dim}(T)\delta^{U^*2}_{1H/15N}\end{aligned} \qquad (2b)$$

for three-state transitions, where x^k is the molar fraction of state k at a given temperature and δ^k is its $^{1}H/^{15}N$ chemical shift ($k = F, I, U^*, U^*_2$) [34]. Values of x^k as a function of temperature were determined using the Gibbs–Helmholtz equation. Specifically, the free energy change in an unfolding transition at a specific temperature can be given in terms of the changes in enthalpy (ΔH), heat capacity (ΔC_p) and the melting temperature (T_m) as follows.

(i) For a two-state $F \leftrightarrow U^*$ process:

$$\begin{aligned}\Delta G^{F-U^*}(T) = {} & \Delta H^{F-U^*}(T_m^{F-U^*})\left[1 - \frac{T}{T_m^{F-U^*}}\right] \\ & + \Delta C_p^{F-U^*}\left[T - T_m^{F-U^*} - T\ln\left(\frac{T}{T_m^{F-U^*}}\right)\right]\end{aligned} \qquad (3)$$

where

$$\Delta G^{F-U^*}(T) = -RT\ln K^{F-U^*} \qquad (4)$$

The molar fraction of the components at a specific temperature can be calculated as

$$x^{F}(T) = \frac{1}{1 + K^{F-U^*}} \tag{5a}$$

and

$$x^{U^*}(T) = \frac{K^{F-U^*}}{1 + K^{F-U^*}} \tag{5b}$$

(ii) For a three-state $F \leftrightarrow I \leftrightarrow U^*$ process:

$$\Delta G^{F-I}(T) = \Delta H^{F-I}(T_{\mathrm{m}}^{F-I})\left[1 - \frac{T}{T_{\mathrm{m}}^{F-I}}\right] + \Delta C_{\mathrm{p}}^{F-I}\left[T - T_{\mathrm{m}}^{F-I} - T\ln\left(\frac{T}{T_{\mathrm{m}}^{F-I}}\right)\right] \tag{6a}$$

$$\Delta G^{I-U^*}(T) = \Delta H^{I-U^*}(T_{\mathrm{m}}^{I-U^*})\left[1 - \frac{T}{T_{\mathrm{m}}^{I-U^*}}\right] + \Delta C_{\mathrm{p}}^{I-U^*}\left[T - T_{\mathrm{m}}^{I-U^*} - T\ln\left(\frac{T}{T_{\mathrm{m}}^{I-U^*}}\right)\right] \tag{6b}$$

$$\Delta G^{F-I}(T) = -RT\ln K^{F-I} \tag{7a}$$

$$\Delta G^{I-U^*}(T) = -RT\ln K^{I-U^*} \tag{7b}$$

The molar fraction of the components at a specific temperature can be calculated as

$$x^{F}(T) = \frac{\frac{1}{K^{F-I}}}{K^{I-U^*} + 1 + \frac{1}{K^{F-I}}} \tag{8a}$$

$$x^{I}(T) = \frac{1}{K^{I-U^*} + 1 + \frac{1}{K^{F-I}}} \tag{8b}$$

and

$$x^{U^*}(T) = \frac{\frac{1}{K^{I-U^*}}}{K^{I-U^*} + 1 + \frac{1}{K^{F-I}}} \tag{8c}$$

(iii) For a three-state process, when $F \leftrightarrow U^*$ is followed by a dimerization of U^*:

$$\Delta G^{F-U^*}(T) = \Delta H^{F-U^*}(T_{\mathrm{m}}^{F-U^*})\left[1 - \frac{T}{T_{\mathrm{m}}^{F-U^*}}\right] + \Delta C_{\mathrm{p}}^{F-U^*}\left[T - T_{\mathrm{m}}^{F-U^*} - T\ln\left(\frac{T}{T_{\mathrm{m}}^{F-U^*}}\right)\right] \tag{9a}$$

