Published in final edited form as:

J Mol Biol. 2013 September 23; 425(18): 3522–3535. doi:10.1016/j.jmb.2013.06.030.

The role of aromatic-aromatic interactions in strand-strand stabilization of β -sheets

Ivan L. Budyak¹, Anastasia Zhuravleva¹, and Lila M. Gierasch^{1,2,a}

¹Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, Massachusetts 01003 USA

²Department of Chemistry, University of Massachusetts Amherst, Amherst, Massachusetts 01003 USA

Abstract

Aromatic-aromatic interactions have long been believed to play key roles in protein structure, folding, and binding functions. Yet we still lack full understanding of the contributions of aromatic-aromatic interactions to protein stability and the timing of their formation during folding. Here, using as a case study an aromatic ladder in the β -barrel protein, cellular retinoic acid binding protein 1 (CRABP1), we find aromatic stacking plays a greater role in the Phe65-Phe71 cross-strand pair while in another pair, Phe50-Phe65, hydrophobic interactions are dominant. The Phe65/Phe71 pair spans β -strands 4 and 5 in the β -barrel, which lack interstrand hydrogen bonding, and we speculate that it compensates energetically for the absence of strand-strand backbone interactions. Using perturbation analysis, we find that both aromatic-aromatic pairs form after the transition state for folding of CRABP1, thus playing a role in the final stabilization of the β -sheet rather than in its nucleation as had been earlier proposed. The aromatic interaction between strands 4–5 in CRABP1 is highly conserved in the intracellular lipid-binding protein (iLBP) family, and several lines of evidence combine to support a model wherein it acts to maintain barrel structure while allowing the dynamic opening that is necessary for ligand entry. Lastly, we carried out a bioinformatic analysis and found 51 examples of aromatic-aromatic interactions across non-hydrogen-bonded β -strands outside the iLBPs, arguing for the generality of the role played by this structural motif.

Keywords

aromatic-aromatic interactions; β -barrel; folding; dynamics; hydrogen bond; stacking

INTRODUCTION

Considerable effort has been expended over several years to understand how aromatic-aromatic interactions contribute to protein folding, protein stability, and protein-protein or

© 2013 Elsevier Ltd. All rights reserved.

^aCorresponding author: Lila M. Gierasch, Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, 710 N. Pleasant St., Amherst, MA 01003 USA, Tel. (413) 545-6094, Fax (413) 545-1289, gierasch@biochem.umass.edu.

Backbone assignments for WT CRABP1 have been deposited in the BioMagResBank (BMRB; <http://www.bmrb.wisc.edu>) under accession number 19271.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

protein-ligand recognition.^{1,2} We are particularly interested in the roles aromatic-aromatic interactions may play in β -rich proteins, because the intrinsic secondary structural propensities of aromatic amino acids places them with higher frequency in β -structure than in helix.³ Moreover, Thomas *et al.* noted that 74% of aromatic pairs are between residues far apart in sequence (greater than five residues), and that approximately one-third of far-sequence aromatic pairs occur in β -structure.⁴ Secondary and tertiary structure formation is coupled in β -proteins, and β -rich proteins are stabilized to a greater extent than helical proteins by long-range interactions (i.e., between residues separated in sequence by five or more amino acids): Thus, one may ask whether there is a specific role for aromatic-aromatic interactions in the formation and stability of β -structure. If so, what kinds of interactions do cross-strand aromatic pairs contribute most to the stability of a β -sheet – hydrophobic or aromatic? Do cross-strand pairs preferentially stabilize neighboring strands when usual stabilizing interactions such as interstrand hydrogen bonds (H-bonds) are absent? One would envision this to be the case since aromatic amino acids may both increase the overall sheet propensity and contribute highly specific interactions at distances larger than H-bonds typically do.

The topological complexity of β -proteins leads to rugged folding energy landscapes with long-lived intermediate states,^{5,6} and relatively slow folding.^{7,8,9} When do specific interactions between aromatic rings form in the course of β -sheet folding – do they actively drive folding or fine tune and lock in sheet structure late in folding? Not surprisingly, β -proteins generally display higher vulnerability to aggregation than helical proteins. Are aromatic-aromatic interactions protective against exposure of aggregation-prone regions?

Previous studies have explored whether aromatic interactions play a particular role in the stabilization and folding of β -structure without a clear consensus. In many peptide models aromatic cross-strand pairs have been reported to stabilize β -hairpins,^{10,11,12,13} but some studies argue for the absence of favorable cross-strand aromatic interactions.¹⁴ Currently suggested contributions of aromatic interactions to protein folding range from participation in early folding events such as nucleation¹⁵ to late stabilization of tertiary structure.⁴ Perturbation analysis on a solvent-exposed aromatic cluster of bacterial cold shock protein A yielded non-canonical negative ΔG -values¹⁶ possibly due to perturbation of the denatured and/or transition states (TS). A cross-strand Phe stack has been found to be important for the folding of the parallel β -helix domain of the phage P22 tailspike,¹⁷ but its energetic contribution to the stability of the protein has not been measured. Investigation of the role of aromatic pairs in the Greek key motifs of human D-crystallin showed that the four Greek key aromatic pairs may play a variety of roles early and late in the complex folding of this two-domain protein.¹⁸

We seek to better understand the sequence information that encodes β -protein folding mechanisms, susceptibility to aggregation, and specificity of sequence information, and to what extent, if any, aromatic-aromatic interactions play specialized roles in β -proteins. The present study addresses these issues through a multipronged analysis of the roles played by an aromatic array made up of three phenylalanines in a model β -barrel protein, mouse cellular retinoic acid-binding protein 1 (CRABP1), a member of the large family of intracellular lipid-binding proteins (iLBPs).¹⁹ CRABP1 serves as an apt model system to explore the roles of aromatic-aromatic interactions in β -structure because of its well-characterized folding pathway with distinct kinetic phases corresponding to sequential steps of β -sheet formation.^{20,21} Furthermore, the aromatic cross-strand ladder is a conserved feature in the iLBP family,²² and hence any findings on CRABP1 are more generally applicable. We used protein engineering methods to dissect the nature of the energetic contributions of the aromatic interactions within the Phe array in CRABP1 and to determine the timeline of the formation of these cross-strand aromatic interactions in CRABP1 folding.

Intriguingly, we found that one of the two aromatic pairs in CRABP1 establishes primarily a hydrophobic packing interaction, while the other is stabilized as well by true aromatic-aromatic interaction. Based on NMR analysis, comparisons with crystal structures of homologues, and molecular dynamics (MD) simulations, we observe that the former has less stereochemical restriction than the latter. We conclude from several lines of evidence that the differing nature of these two aromatic pairs correlates with their distinct roles in the structure of the native protein: One (between F50 and F65) connects strands stabilized by canonical hydrogen bonds, while the other (between F65 and F71) constitutes a key linkage between two strands that are non-hydrogen bonded. Along with well-documented mobility of the helical portal,^{23,24,25} this latter aromatic pair appears to allow a functionally critical conformational change to take place in the iLBPs, opening the barrel and enabling ligand entry, without losing the integrity of the barrel. Our observations suggest that this role as a key stabilizing feature across non-hydrogen bonded strands may be operative in several other protein families and may be a common structural role of aromatic cross-strand interactions.

