J Mol Model (2017) 23:206
DOI 10.1007/s00894-017-3376-y

ORIGINAL PAPER

Salt-bridge networks within globular and disordered proteins: characterizing trends for designable interactions

Sankar Basu[1] · **Debasish Mukharjee**[2]

Received: 5 March 2017 / Accepted: 17 May 2017

Abstract There has been considerable debate about the contribution of salt bridges to the stabilization of protein folds, in spite of their participation in crucial protein functions. Salt bridges appear to contribute to the activity–stability trade-off within proteins by bringing high-entropy charged amino acids into close contacts during the course of their functions. The current study analyzes the modes of association of salt bridges (in terms of networks) within globular proteins and at protein–protein interfaces. While the most common and trivial type of salt bridge is the isolated salt bridge, bifurcated salt bridge appears to be a distinct salt-bridge motif having a special topology and geometry. Bifurcated salt bridges are found ubiquitously in proteins and interprotein complexes. Interesting and attractive examples presenting different modes of interaction are highlighted. Bifurcated salt bridges appear to function as molecular clips that are used to stitch together large surface contours at interacting protein interfaces. The present work also emphasizes the key role of salt-bridge-mediated interactions in the partial folding of proteins containing long stretches of disordered regions. Salt-bridge-mediated interactions seem to be pivotal to the promotion of "disorder-to-order" transitions in small disordered protein fragments and their stabilization upon binding. The results obtained in this work should help to guide efforts to elucidate the modus operandi of these partially disordered proteins, and to conceptualize how these proteins manage to maintain the required amount of disorder even in their bound forms. This work could also potentially facilitate explorations of geometrically specific designable salt bridges through the characterization of composite salt-bridge networks.

Keywords Salt bridges · Ioinic bonds · Motif identifier · Bifurcation angle · Bifurcated salt bridges · Molecular clips · Disordered proteins · Designable interactions

Introduction

Historically, it was believed that the role of salt bridges was primarily to maintain the global electrostatic balance within correctly folded globular proteins and at protein–protein interfaces [1, 2]. However, that theory had to be revised as it was found to be only partially true—at least in the context of protein–protein interfaces [3]—according to continuum electrostatic models that were employed to numerically solve the Poisson–Boltzmann equation for the protein–solvent system [4, 5]. It was categorically shown that the electrostatic complementarity (EC) at the protein–protein interface can remain high even when the salt bridge is computationally neutralized (or switched off). The reason for this is that electrostatic complementarity is a non-local effect governed by a long-range force wherein partial charges from all of the atoms in the two interacting molecules collectively contribute to the potentials; hence, switching off the salt bridge does not significantly alter the overall correlation. In fact, it was in the very first study of electrostatic complementarity at protein–protein

Electronic supplementary material The online version of this article (doi:10.1007/s00894-017-3376-y) contains supplementary material, which is available to authorized users.

✉ Sankar Basu
nemo8130@gmail.com

[1] Department of Biochemistry, University of Calcutta, 35, Ballygunge Circular Rd, Ballygunge, Kolkata, West Bengal 700019, India

[2] Computational Science Division, Saha Institute of Nuclear Physics, 1/AF, BidhanNagar, Kolkata, West Bengal 700064, India

Published online: 19 June 2017

interfaces [3] that the earlier idea of charge complementarity was virtually ruled out and replaced with the concept of electrostatic potential complementarity. In other words, the complementarity is actually attained in the anticorrelation of surface electrostatic potentials realized at protein–protein interfaces. A later study extended this concept to the folding of single-domain globular proteins and, for the first time, characterized the trends in electrostatic complementarity in a residue-wise manner within the protein interior [6]. In line with the results seen for protein–protein interfaces, elevated values of electrostatic complementarity (E_m) were attained for all completely or partially buried amino acid residues, irrespective of their chemical identity, in both the presence and absence of contributions from salt bridges. All said and done, the different functions attributed to these ionic bonds (ranging from anchoring protein subunits [7] to metal coordination and cooperativity in networks of salt bridges [8]), as well as their impact on stability [9, 10], cannot be undermined.

There does not appear to be a universal rule regarding the role of salt bridges in stabilizing protein structures, however. Due to desolvation effects, they are generally considered to be destabilizing [11], though instances have been observed where networks of ionic bonds contribute favorably to the thermal stabilization of proteins [10, 12, 13]. For instance, arginine-glutamate isolated salt bridges have been found to contribute favorably to the stability of helices [10], whereas bifurcated salt bridges have been shown to be instrumental in originating photoacoustic signals in cytochromes [14].