$$\Delta G^{\mathrm{dim}}(T) = \Delta H^{\mathrm{dim}}(T_{\mathrm{dim}})\left[1 - \frac{T}{T_{\mathrm{dim}}}\right] + \Delta C_{\mathrm{p}}^{\mathrm{dim}}\left[T - T_{\mathrm{dim}} - T\ln\left(\frac{T}{T_{\mathrm{dim}}}\right)\right] \tag{9b}$$

where

$$\Delta G^{F-U^*}(T) = -RT\ln K^{F-U^*} \tag{10a}$$

$$\Delta G^{\mathrm{dim}}(T) = -RT\ln K^{\mathrm{dim}} \tag{10b}$$

The molar fraction of the components at a specific temperature can be calculated as

$$x^{F}(T) = \frac{-(1 + c_1K^{F-U^*}) + \sqrt{(1 + c_1K^{F-U^*})^2 + 8c_1^2c_2K^{\mathrm{dim}}K^{F-U^*}K^{F-U^*}}}{4c_1^2c_2K^{\mathrm{dim}}K^{F-U^*}K^{F-U^*}} \tag{11a}$$

$$x^{U^*}(T) = \frac{-(1 + c_1K^{F-U^*}) + \sqrt{(1 + c_1K^{F-U^*})^2 + 8c_1^2c_2K^{\mathrm{dim}}K^{F-U^*}K^{F-U^*}}}{4c_1^2c_2K^{\mathrm{dim}}K^{F-U^*}} \tag{11b}$$

and

$$x^{\mathrm{dim}}(T) = \left(\frac{-(1 + c_1K^{F-U^*}) + \sqrt{(1 + c_1K^{F-U^*})^2 + 8c_1^2c_2K^{\mathrm{dim}}K^{F-U^*}K^{F-U^*}}}{4c_1^2c_2K^{\mathrm{dim}}K^{F-U^*}}\right)^2 K_{\mathrm{dim}} \tag{11c}$$

Scaling factors c_1 and c_2 were introduced into Eqns (2b) and (11a–c), to correct for partially unfolded monomers and dimers that turned into larger aggregates and due to their high molecular weight (slow rotational correlation) and heterogeneity are not expected to contribute. c_1 and c_2 were optimized by performing a grid search for a subset of 15 residues located in various secondary structure elements throughout the protein yielding an average value of $c_1 = 0.96 \pm 0.06$ and $c_2 = 0.54 \pm 0.08$ indicating that on average, monomeric species are not much affected, whereas about half of the formed dimers do not report. As c_1 and c_2 should be temperature dependent, which to keep the number of fitted parameters minimal we did not take into account, this is a crude but reasonable estimate. We note that while the scaling factors naturally influenced the exact values of fitted parameters, they had only a marginal effect on the $\Delta\delta_{F-U^*}$ chemical shift difference.

For a joint analysis of NMR thermal melting and conformational exchange, relaxation compensated CPMG [70] relaxation dispersion data were collected and analysed as described previously [14].

For the cluster analysis of individual amide ^{1}H unfolding profiles, chemical shift changes as a function of temperature were normalized for each residue using

$$\delta_{\mathrm{norm},T}^{\mathrm{HN}} = \frac{\delta_{\mathrm{exp},T} - \delta_{\mathrm{exp},5^\circ\mathrm{C}}}{\delta_{\mathrm{exp},5^\circ\mathrm{C}} - \delta_{\mathrm{rc},5^\circ\mathrm{C}}}$$

where $\delta_{\mathrm{exp},T}$ and $\delta_{\mathrm{exp},5\ ^\circ\mathrm{C}}$ are chemical shifts detected at temperature T and 5 °C, respectively, and $\delta_{\mathrm{rc},5\ ^\circ\mathrm{C}}$ is a sequence-corrected calculated random coil chemical shift at 5 °C (pH = 6.3) [71,72]. The analysis was performed by clustering the temperature series of normalized chemical shifts according to the root-mean-square deviations using an in-house program.