RESULTS

The Phe ladder in CRABP1

The native structure of CRABP1 is comprised of a 10-stranded α -barrel made up of two orthogonal antiparallel up-down five-stranded β -sheets and closed by interactions between strand 1 and 10. Contiguous strands are separated by reverse turns, with the exception of strands 1 and 2, which are connected by a helix-turn-helix, and strands 7 and 8, which are connected by an α loop. Phe residues at positions 50, 65, and 71 form an array of two linked aromatic pairs (a 'Phe ladder'²²) across strands 3, 4 and 5 (Fig. 1a). Inspection of both the F50-F65 pair across strands 3 and 4 and the F65-F71 pair across strands 4 and 5 reveals that they meet all criteria for aromatic-aromatic pairs: Using the holoenzyme structure (PDB ID, 1CBI; the numbers are only slightly altered in the structure of apo-CRABP1, 1CBR), the center-of-mass (COM) to COM distances are 5.18 Å for F50 to F65, and 4.91 Å for F65 to F71, the closest distances between the rings are 3.79 Å (F50 to F65) and 3.99 Å (F65 to F71) and the interring dihedral angles, 9.7° (F50 to F65) and 17.8° (F65 to F71), show these residues to be in a stacked aromatic trimer (see Materials and Methods). Note that in an earlier study²² we included F3 as a participant in the aromatic cluster in CRABP1. To enable clear comparisons among related proteins and to use a definition that can be applied generally, we have focused the present analysis on the three Phe residues that reside on adjacent strands and interact in a cross-strand manner: F50, F65, and F71.

Aromatic pairs on strands 3 and 4 and strands 4 and 5 differentially contribute to CRABP1 native state stability

In order to dissect the relative contributions of the aromatic interactions in the cross-strand Phe ladder in CRABP1 we mutated each of the three Phe residues, F50, F65 and F71, to Met. Substitution of Phe by Met preserves the ability of the side chain to participate in hydrophobic interactions, because Met is hydrophobic in nature and has the closest accessible surface area (ASA) to Phe,^{26,27,28} but removes the contribution from aromaticity. Hence, if the dominant interactions of a particular Phe side chain are hydrophobic then its substitution to Met should result in smaller destabilization compared to the case when interactions are predominantly aromatic. All three mutants were destabilized compared to the WT CRABP1 (Fig. 1b and Table 1). As expected, the strongest effect was associated with the F65M substitution (3.3 ± 0.2 kcal/mol), because this mutation results in simultaneous loss of both aromatic pairs (F50-F65 and F65-F71), ultimately breaking the whole aromatic ladder. The stability loss in CRABP1 is close to the estimates in the literature of the stabilization energy for a ladder-arranged aromatic trimer (4.8 kcal/mol).²⁹ Notably, the

destabilization caused by disruption of the F50-F65 pair by the F50M substitution was significantly smaller (2.0 ± 0.2 kcal/mol) than that caused by breaking the F65-F71 pair by the F71M substitution (2.7 ± 0.2 kcal/mol). This result argues for a greater aromatic contribution to the F65-F71 interaction, and a greater hydrophobic contribution to the F50-F65 pair interaction. While an earlier analysis found that all three Phe residues constituting the ladder participate in the highly networked conserved hydrophobic core of CRABP1,²² F50 was found to have the highest number of conserved hydrophobic contacts of the three Phe's. In terms of the respective roles of the Phe-Phe pairs in the native structure of CRABP1, the pair bridging strands 4 and 5 provides more stability than the one linking strands 3 and 4.

True aromatic interactions have well-defined geometry because of the spatial distribution of the ring electrons. One would therefore expect that aromatic pairs interacting via their systems would be more restricted in terms of ring-to-ring distance and dihedral angles than those interacting primarily hydrophobically. Conversely, the pairs where hydrophobic interactions are dominant would be expected to allow more distance and angle variation. Indeed, in an 8 ns MD simulation of CRABP1 starting with the X-ray structure of apo-CRABP1 (1CBI) the distance between the phenyl rings of F65 and F71 and their mutual ring orientation remained constant, while the distance between the COM of F50 and F65 and their relative ring orientation fluctuated significantly (Fig. 1c). These MD results support our conclusion based on the Met substitution that the Phe residues and corresponding pairs constituting the cross-strand ladder act differently from one another.

The timing of formation of the aromatic ladder in CRABP1 folding

How critical are cross-strand aromatic interactions for the folding of the β -sheet as a whole and when do they contribute to folding? CRABP1 folds via several kinetic phases, with an early hydrophobic collapse (μ s to ms), an intermediate timescale phase (100 ms) in which native-state topology forms but without interstrand hydrogen bonding or specific side chain packing, and a rate-determining 1 s phase in which barrel hydrogen bonding and all specific side chain packing develop.^{20,21} Here, we asked whether specific interactions between aromatic rings occur before or after the consolidation of the β -sheet in CRABP1. In other words: Are these interactions present in the transition state (TS) or not?

We probed the timing of formation of aromatic interactions in the cross-strand Phe ladder during folding of CRABP1 using perturbation analysis (also referred to as Δ -value analysis).³⁰ As noted above, the rate-determining TS of CRABP1 corresponds to the onset of interstrand H-bonding in the β -barrel.^{20,21} To probe the rate-determining TS of CRABP1 we first measured unfolding kinetics. Unfolding of all variants was significantly accelerated in comparison to that of WT CRABP1 (Fig. 2a and Table 1). Using G_{U-N} from equilibrium unfolding and $G_{\ddagger-N}$ from unfolding kinetics data, we calculated Δ -values (the ratio of $G_{\ddagger-N}$ to G_{U-N}). All Δ -values were close to 1.0 (Table 1) indicating that specific native-like interactions involving the Phe residues formed after the rate-determining TS.

Note that the slope of the unfolding branch of the chevron plot (m_U) for F71M was significantly decreased from that of WT, while m_U values for the other Phe mutants were relatively unchanged (within 14%) from that of WT. The smaller m_U for F71M CRABP1 indicates a smaller change in accessible surface area (ASA) between the native and rate-determining TS.³¹ The ASA of both the denatured and native states of all variants were unaltered from those of WT as judged by minimal and local perturbations of their denatured states (Fig. S2) and unperturbed equilibrium m -values (Fig. 1b). Therefore, we conclude that the F71M mutation increases the compactness of the rate-determining TS (Hammond behavior);³² similar effects have been previously observed for other CRABP1 variants.³³

Refolding kinetics for the Phe mutants fully validate the conclusion that the specific interactions involving the Phe side chains develop late. All Phe to Met variants showed minimally perturbed refolding behavior in stopped-flow experiments (Fig. 2b): Neither the fast phase nor the medium one was significantly affected by the mutation. Thus, the refolding data further argue for the onset of specific aromatic interactions within the Phe ladder after the formation of stable interstrand H-bonds in the β -sheets and therefore fully support the results of perturbation analysis based on $\Delta\Delta G$ -values. Together these pieces of evidence strongly suggest minimal perturbation of the rate-determining TS of F50M, F65M and F71M relative to the WT CRABP1, which corresponds to the onset of specific aromatic interactions after the formation of interstrand H-bonds. It is necessary to underline that perturbation analysis reports on specific side chain interactions and that aromatic residues may engage in non-specific hydrophobic interactions as early as during hydrophobic collapse.