Although some empirical rules have been proposed regarding the geometry of hydrogen bonds within salt bridges [7, 13], and for distinguishing between simple and complex salt bridges [7], a comprehensive characterization of the various possible topological assemblies of salt-bridge networks (consistent with the growing number of high-resolution protein structures) appears to be lacking.

Against this background, the study reported in the present paper was performed to characterize salt-bridge networks within proteins in terms of topology, geometry, and composition. The study recognized bifurcated salt bridges to be a key motif found mostly within proteins and at protein interfaces. Statistical analyses of the functional attributes of bifurcated salt bridges were carried out; an exhaustive list of functionally relevant interactions was put forward; and some unique and interesting patterns were highlighted. The study also attempted to shed some light on the potential contributions of ionic bonds and charged residue contacts to the partial folding/stabilization of intrinsically disordered proteins (IDPs) in both their free and bound forms. The results of the study should aid the design of salt bridges in recombinant proteins and the computational modeling of interactions involving disordered protein regions.

Materials and methods

Databases

Globular proteins Following our previous reports in the related area of protein electrostatics [2], we accumulated salt bridges from a previously reported [6, 15] database (https://github.com/nemo8130/DB2) of high-resolution crystal structures of native globular proteins. The structures in that database (resolution ≤2 Å; R factor ≤ 20%; pairwise sequence similarity ≤30%) were devoid of any bulky prosthetic groups (e.g., protoporphyrins) and any missing atoms. That database was used to characterize the topological and geometric preferences of networks involving salt bridges.

Protein–protein complexes To study the role of salt bridges in triggering and stabilizing protein–protein interactions, another database containing high-resolution (≤2 Å) crystal structures of 1879 native protein–protein complexes was assembled. None of these structures contained any missing backbone atoms. This database was previously used to validate the structures of protein complexes [16] and also as a reality check when scoring protein–protein docking [17]. For structures containing more than two chains, the two largest interacting chains were considered for all subsequent calculations.

Proteins with intrinsically disordered regions To study the role played by salt bridges and atomic contacts involving charged residues in the stabilization of intrinsically disordered proteins, yet another nonredundant database of 109 polypeptide chains containing missing disordered loops was assembled. This database was a subset of a previously assembled set that was used in a recent study of the conformational entropy of intrinsically disordered proteins calculated from amino acid triads [18]. The original database [18] contained various types of possible experimental structures solved by X-ray crystallography, NMR, and electron microscopy (EM). Missing disordered residues were identified by comparing the SEQRES and ATOM records in the PDB files and then cross-checking the difference between the two sets with the missing residues declared in the REMARK 465 list. The database was initially compiled using the PISCES server [19], applying a sequence similarity of ≤25% and a length of ≥ 40 residues as selection criteria, which resulted in a total of 138 chains with >50% structural disorder [18]. As well as implementing the selection criteria above, we also discarded the EM structures and structures containing more than 700 residues, reducing the number of chains to 109.

Apart from the systematic calculations performed on the above database, two case studies of two popular individual systems involving disorder were also explored. The first of the two was the most popular tau protein (PDB IDs: 2N4R,

4TQE, 5DMG, and 5BTV); this protein is a potential causal factor in Alzheimer's disease, as abnormal phosphorylation of the protein results in paired helical filament and neurofibrillary tangles. The second candidate was the plant protein crambin, which has been solved at an ultrahigh resolution, highlighting the valence electron density (PDB ID: 1EJG); interestingly, all isoforms of crambin appear to be individually disordered (http://deposit.rcsb.org/format-faq-v1.html).

Modeling intrinsically disordered protein regions (IDPRs) in existing experimental structures

Disordered residues were identified and distinguished by the selection procedure elaborated in the "Databases" section. The missing disordered residues (as listed in the REMARK 465 of the corresponding PDB file) were then built by MODELER (version 9.17) [20], implementing its general "automodel" or advanced "loopmodel" module (the latter with the "fast refinement" option) as appropriate for each individual protein system. The experimental structure was used as a template for modeling, and the correct position of each missing residue was determined by a pairwise sequence alignment performed using the high-accuracy alignment program MUSCLE (version 3.8) [21]. Both the template and the alignment were then fed to MODELER. The root mean square (rms) deviation calculated from all pairs of corresponding atoms (i.e., the atoms solved experimentally) upon aligning the two structures was found to be ≤1.5 Å for all models. In each case, the appropriate MODELER module (automodel/loopmodel) was determined by performing a comparative study of the two initial models produced by the two methods; the one resulting in the lowest rms deviation was chosen. Once the correct module was determined for each protein, the program was rerun, generating 10 models. The model with the lowest rms deviation upon alignment with the corresponding experimental template was chosen (Fig. 1). Residues in the modeled structures were then renumbered (from 1) according to the SEQRES records in order to naturally accommodate the missing disordered residues while avoiding negative integers, and are reported that exact way here. Stride [22] was used to calculate the relative content of secondary structural elements in each modeled structure.