Analysis of conserved pairwise interactions

Amino acid sequence of 94 members of the iLBP family with different organism-, tissue- and ligand specificity was analysed. Specifically, about equal number of FABP (intestinal/brain/adipocyte), CR(A)BP I/II and BABP proteins were selected from rat, mouse, bovine, chicken, frog, fish and human from the SWISSPROT database sharing a sequence identity > 20% with human I-BABP. Multiple sequence alignment of the sequences was carried out using the CLUSTAL OMEGA program [73]. Amino acids responsible for structural conservation within the family were analysed based on identity and similarity, for the latter relying on the Gonnet Pam250 matrix [74]. Accordingly, using the groupings of 'STA', 'NEQK'. 'NHQK', 'NDEQ', 'QHRK', 'MILV', 'MILF', 'HY', 'FYW', 'C', if an amino acid is replaced by another one belonging to the same group, then that substitution was considered to be a conserved mutation. The identification of conserved pairwise interactions was based on the solution NMR structure of human I-BABP (PDB: 1O1U) [8]. Specifically, residues i and j were considered to be interacting if any side-chain atom of residue i is within 4.0 Å of any side-chain atom of residue j in more than 50% of the lowest energy structural ensemble. Interactions with $|i - j| > 5$ and $|i - j| < 6$ were classified as long- and short-range interactions respectively.

Molecular dynamics simulations

The solution NMR structure of human I-BABP [8] was subjected to MD simulations as implemented in GROMACS [75], using the AMBER-ff99SB*-ILDNP force field [76]. The system was solvated by ~ 5016 water molecules with TIP4P parametrization [77]. The total charge of the system was neutralized, and the physiological salt concentration was set by placing Na^+ and Cl^- ions. Energy minimization of starting structures was followed by sequential relaxation of constraints on protein atoms in three steps and an additional NVT step (all of 200 ps) to stabilize pressure. One microsecond trajectories of NPT simulations at 300, 350 and 370 K and a 400-ns trajectory at 400 K at 1 bar were recorded (collecting snapshots at every 20 ps). Secondary structure compositions of the frames of MD trajectories were determined by DSSP algorithm [47]. Molecular graphics was performed with the UCSF CHIMERA package (University of California, San Francisco) [78]. The VMD [79] and BIO3D R-package [48] were used to analyse the correlated motions and for the principal component analysis (PCA). All Cα atoms were used for trajectory frame superposition and for PCA resulting principal components (orthogonal eigenvectors) describing the axes of maximal variance of the distribution of movement of the structures. To analyse which atomic displacements in the trajectory are correlated with one another, we used Bio3D cross-correlation analysis module [48]. The first 100 ns of the trajectories were excluded from the analysis. We used an in-house program to find out which regions of the molecule move together. The calculation was based on the clustering of the time series of alterations of pairwise distances between Cα atom positions. To focus on the slow movements of the system, fast motions were averaged by 5 and 10 ns (i.e. 250 and 500 frames) averaging to determine the Cα atom positions before the calculation. Hydrophilic and hydrophobic surfaces were calculated using Getarea with a water probe of size 1.4 Å [80].

Acknowledgements

This work was supported by National Research, Development and Innovation Office (NKFIH) Grants K-109035 (to OT), K-123995 (to LB), K-117062 (to MK), K-120391 (to JK and AM), KTIA_NAP_13-2-2014-0017 (to JK and AM), ERC_HU 117680 (to MK), the 'Momentum' Program of the Hungarian Academy of Sciences (LP2011-006/2011 to MK), the 'János Bolyai Research Scholarship' Program of the Hungarian Academy of Sciences (to OT and AM), and a MedinProt Grant (to LB, MK and OT). The authors thank Prof David Cistola for the pMON5840 plasmid and Dr Nikolett T. Nagy for assistance with the stopped-flow experiments.