Loss of aromatic interactions between non-H-bonded strands results in minor local perturbation, but widening of the strand 4 – strand 5 ‘gap’

To further explore the role of cross-strand aromatic interactions in the β -sheet structure of CRABP1 we asked whether the loss of these interactions causes long-range or local structural perturbation. Amide chemical shifts provide site-specific probes that are extremely sensitive to minor changes in protein conformation. Strikingly, chemical shift perturbation analysis (CSP) of backbone amides of the three Phe to Met CRABP1 variants showed no significant changes beyond a radius of 12 Å from the corresponding mutation site for the F50M, F65M and F71M variants (red regions in Fig. 3). The patterns of residues involved in H-bonds also remain virtually unperturbed in all three mutant proteins with respect to WT CRABP1 as revealed by amide proton temperature coefficients and ^1H - ^{15}N HSQC hydrogen-deuterium exchange experiments (Fig. S3). Finally, each of the three substitutions had only a minor effect on the overall integrity of the CRABP1 native structure as judged by far-UV CD, ligand binding, and Trp fluorescence emission spectra (Fig. S4). Combined, these results provide strong evidence that the loss of either of the two cross-strand aromatic interactions in CRABP1 does not perturb sheet structure globally, but instead results in an overall destabilization of the native structure.

While the Phe mutations do not cause significant change in the average global structure of CRABP1, loss of these aromatic interactions may lead to local structural perturbation and/or increased dynamics not revealed in the battery of experiments above. To explore this possibility, we probed potential local structural and dynamic perturbations by MD simulations of WT CRABP1 and the Phe to Met variants. MD simulations of the Phe to Met variants (F50M, F65M or F71M) showed that the loss of the aromatic side chain at any of these Phe positions results in significant reorientation of strands 3 and 4 relative to strand 5 without disrupting hydrogen bonding (as illustrated for the F71M variant in Fig. 4a and b). The consequence of this change is a widening of the ‘gap’ between strands 4 and 5, the only pair of strands in CRABP1 without stable hydrogen-bonding. Indeed, the reorientation of the gap neighboring strands observed by MD simulations is highly reminiscent of the conformational transition between apo- and holo-CRABP1 observed in the corresponding X-ray structures^{34,35} (compare Fig. 4b and c), suggesting that either the presence of the aromatic-aromatic interactions across the gap or the presence of the bound ligand offer stabilizing contributions to maintain the barrel structure.

The cross-strand aromatic pair on non-H-bonded strands 4 and 5 is a conserved feature in the iLBP family

Can these findings be extrapolated to other iLBPs? Phe ladders similar to the one present in CRABP1²² have been found in several iLBP family members, e.g. intestinal fatty-acid

binding protein (IFABP)³⁶ and cellular retinol-binding proteins (CRBPs) 1 and 2³⁷. To assess how conserved this ladder is in other iLBPs we performed a systematic search on the level of primary and tertiary structures. First, we asked how often aromatic residues (Phe, Tyr, Trp) occur in positions corresponding to the cross-strand Phe ladder in CRABP1 in the primary sequences of iLBPs. Multiple sequence alignment (MSA) of iLBPs by ConSurf³⁸ yielded 70 unique sequences. The conservation of aromatic side chains at the positions corresponding to the Phe ladder in CRABP1 was 36.2% at position 50, 91.3% at position 65 and 95.6% at position 71. As noted above, the Phe residues constituting the ladder are in the conserved hydrophobic core with the residue at the position 50 being its hub in terms of hydrophobic contacts (six), unlike the peripheral residues at positions 65 and 71 (five and two contacts, respectively).²² Thus, the conservation of aromaticity is anti-correlated with the number of conserved hydrophobic contacts, suggesting differential roles of aromatic residues in the ladder in iLBPs other than CRABP1. We therefore conclude that the distinct nature of the interactions of the two cross-strand aromatic-aromatic pairs we observed in CRABP1 is a general feature of the iLBP aromatic ladders: In the pair across strands 3 and 4 overall hydrophobicity is more important than aromaticity, which is consistent with a lower conservation of aromatic residues at position 50 in CRABP1. By contrast, the strand 4 – strand 5 aromatic-aromatic pair (positions 65 and 71 in CRABP1 numbering) interacts by electron clouds, as well as by hydrophobicity.

As mentioned above, one would expect that aromatic pairs interacting primarily by hydrophobic effects would have less restrictive geometry than those involved in aromatic interactions. Indeed, the geometries of cross-strand aromatic pairs in a representative set of 42 iLBP structures showed considerably higher variations in COM-COM distances and dihedral angles between side chains for the pair on strands 3 and 4 than for the pair on strands 4 and 5 (Supplementary Table S1). Note that this conclusion is based on a significant sample size: The pair on H-bonded strands 3 and 4 (corresponding to F50-F65 in CRABP1) was present in 24 structures (57%), while the one on non-H-bonded strands 4 and 5 (F65-F71 in CRABP1) – in 31 structures (74%). The full two-pair cross-strand ladder was found in one third (14 out of 42) of iLBP structures (Fig. 5). Average COM-COM distances in the strand 4 – strand 5 pair are similar to those for the strand 3 – strand 4 one (5.8 Å vs. 5.5 Å), but the standard deviation is almost threefold smaller (0.5 Å vs. 1.3 Å). The average dihedral angle for the strand 4 – strand 5 pair compared to the strand 3 – strand 4 pair is smaller (15° vs. 20°) as well as its standard deviation (11° vs. 15°). Thus, the conserved pair on non-H-bonded strands is more co-planar and allows significantly less angle and COM-COM distance variation, which is in excellent agreement with MD simulations. Interestingly, we observed that strand 5 is significantly shorter in the absence of the cross-strand aromatic pair between strands 4 and 5 (at least in three cases, PDB IDs 1LFO, in which strand 5 is disordered, and 1O1U and 1P6P, in which strand 5 is three residues in length as compared to the seven residue length of CRABP1); in the presence of the pair, strand 5 always has four or more residues.

Cross-strand aromatic pairs on non-H-bonded strands – beyond iLBPs

What can be learned from cross-strand aromatic pairs in non-iLBP proteins? In particular, is there a general role of cross-strand aromatic pairs in stabilizing non-H-bonded strand-strand arrangements? We sought to further generalize our hypothesis by analyzing the frequency of occurrence of cross-strand pairs as a function of interstrand distance. First we calculated the average C – C distances between two strands for all neighboring strands in the PDBSelect³⁹ database and for those that carry a cross-strand aromatic pair. Distance distribution plots each yielded a single dominant peak at about 5.0 Å corresponding to the average C – C distance between H-bonded strands (Fig. S5). Dividing the number of neighboring strands carrying a cross-strand aromatic pair by the total number of neighboring

strands we calculated the frequency of occurrence of cross-strand pairs at each interstrand distance (Fig. 6). The frequency plot shows a small but significant preference for cross-strand aromatic pairs on non-H-bonded strands with a maximum at interstrand distances about 6.5 Å. Thus, we conclude that cross-strand pairs indeed preferentially occur on non-H-bonded strands.

Finally, we addressed the question whether there are protein families besides iLBPs having a conserved cross-strand aromatic pair residing on non-H-bonded strands as a conserved feature of their structures. We considered the strands to be paired but non-H-bonded if they were more than 6.5 Å apart, setting a rigorous cut-off distance which is well above the average value of about 5.0 Å for H-bonded strands (the average C-C distance between the gap strands 4 and 5 in CRABP1 is 8.3 Å). We used an upper limit for interstrand distance of 10 Å (twice the normal distance for H-bonded strands), and inspected each example to insure that we were comparing appropriate structures. The search in the PDBSelect yielded 51 protein structures with aromatic pairs on non-H-bonded strands belonging to the same β -sheet (Table S2). We checked the conservation of the aromatic nature of corresponding side chains using ConSurf³⁸ in a way analogous to that for iLBPs. This allowed us to select a set of 17 proteins (Table 2) where aromaticity at the positions constituting the aromatic pair was more than 50% conserved. Importantly, this fully unbiased search over a non-redundant dataset identified an iLBP as a hit (rat IFABP, PDB ID 1IFC, Pfam family 00061). Another notable result is C-type lectin protein family (Pfam 00059), which showed three hits per family. Currently, for the majority of the pairs and/or proteins in the list there is no folding or energetic information to enable dissection of energetic contributions and timing of formation.

Overall, the results of the PDB-wide analysis strongly support the idea that aromatic interactions are widely employed by nature to stabilize β -strand – β -strand pairings in the absence of interstrand H-bonds.

DISCUSSION

Our results using CRABP1 as a case study to investigate how aromatic-aromatic interactions contribute to the formation and stabilization of β -structure highlight several important and general roles. We conclude that cross-strand aromatic interactions within the conserved Phe ladder in CRABP1 stabilize the native state of the protein; removal of the aromatic interactions in the Phe ladder by substituting Met, which has comparable solvent-accessible surface to Phe, led to reduction of native state stability of 2.0 and 2.7 kcal/mol, for the F50-F65 and F65-F71 pairs, respectively. Our kinetic analysis shows that these interactions form after the rate-determining TS. Thus, these interactions neither nucleate nor guide the specific folding of the β -barrel, as had been suggested early on by Burley and Petsko.¹⁵ We note that interpreting the role and timing of aromatic interactions during folding may depend on whether one is observing a specific or non-specific aromatic interaction. For example, studies on the denatured state of human bile acid-binding protein (BABP) argued that residues F47 (corresponding to F50 in CRABP1) and F63 (corresponding to F65 in CRABP2) formed part of an early-forming hydrophobic core in folding.⁴⁰ These observations implicate Phe residues in hydrophobic collapse, but do not report on a specific interaction, as revealed by π value analysis as done here. Similarly, detailed NMR studies using ¹⁹F-Phe to follow behavior of eight phenylalanines upon folding of a slow-folding variant of another iLBP, rat IFABP, found that there were early non-native clusters of Phe's, with later rearrangements to the final native structure.⁴¹ The model described by these authors is consistent with our conclusion that specific aromatic-aromatic interactions develop late. Late formation as we found in the case of CRABP1 implies that specific aromatic interactions *per*

se do not drive folding of β -sheets but instead are responsible for native state stability and for final fine-tuning of the sheet structure.

Notably, different aromatic pairs may have distinct roles depending on the structural and functional context. In the case of CRABP1, the cross-strand pair between non-H-bonded strands 4 and 5 is stabilized via aromatic interactions while that between strands 3 and 4 seems to be stabilized predominantly by hydrophobic interactions. Evidence for this conclusion included the lower cost of a Phe to Met substitution at this site, the observation of significantly less restricted geometry (indicated by variations among crystal structures and mobility in MD trajectories), lower conservation of aromatic residues at this site, and lower stabilizing energy to the protein structure.

Moreover, our analysis showed that aromaticity and interaction geometry of the aromatic pair spanning the non-H-bonded strands 4 and 5 is highly conserved in iLBPs. This pair appears to play a number of critical roles in this family. First, it enables the structural stabilization of a barrel without full H-bonding, which allows a larger strand-strand distance to stably exist: the 'gap'. We conclude that in this role aromatic interactions at least partially compensate for the absence of interstrand H-bonds. Second, it stabilizes the protein fold in the absence of a ligand. Ligands of iLBPs make contacts throughout the region of their β -barrels closest to the helix-turn-helix portal, as well as with residues on the helices themselves. As expected, in their apo-forms iLBPs are considerably less stable and more dynamic than when ligand-bound.^{23,24,25} The structural changes and increased dynamics of apo-iLBPs, including CRABP1, map to the helical portal, the turns between strands 3 and 4, and 5 and 6, and the interstrand distance and angle between strands 4 and 5. Nonetheless, the aromatic pair across strands 4 and 5 retains approximately the same geometry, suggesting that it plays a key role in tethering these two strands and maintaining the β -barrel architecture. In a related third role, the aromatic pair across strands 4 and 5 allows iLBPs to dynamically open to enable ligand entry and egress. The concerted motion of the helical portal, the turns, and the gap allows the typical iLBP hydrophobic ligand to gain access to its binding partner residues in the relatively inaccessible binding site. Lastly, strands 3 and 4 are predicted to be highly aggregation-prone in CRABP1, and this tendency is present in other iLBPs.³³ Increasing the gap opening and dynamics are expected to increase the risk of aggregation (M. C. Ferrolino, A. Zhuravleva, I. L. Budyak, and L. M. Gierasch, manuscript in preparation). It is thus tempting to speculate that another role of the conserved aromatic ladder in iLBPs is to counteract the increased aggregation propensity from exposure of aggregation-prone strands.

Interestingly, the degree of gap opening (i.e., the distance and orientation of strands 3–4 relative to the strand 5) varies between different iLBP family members and has been proposed to be tuned for ligand binding.³⁷ Our results are consistent with conclusions from previous NMR relaxation and NOE measurements on bovine heart FABP and porcine ILBP.⁴² Both proteins belong to the iLBP family, but in bovine heart FABP strands 4 and 5 are stabilized by a cross-strand aromatic pair (F64-F70) while in porcine ILBP the pair is absent. The authors have demonstrated that bovine heart FABP possesses a rather compact backbone structure with overall low flexibility. In contrast, individual residues in porcine ILBP displayed high backbone mobility with a large number of residues showing conformational exchange. Remarkably, the strongest dynamics were observed in β -strands 3 to 6. The greater flexibility of ILBP enables it to bind larger ligands as bile acid ligands are larger compared to fatty acids and need more space to enter and exit the protein. Interestingly, chicken liver BABP, which binds two ligand molecules in a form of a ternary complex⁴³ also lacks the cross-strand pair between the non-H-bonded strand.

Aromatic-aromatic cross-strand interactions emerge as a stabilizing feature that play a role comparable to H-bonding in specific contexts of protein structures. The particular example of stabilization across non-H-bonded β -strands, revealed here in the iLBPs, also occurs in other protein families, as our bioinformatics analysis revealed. We suggest that the range of possible structural and dynamic functions highlighted in the iLBPs be taken into consideration whenever this motif is identified in new protein families.

We need to underline here that stabilization brought by aromatic interactions is not restricted to β -structure or a certain class of proteins. On the level of secondary structure aromatic-aromatic interactions have been proposed to stabilize β -helices^{44,45} and helical bundles⁴⁶ including oligomers of transmembrane helices^{47,48}. On the level of folded proteins, aromatic clusters were found to be stability determinants of thermophilic proteins.⁴⁹ As for intermolecular interactions, protein-protein and protein-DNA recognition often involves aromatic-aromatic interactions. For instance, intrinsically disordered proteins frequently rely on aromatic residues in binding.⁵⁰ Moreover, π -stacking has been proposed to mediate intermolecular recognition and subsequent fibril assembly in amyloidogenic peptides.^{51,52} The wide spectrum of roles suggests high versatility of aromatic-aromatic interactions, thus offering a powerful tool in molecular design – from small molecules to protein complexes.