Author contributions

GH produced the isotopically labelled proteins and analysed the relaxation NMR experiments. LB and MK performed the fluorescence experiments. ZM carried out the CD experiments. AM and JK performed and analysed the MD simulations and wrote part of the manuscript. OT carried out and analysed the NMR experiments, designed the research and wrote the manuscript.

References

1 Banaszak L, Winter N, Xu Z, Bernlohr DA, Cowan S & Jones TA (1994) Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv Protein Chem* **45**, 89–151.

2 Veerkamp JH & Maatman RG (1995) Cytoplasmic fatty acid-binding proteins: their structure and genes. *Prog Lipid Res* **34**, 17–52.

3 Small DM, Dowling RH & Redinger RN (1972) The enterohepatic circulation of bile salts. *Arch Intern Med* **130**, 552–573.

4 Nakahara M, Furuya N, Takagaki K, Sugaya T, Hirota K, Fukamizu A, Kanda T, Fujii H & Sato R

(2005) Ileal bile acid-binding protein, functionally associated with the farnesoid X receptor or the ileal bile acid transporter, regulates bile acid activity in the small intestine. *J Biol Chem* **280**, 42283–42289.

5 Prawitt J, Caron S & Staels B (2011) Bile acid metabolism and the pathogenesis of type 2 diabetes. *Curr Diab Rep* **11**, 160–166.

6 Payne CM, Bernstein C, Dvorak K & Bernstein H (2008) Hydrophobic bile acids, genomic instability. Darwinian selection, and colon carcinogenesis. *Clin Exp Gastroenterol* **1**, 19–47.

7 Fang C, Dean J & Smith JW (2007) A novel variant of ileal bile acid binding protein is up-regulated through nuclear factor-kB activation in colorectal adenocarcinoma. *Cancer Res* **67**, 9039–9046.

8 Kurz M, Brachvogel V, Matter H, Stengelin S, Thüring H & Kramer W (2003) Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholyltaurine complex and its comparison with homologous structures. *Proteins* **50**, 312–328.

9 Horváth G, Bencsura Á, Simon Á, Tochtrop GP, DeKoster GT, Covey DF, Cistola DP & Toke O (2016) Structural determinants of ligand binding in the ternary complex of human ileal bile acid binding protein with glycocholate and glycochenodeoxycholate obtained from solution NMR. *FEBS J* **283**, 541–555.

10 Tochtrop GP, Richter K, Tang C, Toner JJ, Covey DF & Cistola DP (2002) Energetics by NMR: site-specific binding in a positively cooperative system. *Proc Natl Acad Sci USA* **99**, 1847–1852.

11 Tochtrop GP, DeKoster GT, Covey DF & Cistola DP (2004) A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein. *J Am Chem Soc* **126**, 1102–11029.

12 Horváth G, Király P, Tárkányi G & Toke O (2012) Internal motions and exchange processes in human ileal bile acid binding protein as studied by backbone ^{15}N nuclear magnetic resonance spectroscopy. *Biochemistry* **51**, 1848–1861 (erratum appears in Biochemistry 51, 10119).

13 Ragona L, Catalano M, Luppi M, Cicero D, Eliseo T, Foote J, Fogolari F, Zetta L & Molinari H (2006) NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. *J Biol Chem* **281**, 9697–9709.

14 Horváth G, Egyed O & Toke O (2014) Temperature dependence of backbone dynamics in human ileal bile acid-binding protein: implications for the mechanism of ligand binding. *Biochemistry* **53**, 5186–5198.

15 Ragona L, Pagano K, Tomaselli S, Favretto F, Ceccon A, Zanzoni S, D'Onofrio M, Assfalg M & Molinari H (2014) The role of dynamics in modulating ligand exchange in intracellular lipid binding proteins. *Biochim Biophys Acta* **1844**, 1268–1278.