MATERIALS AND METHODS

Search for aromatic interactions between aromatic side chains in proteins

A non-redundant protein structure dataset PDBselect³⁹ (March 2012 release) was used for analysis. The set was populated by structures of predominantly soluble proteins with $\geq 25\%$ sequence identity, chain length 25–100,000 and ≤ 3.5 Å resolution. Corresponding atomic coordinates were retrieved from the Protein Data Bank (PDB, www.pdb.org). Phe, Tyr and Trp were considered aromatic; His was excluded due to the charge bias introduced.²⁹ Two aromatic residues were defined to form an aromatic pair through π -stacking if the closest distance between any two atoms (except hydrogens) of the rings was < 5.5 Å and the distance between ring COMs (calculated as a center of mass of six carbon atoms of the benzene ring for Phe and Tyr and as a center of mass of the five atoms of the pyrrole ring for Trp) was < 7.5 Å. The definition was based on the statistical analysis of all pairwise aromatic-aromatic interactions in the PDBSelect (Fig. S1) and is essentially the same as reported earlier.⁵³ For calculation of dihedral angles the ring planes were defined by ϕ_1 and ϕ_2 carbon atoms.

Interstrand distance calculation

To calculate the distance between two neighboring β -strands first the shorter strand was identified. Then the minimal distances between each of its carbon atoms and the carbon atoms of the longer strand were calculated, summed and averaged over the number of residues. Only strands belonging to the same β -sheet were considered. Secondary structure boundaries were taken from the corresponding PDB file records.

Sequence and structure conservation analysis

Query sequences were retrieved from the corresponding PDB files by ConSurf server³⁸ which then applied a BLAST search against Swiss-Prot database using a query sequence and performed MSA by MAFFT.⁵⁴ Identity cut-off limits were set to 95% (maximal) and 30% (minimal). Representative structures of iLBPs were acquired from the PDB using mouse CRABP1 sequence as a query; homologs identity cut-off was set to 95%. Three structures (1FTR, formylmethanofuran: tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri*; 3FLP, SAP-like pentraxin from *Limulus polyphemus*, and 3IO5, tryptophanyl-tRNA synthetase from *Trypanosoma brucei*) were discarded as these proteins

do not belong to the iLBP family; one structure, 1SA8, a helix-less variant of rat ILBP), was considered a duplicate of 1A57, a full-length rat apo-ILBP, and was also discarded.

MD simulations

All MD simulations were performed using GROMACS (version 4.5).⁵⁵ X-ray structure of apo-CRABP1 (PDB ID 1CBI, chain B) was taken as a starting point for the simulation of the WT CRABP1, starting structures for the F50M, F65M and F71M variants were generated in PyMol (www.pymol.org) by introducing the corresponding mutation. The starting structures were solvated in Simple Point Charge (Extended) (SPC/E) water; thickness of the explicit water layer was 1.4 nm from the protein molecule. The systems were neutralized with Na⁺ and Cl⁻ ions to bring the final salt concentration to 0.1 M and were equilibrated before and after insertion of ions. Production simulations were performed at constant temperature (300 K), pressure and number of particles. The all-hydrogen force field GROMOS96 43a2 were used for the simulations. Temperature and pressure were controlled using Berendsen weak coupling. All bonds with hydrogen atoms were constrained by the LINCS algorithm. Coulomb interactions were computed with the PME method. A 10 Å and 14 Å cut-offs were used for the electrostatic and van der Waals interactions; a neighbor list for non-bonded interactions based on a cut-off of 10 Å.

Protein design, expression and purification

Site-directed mutagenesis was done using QuikChange protocol (*Stratagene*). CRABP1 (a variant of CRABP1 with N-terminal 10His-tag and a stabilizing R131Q mutation,⁵⁶ referred to as WT CRABP1 throughout this article) and its variants were expressed in *Escherichia coli* strain BL21(DE3) (*Novagen*) and purified via Me-affinity chromatography as described elsewhere.²¹ Protein concentration was determined spectrophotometrically ($\epsilon_{280} = 20,970 \text{ M}^{-1}\text{cm}^{-1}$).

CD spectroscopy

Protein samples were prepared in 10 mM Tris-HCl pH 8.0, 1 mM DTT. CD spectra were collected using a Jasco J-715 spectropolarimeter in a 2 mm cuvette at 25 °C.

NMR spectroscopy

All NMR experiments were measured on a 700 MHz Varian NMR spectrometer equipped with a triple-resonance cryoprobe. Spectra were processed using NMRPipe⁵⁷ and analyzed using Cara⁵⁸. Backbone assignments for the apo-WT CRABP1 (500 μM ¹³C, ¹⁵N-labeled protein in 10 mM Tris-HCl pH 8.0, 1 mM DTT, 5% D₂O at 25 °C) were obtained from a standard set of triple-resonance experiments, including HNCO, HNCA, HNCACB, HN(CO)CACB and HN(CA)CO. Denatured samples were prepared in 8 M urea, 10 mM Tris-HCl pH 4.0 (adjusted with deuterated acetic acid), 5 mM DTT, 10% D₂O.

Chemical shift perturbation analysis

For each residue, the total chemical shift difference was calculated as $\{\Delta H^2 + (0.154 \Delta N^2)\}$, where ΔH and ΔN are ¹H and ¹⁵N chemical shift differences between WT CRABP1 and its corresponding variant. Chemical shift differences were considered significant if either ΔH or ΔN was twofold larger than the corresponding chemical shift error (0.02 and 0.2 ppm for ¹H and ¹⁵N atoms, respectively).

Temperature coefficients

Sets of ¹H-¹⁵N HSQC spectra were acquired at 5, 10, 15, 20, 25 and 30 °C on samples containing 150 μM of ¹⁵N-labeled WT CRABP1 or its variant in 20 mM deuterated Tris-HCl pH 8.0, 1 mM DTT, 5% D₂O. Assignments were gradually transferred from backbone

assignments obtained at 25 °C using a peak-working pattern; assignments at 10 °C were further validated using the HNCA experiment. At each temperature the ^1H chemical shifts were referenced externally to DSS. Temperature coefficients were calculated as the best fit to measured ^1H chemical shifts versus temperature.

Hydrogen-deuterium exchange

Samples of ^{15}N -labeled WT CRABP1 and its variants lyophilized in 20 mM deuterated Tris-HCl pH 8.0, 1 mM DTT were reconstituted in D_2O up to yield the final protein concentration of 100 μM . Sets of ^1H - ^{15}N HSQC spectra were recorded at 10 °C at three time points (1, 2 and 18 hours upon reconstitution in D_2O).

Ligand binding assay

Ligand binding assay was performed following the fluorescence quenching method.⁵⁹ Briefly, Trp fluorescence spectra (excitation at 280 nm, emission at 290–400 nm) were acquired either with or without equimolar amount of RA present in solution. The added volume of concentrated RA stock in absolute ethanol did not exceed 1% of the final volume. RA stock concentrations were determined from the absorbance at 336 nm ($\epsilon_{336} = 45,000 \text{ M}^{-1}\text{cm}^{-1}$).

Equilibrium unfolding

Protein samples were equilibrated at 25 °C overnight in solutions of 10 mM Tris-HCl pH 8.0, 1 mM DTT containing varying concentrations of urea. Unfolding transitions were monitored by Trp fluorescence (excitation at 280 nm, emission at 350 nm) at 25 °C. The data were analyzed by a two-state model according to the linear extrapolation method.⁶⁰

Unfolding and refolding kinetics

To induce unfolding, protein samples in 10 mM Tris-HCl pH 8.0, 1 mM DTT were manually added to urea solutions in the same buffer and rapidly mixed. The unfolding process was monitored following Trp fluorescence (excitation at 280 nm, emission at 350 nm) at 25 °C. The rate constants of unfolding (k_U) were determined from single-exponential fits with a linear component introduced to account for bleaching.