16 Villareal MA, Perduca M, Monaco HL & Montich GG (2008) Binding and interactions if L-BABP to lipid membranes studied by molecular dynamics simulations. *Biochim Biophys Acta* **1778**, 1390–1397.

17 Pedò M, Löhr F, D'Onofrio M, Assfalg M, Dötsch V & Molinari H (2009) NMR studies reveal the role of biomembranes in modulating ligand binding and release by intracellular bile acid binding proteins. *J Mol Biol* **394**, 852–863.

18 Sessler RJ & Noy N (2005) A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Mol Cell* **18**, 343–353.

19 Kim K, Ramanathan R & Frieden C (1997) Intestinal fatty acid binding protein: a specific residue in one turn appears to stabilize the native structure and be responsible for slow refolding. *Protein Sci* **6**, 364–372.

20 Dalessio PM & Ropson IJ (2000) Beta-sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry* **39**, 860–871.

21 Burns LL & Ropson IJ (2001) Folding of intracellular retinol and retinoic acid binding proteins. *Proteins* **43**, 292–302.

22 Yeh SR, Ropson IJ & Rousseau DL (2001) Hierarchical folding of intestinal fatty acid binding protein. *Biochemistry* **40**, 4205–4210.

23 Chattopadhyay K, Zhong S, Yeh SR, Rousseau DL & Frieden C (2002) The intestinal fatty acid binding protein: the role of turns in fast and slow folding processes. *Biochemistry* **41**, 4040–4047.

24 Dalessio PM, Boyer JA, McGettigan JL & Ropson IJ (2005) Swapping core residues in homologous proteins swaps folding mechanism. *Biochemistry* **44**, 3082–3090.

25 Li H & Frieden C (2007) Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and ^{19}F NMR. *Proc Natl Acad Sci USA* **104**, 11993–11998.

26 D'Onofrio M, Ragona L, Fessas D, Signorelli M, Ugolini R, Pedò M, Assfalg M & Molinari H (2009) NMR unfolding studies on a liver bile acid binding protein reveal a global two-state unfolding and localized singular behaviors. *Arch Biochem Biophys* **481**, 21–29.

27 Ropson IJ, Boyer JA, Schaeffer BA & Dalessio PM (2009) Comparison of the folding mechanism of highly homologous proteins in the lipid-binding protein family. *Proteins* **75**, 799–806.

28 Ogbay B, DeKoster GT & Cistola DP (2004) The NMR structure of a stable and compact all β-sheet variant of intestinal fatty acid-binding protein. *Protein Sci* **13**, 1227–1237.

29 Haring D & Distefano MD (2001) Enzymes by design: chemogenetic assembly of transamination active sites containing lysine residues for covalent catalysis. *Bioconjug Chem* **12**, 385–390.

30 Micsonai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M & Kardos J (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci USA* **112**, E3095–E3103.

31 Basehore HK & Ropson IJ (2011) Residual interaction in unfolded bile acid-binding protein by ^{19}F NMR. *Protein Sci* **20**, 327–335.

32 Cordier F & Grzesiek S (2002) Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *J Mol Biol* **317**, 739–752.

33 Wintrode PL, Makhatadze GI & Privalov PL (1994) Thermodynamics of ubiquitin unfolding. *Proteins* **18**, 246–253.

34 Heyda J, Kozisek M, Bednarova L, Thompson G, Konvalinka J, Vondrasek J & Jungwirth P (2011) Urea and guanidinium induced denaturation of a Trp-cage miniprotein. *J Phys Chem B* **115**, 8910–8924.

35 Maatman RGHJ, van Kuppevelt THMSM & Veerkamp JH (1991) Two types of fatty acid-binding protein in human kidney. *Biochem J* **273**, 759–766.

36 Gillilian RE, Ayers SD & Noy N (2007) Structural basis for activation of fatty acid binding protein FABP4. *J Mol Biol* **372**, 1246–1260.