Refolding kinetics were followed by Trp fluorescence (excitation at 280 nm, emission at 350 nm) at 25 °C using an SFM-400 stopped-flow device (BioLogic) with a 0.8 mm cuvette. Urea-denatured protein (about 100 μM) in 20 mM Tris-HCl pH 8.0, 1 mM DTT was refolded by dilution into refolding buffer (20 mM Tris-HCl pH 8.0, 1 mM DTT containing varying urea concentrations). The dead time of the instrument was 2.4 ms. Kinetic traces were averaged, and folding rate constants (k_F) were determined from multi-exponential fits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from the National Institutes of Health (OD000945).

ABBREVIATIONS

ASA	accessible surface area
BABP	bile acid-binding protein

CD	circular dichroism
COM	center-of-mass
CRABP1	cellular retinoic acid-binding protein 1
CRBP	cellular retinol-binding protein
FABP	fatty acid-binding protein
H-bond(ed)	hydrogen bond(ed)
HSQC	heteronuclear single quantum coherence
IFABP	intestinal fatty acid-binding protein
iLBP	intracellular lipid-binding protein
MD	molecular dynamics
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
TS	transition state
WT	wild-type

References

1. Waters ML. Aromatic interactions in model systems. *Curr Opin Chem Biol.* 2002; 6:736–41. [PubMed: 12470725]
2. Chakrabarti P, Bhattacharyya R. Geometry of nonbonded interactions involving planar groups in proteins. *Prog Biophys Mol Biol.* 2007; 95:83–137. [PubMed: 17629549]
3. Malkov SN, Zivkovic MV, Beljanski MV, Hall MB, Zaric SD. A reexamination of the propensities of amino acids towards a particular secondary structure: classification of amino acids based on their chemical structure. *J Mol Model.* 2008; 14:769–75. [PubMed: 18504624]
4. Thomas A, Meurisse R, Charlotiaux B, Brasseur R. Aromatic side-chain interactions in proteins. I Main structural features. *Proteins.* 2002; 48:628–34. [PubMed: 12211030]
5. Horwich A. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J Clin Invest.* 2002; 110:1221–32. [PubMed: 12417558]
6. Jahn TR, Radford SE. The Yin and Yang of protein folding. *FEBS J.* 2005; 272:5962–70. [PubMed: 16302961]
7. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol.* 1998; 277:985–94. [PubMed: 9545386]
8. Istomin AY, Jacobs DJ, Livesay DR. On the role of structural class of a protein with two-state folding kinetics in determining correlations between its size, topology, and folding rate. *Protein Sci.* 2007; 16:2564–9. [PubMed: 17962408]
9. Chavez LL, Onuchic JN, Clementi C. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc.* 2004; 126:8426–32. [PubMed: 15237999]
10. Cochran AG, Skelton NJ, Starovasnik MA. Tryptophan zippers: stable, monomeric beta -hairpins. *Proc Natl Acad Sci USA.* 2001; 98:5578–83. [PubMed: 11331745]
11. Tatko CD, Waters ML. Selective aromatic interactions in beta-hairpin peptides. *J Am Chem Soc.* 2002; 124:9372–3. [PubMed: 12167022]
12. Mahalakshmi R, Raghothama S, Balaram P. NMR analysis of aromatic interactions in designed peptide beta-hairpins. *J Am Chem Soc.* 2006; 128:1125–38. [PubMed: 16433528]

13. Sonti R, Rai R, Ragothama S, Balam P. NMR analysis of cross strand aromatic interactions in an 8 residue hairpin and a 14 residue three stranded beta-sheet peptide. *J Phys Chem B*. 2012; 116:14207–15. [PubMed: 23163345]
14. Chung DM, Dou Y, Baldi P, Nowick JS. The absence of favorable aromatic interactions between beta-sheet peptides. *J Am Chem Soc*. 2005; 127:9998–9. [PubMed: 16011353]
15. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*. 1985; 229:23–8. [PubMed: 3892686]
16. Rodriguez HM, Vu DM, Gregoret LM. Role of a solvent-exposed aromatic cluster in the folding of *Escherichia coli* CspA. *Protein Sci*. 2000; 9:1993–2000. [PubMed: 11106173]
17. Betts S, Haase-Pettingell C, Cook K, King J. Buried hydrophobic side-chains essential for the folding of the parallel beta-helix domains of the P22 tailspike. *Protein Sci*. 2004; 13:2291–303. [PubMed: 15322277]
18. Kong F, King J. Contributions of aromatic pairs to the folding and stability of long-lived human gammaD-crystallin. *Protein Sci*. 2011; 20:513–28. [PubMed: 21432932]
19. Schaap FG, van der Vusse GJ, Glatz JF. Evolution of the family of intracellular lipid binding proteins in vertebrates. *Mol Cell Biochem*. 2002; 239:69–77. [PubMed: 12479570]
20. Clark PL, Liu ZP, Rizo J, Gierasch LM. Cavity formation before stable hydrogen bonding in the folding of a beta-clam protein. *Nat Struct Biol*. 1997; 4:883–6. [PubMed: 9360599]
21. Clark PL, Weston BF, Gierasch LM. Probing the folding pathway of a beta-clam protein with single-tryptophan constructs. *Fold Des*. 1998; 3:401–12. [PubMed: 9806942]
22. Gunasekaran K, Hagler AT, Gierasch LM. Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins*. 2004; 54:179–94. [PubMed: 14696180]
23. Krishnan VV, Sukumar M, Gierasch LM, Cosman M. Dynamics of cellular retinoic acid binding protein I on multiple time scales with implications for ligand binding. *Biochemistry*. 2000; 39:9119–29. [PubMed: 10924105]
24. Hodsdon ME, Cistola DP. Discrete backbone disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid-binding protein: implications for the mechanism of ligand entry. *Biochemistry*. 1997; 36:1450–60. [PubMed: 9063893]
25. Eberini I, Rocco AG, Ientile AR, Baptista AM, Gianazza E, Tomaselli S, Molinari H, Ragona L. Conformational and dynamics changes induced by bile acids binding to chicken liver bile acid binding protein. *Proteins*. 2008
26. Creamer TP, Srinivasan R, Rose GD. Modeling unfolded states of proteins and peptides. II Backbone solvent accessibility. *Biochemistry*. 1997; 36:2832–5. [PubMed: 9062111]
27. Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. *Protein Sci*. 2003; 12:1406–17. [PubMed: 12824487]
28. Gong H, Rose GD. Assessing the solvent-dependent surface area of unfolded proteins using an ensemble model. *Proc Natl Acad Sci USA*. 2008; 105:3321–6. [PubMed: 18305164]
29. Lanzarotti E, Biekofsky RR, Estrin DA, Marti MA, Turjanski AG. Aromatic-aromatic interactions in proteins: beyond the dimer. *J Chem Inf Model*. 2011; 51:1623–33. [PubMed: 21662246]
30. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol*. 1992; 224:771–82. [PubMed: 1569556]
31. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci*. 1995; 4:2138–48. [PubMed: 8535251]
32. Matouschek A, Fersht AR. Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proc Natl Acad Sci USA*. 1993; 90:7814–8. [PubMed: 8356089]
33. Budyak IL, Krishnan B, Marcelino-Cruz AM, Ferrolino MC, Zhuravleva A, Gierasch LM. Early folding events protect aggregation-prone regions of a beta-rich protein. *Structure*. 2013; 21:476–485. [PubMed: 23454187]