37 Sanson B, Wang T, Sun J, Wang L, Kaczocha M, Ojima I, Deutsch D & Li H (2014) Crystallographic study of FABP5 as an intracellular endocannabinoid transpoter. *Acta Crystallogr D* **70**, 290–298 (dimer).

38 Kim H, Gil G, Lee S, Kwak A, Jo S, Kim E, Nguyen TT, Kim S, Jhun H, Kim S *et al.* (2016) Cytokine-like activity of liver type fatty acid binding protein (L-FABP) inducing inflammatory cytokine interleukin-6. *Immune Netw* **16**, 296–304.

39 Palmer AG III, Kroenke CD & Loria JP (2001) Nuclear magnetic resonane methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* **339**, 204–238.

40 Vallurupalli P, Hansen DF & Kay LE (2008) Structures of invisible, excited states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci USA* **105**, 11766–11771.

41 Toke O, Monsey JD & Cistola DP (2007) Kinetic mechanism of ligand binding in human ileal bile acid binding protein as determined by stopped-flow fluorescence analysis. *Biochemistry* **46**, 5427–5436.

42 Evenäs J, Tugarinov V, Skrynnikov NR, Goto NK, Muhandiram R & Kay LE (2001) Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. *J Mol Biol* **309**, 961–974.

43 Gunasekaran K, Hagler AT & Gierasch LM (2004) Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins* **54**, 179–194.

44 Hodsdon ME & Cistola DP (1997) Ligand binding alters the backbone mobility of intestinal fatty acid-binding protein as monitored by ^{15}N NMR relaxation and ^{1}H exchange. *Biochemistry* **36**, 2278–2290.

45 Wagner G, Pardi A & Wüthrich K (1983) Hydrogen-bond length and H-1-NMR chemical-shifts in proteins. *J Am Chem Soc* **105**, 5948–5949.

46 Sadqi M, Fushman D & Munoz V (2006) Atom-by-atom analysis of global downhill protein folding. *Nature* **442**, 317–321.

47 Kabsch W & Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogenbonded and geometrical features. *Biopolymers* **22**, 2577–2637.

48 Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA & Caves LS (2006) Bio3D: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696.

49 Stefani M (2004) Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim Biophys Acta* **1739**, 5–25.

50 Chiti F & Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333–366.

51 Whitney DS & Volkman BF (2015) Some (dis)assembly required: partial unfolding in the Par-6 allosteric switch. *Biophys Rev* **7**, 183–190.

52 Mitrea DM & Kriwacki RW (2013) Regulated unfolding of protein sin signaling. *FEBS Lett* **587**, 1081–1088.

53 Foit L, George JS, Zhang BW, Brooks CL III & Bardwell JCA (2013) Chaperone activation by unfolding. *Proc Natl Acad Sci USA* **110**, E1254–E1262.

54 Arrigoni C, Rohaim A, Shaya D, Findeisen F, Stein RA, Nurva SR, Mishra S, Mchaourab HS & Minor DL Jr (2016) Unfolding of a temperature-sensitive domain controls voltage-gated channel activation. *Cell* **164**, 922–936.

55 Chatterjee A, Krishna Mohan PM, Prabhu A, Ghosh-Roy A & Hosur RV (2007) Equilibrium unfolding of DLC8 monomer by urea and guanidine hydrochloride: distinctive global and residue level features. *Biochimie* **89**, 117–134.

56 Dalessio PM, Fromholt SE & Ropson IJ (2005) The role of Trp-82 in the folding of intestinal fatty acid binding protein. *Proteins* **61**, 176–183.

57 Kim K, Cistola DP & Frieden CF (1996) Intestinal fatty acid-binding protethe structure and stability of a helix-less variant. *Biochemistry* **35**, 7553–7558.

58 Ogbay B, DeKoster GT & Cistola DP (2003) The NMR structure of a stable and compact all-beta-sheet variant of intestinal fatty acid-binding protein. *Protein Sci* **13**, 1227–1237.