34. Thompson JR, Bratt JM, Banaszak LJ. Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular beta-sheet. *J Mol Biol.* 1995; 252:433–46. [PubMed: 7563063]
35. Kleywegt GJ, Bergfors T, Senn H, Le Motte P, Gsell B, Shudo K, Jones TA. Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure.* 1994; 2:1241–58. [PubMed: 7704533]
36. Li H, Frieden C. NMR studies of 4-19F-phenylalanine-labeled intestinal fatty acid binding protein: evidence for conformational heterogeneity in the native state. *Biochemistry.* 2005; 44:2369–77. [PubMed: 15709749]
37. Franzoni L, Cavazzini D, Rossi GL, Lucke C. New insights on the protein-ligand interaction differences between the two primary cellular retinol carriers. *J Lipid Res.* 2010; 51:1332–43. [PubMed: 19965581]
38. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 2010; 38:W529–33. [PubMed: 20478830]
39. Griep S, Hobohm U. PDBselect 1992–2009 and PDBfilter-select. *Nucleic Acids Res.* 2009; 37:D318–19. [PubMed: 19783827]
40. Basehore HK, Ropson IJ. Residual interactions in unfolded bile acid-binding protein by 19F NMR. *Protein Sci.* 2011; 20:327–35. [PubMed: 21280124]
41. Li H, Frieden C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and 19F NMR. *Proc Natl Acad Sci USA.* 2007; 104:11993–8. [PubMed: 17615232]
42. Lucke C, Fushman D, Ludwig C, Hamilton JA, Sacchettini JC, Ruterjans H. A comparative study of the backbone dynamics of two closely related lipid binding proteins: bovine heart fatty acid binding protein and porcine ileal lipid binding protein. *Mol Cell Biochem.* 1999; 192:109–21. [PubMed: 10331665]
43. Eliseo T, Ragona L, Catalano M, Assfalg M, Paci M, Zetta L, Molinari H, Cicero DO. Structural and dynamic determinants of ligand binding in the ternary complex of chicken liver bile acid binding protein with two bile salts revealed by NMR. *Biochemistry.* 2007; 46:12557–67. [PubMed: 17929837]
44. Bhattacharyya R, Samanta U, Chakrabarti P. Aromatic-aromatic interactions in and around alpha-helices. *Protein Eng.* 2002; 15:91–100. [PubMed: 11917145]
45. Butterfield SM, Patel PR, Waters ML. Contribution of aromatic interactions to alpha-helix stability. *J Am Chem Soc.* 2002; 124:9751–5. [PubMed: 12175233]
46. Aravinda S, Shamala N, Das C, Sriranjini A, Karle IL, Balaram P. Aromatic-aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues. *J Am Chem Soc.* 2003; 125:5308–15. [PubMed: 12720442]
47. Johnson RM, Hecht K, Deber CM. Aromatic and cation-pi interactions enhance helix-helix association in a membrane environment. *Biochemistry.* 2007; 46:9208–14. [PubMed: 17658897]
48. Haeger S, Kuzmin D, Detro-Dassen S, Lang N, Kilb M, Tsetlin V, Betz H, Laube B, Schmalzing G. An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nat Struct Mol Biol.* 2010; 17:90–8. [PubMed: 20023641]
49. Kannan N, Vishveshwara S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* 2000; 13:753–61. [PubMed: 11161106]
50. Espinoza-Fonseca LM. Aromatic residues link binding and function of intrinsically disordered proteins. *Mol Biosyst.* 2012; 8:237–46. [PubMed: 21863198]
51. Gazit E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* 2002; 16:77–83. [PubMed: 11772939]
52. Bemporad F, Taddei N, Stefani M, Chiti F. Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase. *Protein Sci.* 2006; 15:862–70. [PubMed: 16600970]
53. McGaughey GB, Gagne M, Rappe AK. pi-Stacking interactions. Alive and well in proteins. *J Biol Chem.* 1998; 273:15458–63. [PubMed: 9624131]

54. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002; 30:3059–66. [PubMed: 12136088]
55. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013; 29:845–854. [PubMed: 23407358]
56. Zhang J, Liu ZP, Jones TA, Gierasch LM, Sambrook JF. Mutating the charged residues in the binding pocket of cellular retinoic acid-binding protein simultaneously reduces its binding affinity to retinoic acid and increases its thermostability. *Proteins.* 1992; 13:87–99. [PubMed: 1377826]
57. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–93. [PubMed: 8520220]
58. Keller, R. PhD thesis. Swiss Federal Institute of Technology; 2004. Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment.
59. Cogan U, Kopelman M, Mokady S, Shinitzky M. Binding affinities of retinol and related compounds to retinol binding proteins. *Eur J Biochem.* 1976; 65:71–8. [PubMed: 945163]
60. Pace CN. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 1986; 131:266–80. [PubMed: 3773761]

HIGHLIGHTS

- a trimeric aromatic ladder is a conserved feature in iLBPs and forms late in folding
- one aromatic pair interacts primarily hydrophobically and the other also by stacking
- the aromatic pair across non-H-bonded strands 4 and 5 controls protein dynamics
- aromatic pairs may be a general mechanism to stabilize non-H-bonded strands

Fig. 1. Conserved Phe ladder in CRABP1, equilibrium stabilities of CRABP1 Phe variants, and MD-simulated geometries of the corresponding aromatic pairs

(a) Ribbon representation of mouse holo-CRABP1 X-ray crystal structure (PDB ID 1CBR) with strands 3, 4 and 5 labeled and F50, F65 and F71 constituting the cross-strand parallel-stacked aromatic ladder shown in spacefill and highlighted in orange. (b) Representative curves for urea-induced equilibrium denaturation of WT CRABP1 and the F50M, F65M and F71M variants. The solid lines on equilibrium denaturation plots are fits of the data to a two-state model using a fixed equilibrium m-value of -2.0 kcal/mol·M. (c) COM-COM distance (R_{COM}) and (d) dihedral angle (ϕ) variation within the F50-F65 (black) and the F65-F71 pairs (red) extracted from an 8-ns MD trajectory of WT CRABP1.

Fig. 2. Unfolding and refolding kinetics of CRABP1 Phe variants

(a) Dependence of the logarithm of the unfolding rates on [urea] of WT CRABP1 and the F50M, F65M and F71M variants. The solid lines on $\ln k_U$ plots represent linear fits. (b) Dependence of the logarithms of the rates of I_2 (fast phase, in squares) and N (slow phase, in triangles) formation (fast and medium kinetic phases, correspondingly) on [urea] for WT CRABP1 and the F50M, F65M and F71M variants. Colors are as in (a).

Fig. 3. Chemical shift perturbation analysis of CRABP1 Phe variants

(a) Histograms of chemical shift perturbation, $\Delta\delta$, between WT CRABP1 and the F50M (top panel), F65M (middle panel) and F71M (bottom panel) variants. Bars corresponding to the residues showing significant perturbation are colored in red ($\Delta\delta > 0.1$ ppm, large changes) and yellow ($\Delta\delta \leq 0.1$ ppm, moderate changes); those with insignificant perturbations are shown in gray. (b) The magnitude of $\Delta\delta$ for the F50M, F65M and F71M variants mapped onto the holo-CRABP1 structure (PDB ID 1CBR). Residues with large and moderate $\Delta\delta$ are colored in red and yellow, unperturbed residues are shown in gray (same as in (a)); unassigned residues are colored in black.