59 Kouvatsos N, Meldrum JK, Searle MS & Thomas NR (2006) Coupling of ligand recognition to protein folding

in an engineered variant of rabbit ileal lipid binding protein. *Chem Commun (Camb)* **44**, 4623–4625.

60 Ropson IJ & Frieden C (1992) Dynamic NMR spectral analysis and protein folding: identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by ^{19}F NMR. *Proc Natl Acad Sci USA* **89**, 7222–7226.

61 Hodsdon ME & Cistola DP (1997) Discrete backbone disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid-binding protein: implications for the mechanism of ligand entry. *Biochemistry* **36**, 1450–1460.

62 Lu J, Cistola DP & Li E (2003) Two homologous rat cellular retinol-binding proteins differ in local conformational flexibility. *J Mol Biol* **330**, 799–812.

63 Cogliati C, Ragona L, D'Onofrio M, Günther U, Whittaker S, Ludwig C, Tomaselli S, Assfalg M & Molinari H (2010) Site-specific investigation of the steady-state kinetics and dynamics of the multistep binding of bile acid molecules to a lipid carrier protein. *Chem Eur J* **16**, 11300–11310.

64 Ayers SD, Nedrow KL, Gillilan RE & Noy N (2007) Continuous nucleocytoplasmic shuttling underlies transcriptional activation of PPARγ by FABP4. *Biochemistry* **46**, 6744–6752.

65 Li E, Locke B, Yang NC, Ong DE & Gordon JI (1987) Characterization of rat cellular retinol-binding protein II expressed in *Escherichia coli*. *J Biol Chem* **262**, 13773–13779.

66 Glatz JFC & Veerkamp JH (1983) A radiochemical procedure for the assay of fatty acid binding by proteins. *Anal Biochem* **132**, 89–95.

67 Pace CN, Vajdos F, Fee L, Grimsley G & Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**, 2411–2423.

68 Kay LE, Keifer P & Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* **114**, 10663–10665.

69 Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE & Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Pure Appl Chem* **70**, 117−142.

70 Loria JP, Rance M & Palmer AG III (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc* **121**, 2331–2332.

71 Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE & Dyson HJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **123**, 2970–2978.

72 Kjaergaard M, Brander S & Poulsen FM (2011) Random coil chemical shifts for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* **49**, 139–149.

73 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

74 Dayhoff MO, Schwartz RM & Orcutt BC (1978) A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Dayhoff MO, ed.), 5 Suppl 3, pp. 345–352. National Biomedical Science Foundation, Washington, DC.

75 Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854.

76 Aliev AE, Kulke M, Khaneja HS, Chudasama V, Sheppard TD & Lanigan RM (2014) Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. *Proteins* **82**, 195–215.

77 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935.

78 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.

79 Humphrey W, Dalke A & Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38.

80 Fraczkiewicz R & Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* **19**, 319–333.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Thermodynamic and structural parameters obtained by fitting the temperature-dependent ^{1}H-^{15}N HSQC spectra of [99% ^{15}N-labelled] human I-BABP (20 mM K-phosphate, 50 mM KCl, 0.05% NaN_3, 10% D_2O, pH = 6.3) using a three-state model assuming a main $F \leftrightarrow U^*$ transition and a subsequent dimerization of the U^* state.

Fig. S1. Alignment of 94 iLBP sequences with different organism-, tissue- and ligand specificity.

Movie S1. The first principal component of movements analysing the 1 μs MD simulation at 300 K.

Movie S2. The second principal component of movements analysing the 1 μs MD simulation at 300 K.

Movie S3. The first principal component of movements analysing the 1 μs MD simulation at 350 K.

Movie S4. The second principal component of movements analysing the 1 μs MD simulation at 350 K.

Movie S5. The first principal component of movements analysing the 1 μs MD simulation at 370 K.

Movie S6. The second principal component of movements analysing the 1 μs MD simulation at 370 K.