Fig. 4. Conformational flexibility of CRABP1 strands 3, 4 and 5

(a) Interstrand distance variation between strands 4 and 5 in CRABP1 extracted from an 8-ns MD trajectory of WT CRABP1 (black trace) and the F71M variant (red trace). The interstrand distance was defined as the distance between C_{α} atoms of T61 (strand 4) and E74 (strand 5), both shown as spheres in (b). (b) Overlay of the structures after 1.5 ns of MD simulation: WT CRABP1 (gray) and the F71M variant (red). The C_{α} atoms of T61 (strand 4) and E74 (strand 5) are shown as spheres. (c) Overlay of the X-ray structures of the holo-CRABP1 (PDB ID 1CBR, chain A (gray)) and apo-CRABP1 (PDB ID 1CBI, chains A (green) and B (blue)).

Fig. 5. The conserved aromatic ladder in iLBPs

Pie diagram representing the different arrangements of the cross-strand aromatic ladder and their respective occurrences in the representative set of iLBP structures acquired from the PDB. Black sector corresponds to the iLBPs having no aromatic ladder.

Fig. 6. Frequency of cross-strand aromatic pairs

Distribution of the ratio between the number of neighboring strands carrying a cross-strand aromatic pair over the total number of neighboring strands at a given interstrand distance (step size 0.2 Å). Black bars represent interstrand distances with good statistics (both numbers are > 5); gray bars represent the ones with poor statistics (either nominator or denominator is ≤ 5).

Thermodynamic and kinetic parameters for WT CRABP1 and its F50M, F65M and F71M variants used for perturbation analysis^a

Table 1

variant	$G_{U,N}^a$, kcal/mol	$G_{U,N}^b$, kcal/mol	$\ln k_U^c$	m_U , kcal/mol*M	$G_{\ddagger,N}^d$, kcal/mol	U
WT ^e	10 ± 0.2	n/a	-11.0 ± 0.1	0.64 ± 0.02	n/a	n/a
F50M	8.0 ± 0.2	2.0 ± 0.2	-7.6 ± 0.1	0.55 ± 0.02	1.99 ± 0.09	1.0 ± 0.1
F65M	6.7 ± 0.1	3.3 ± 0.2	-6.2 ± 0.1	0.60 ± 0.02	2.82 ± 0.09	0.86 ± 0.06
F71M	7.3 ± 0.2	2.7 ± 0.2	-6.8 ± 0.2	0.33 ± 0.02	2.5 ± 0.1	0.90 ± 0.09

^aThe average values from at least three independent measurements and the standard deviation are reported; all values reported are extrapolated to 0 M denaturant

^bThe energetic effect of mutations on the free energy of the unfolded state (U) with respect to the native state (N) calculated as $G_{U,N}^{WT} - G_{U,N}^{mut}$; $G_{U,N}$ values were obtained by fitting equilibrium urea unfolding curves using $m_{eq} = -2.0$ kcal/mol.M

^cObtained from linear extrapolation of $\ln k_U$ versus [urea]

^dThe energetic effect of mutation on the free energy of the transition state ($\ddagger$) with respect to the native state (N), calculated as $G_{\ddagger,N}^{WT} - RT \ln (k_U^{WT}/k_U^{mut})$

^eHis-tagged CRABP1 with a stabilizing R131Q mutation was used as wild type (WT)

Table 2

Proteins with conserved cross-strand aromatic pair(s) on non-H-bonded strands (interstrand distance > 6.5 Å) identified in the PDBSelect.

PDB ID <i>Pfam family</i>	protein (source organism)	aromatic pair(s) conservation (unique seq.) ^a	COM-COM distance	interstrand distance
1KT7 A <i>PF08212</i>	holo-RBP (<i>Bos taurus</i>)	W24 - F45 100% – 100% (9)	5.70 Å	6.99 Å
1QDD A <i>PF00059</i>	lithostathine (<i>Homo sapiens</i>)	W77 - F138 99.1% – 93.3% (106)	7.10 Å	8.16 Å
1IFC A <i>PF00061</i>	intestinal fatty acid-binding apoprotein (<i>Rattus norvegicus</i>)	F62 - F68 100% – 96.4% (28)	5.45 Å	8.27 Å
2QIM A <i>PF00407</i>	pathogenesis-related protein LIPR-10.2B (<i>Lupinus luteus</i>)	Y82 - F99 97.1% – 51.4% (35)	4.52 Å	6.58 Å
1HT6 A <i>PF00128</i>	alpha-amylase isozyme 1 (<i>Hordeum vulgare</i>)	F246 - F286 100% – 100% (11)	5.86 Å	6.64 Å
1L2H A <i>PF00340</i>	interleukin 1-beta mutant F42W/W120F (<i>Homo sapiens</i>)	F112 - F146 95.5% – 100% (22)	6.11 Å	8.79 Å
3SEB A <i>PF02876</i>	staphylococcal enterotoxin B (<i>Staphylococcus aureus</i>)	F137 - Y233 75% – 75% (8)	4.57 Å	6.52 Å
2Y7L A <i>PF11766</i>	agglutinin-like ALS9 protein, N- terminal domain residues 18-328 (<i>Candida albicans</i>)	F199 - Y274 100% – 100% (6)	5.71 Å	6.52 Å
1WMZ D <i>PF00059</i>	lectin CEL-I (<i>Cucumaria echinata</i>)	W72 - Y133 100% – 88.5% (28)	7.37 Å	8.09 Å
3QSQ A <i>PF03115</i>	capsid spike protein (human astrovirus 8)	F524 - F638 100% – 100% (8)	6.58 Å	9.61 Å
1R8N A <i>PF00197</i>	Kunitz trypsin inhibitor (<i>Delonix regia</i>)	Y20 - F63 100% – 58.3% (12)	6.54 Å	7.32 Å
3HQC A <i>PF08416</i>	phosphoTyr-binding domain from the tensin-like C1 domain- containing phosphatase (TENC1) (<i>Homo sapiens</i>)	Y1280 - F1384 100% – 100% (8)	4.70 Å	6.88 Å
1CKS C <i>PF01111</i>	cell cycle regulatory protein CksHs2 (<i>Homo sapiens</i>)	Y19 - Y57 100% – 100% (14)	5.74 Å	9.88 Å
1ILR 1 <i>PF00340</i>	interleukin-1 receptor antagonist (<i>Homo sapiens</i>)	F111 - F146 78.6% – 100% F111 - F148 78.6% – 85.7% (14)	6.22 Å 4.68 Å	7.61 Å 7.61 Å
3NO8 B <i>PF08005</i>	PHR domain from BTB/POZ domain- containing protein 2 (<i>Homo sapiens</i>)	F395 - F499 100% – 85.7% F395 - F501 100% – 100% (7)	6.65 Å 6.37 Å	6.97 Å 6.97 Å
2OX8 A <i>PF00059</i>	scavenger receptor with C-type lectin type I (<i>Homo sapiens</i>)	W665 - F728 100% – 87.3% (56)	7.05 Å	8.25 Å
2VQE Q <i>PF00366</i>	30S ribosomal protein S17 (<i>Thermus thermophilus</i>)	Y42 - F71 52.7% – 87.3% (150*) * 218 unique seq.	4.91 Å	9.26 Å

^aThe minimal number of unique sequences for MSA and conservation score calculation by ConSurf³⁸ is five, ten or more is recommended