Burial of solvent exposure

The solvent-accessible surface area (ASA) was calculated for each atom in each protein molecule using NACCESS [23] by rolling a probe sphere of 1.4 Å over the entire protein surface (the Lee and Richards algorithm) [24]. The ASAs were then summed for all atoms pertaining to the same residue. The burial of solvent exposure for a residue, defined as the ASA of that residue located in the protein divided by the ASA of the same amino acid (X) in a Gly–X–Gly peptide fragment in its fully extended conformation, was then calculated according to a previously established method [25]. Finally, the completely exposed residues, i.e., those with burial of solvent exposure values of >0.3 [25], were identified and the atoms were demarcated as the peripheral atoms of the protein.

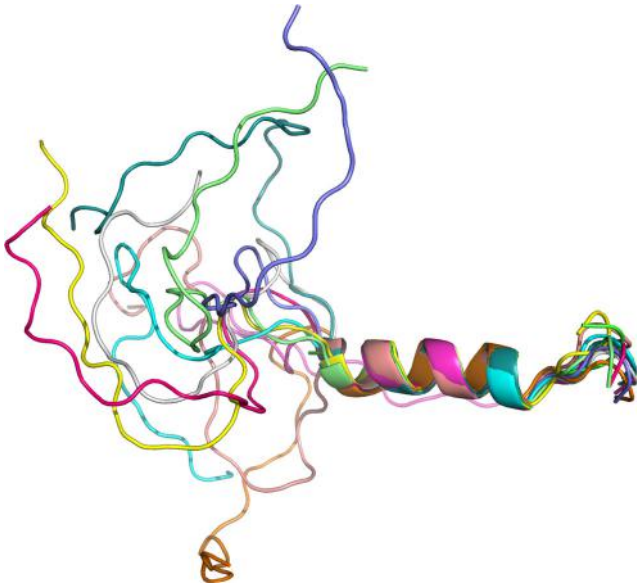

Fig. 1 Modeling IDPRs within partially disordered proteins. The example shows the top ten models created by the 'loopmodel' module of MODELER for the partially disordered protein 2MBH when modeling its missing IDPRs. All models were closely spaced in terms of agreeing with the native experimental template, and gave rise to rms deviations (calculated from all experimentally solved atoms) ranging from 0.461 Å to 1.018 Å upon superposition. The native and the best models (rmsd: 0.461 Å) are colored *light blue* and *light magenta*, respectively

Accessibility score

Depending on the location and extent of the disorder embedded in a protein sequence, the modeled protein can in principle be delirious and unreal. One way to estimate their fury is to check the expected distribution of amino acid residues with respect to their burial of solvent exposure and compare the values with globular proteins that are known to be stable in solution. This distribution was estimated via the accessibility score, *rGb*, which was previously standardized within globular proteins [15] and at protein–protein interfaces [16], and has been successfully used as a coordinate-driven feature when scoring protein–protein docking [17]. Briefly, the definition of the *rGB* score is as follows:

$$rGb = \frac{1}{N_{\mathrm{res}}} \sum_{i=1}^{N} \log_{10}(\mathrm{Pr}_i), \quad (1)$$

where N_{res} is the total number of residues in the polypeptide chain and $\mathrm{Pr_i}$ is the propensity of a particular amino acid (Val, Leu, Asp, etc.) to acquire a particular degree of solvent exposure. Other details of the *rGB* score are defined elsewhere

[15]. The *rGb* score is similar for both correctly folded globular proteins (0.055 ± 0.022) and protein–protein interfaces (0.058 ± 0.022). Based on these observations, the threshold for the successful validation of *rGb* was set to 0.011 ($\mu - 2\sigma$), where μ is the mean (0.055) and σ is the standard deviation (0.022) calculated from the database of globular proteins. Any structure attaining a value equal to or higher than this threshold should be a fairly well-folded globular protein or a well-packed interface. In principle, the score is also applicable to peptides, and a lower value than the threshold (which can also be negative) is indicative that the protein/peptide is unstable in solution. Such molecules will be highly reactive.

Defining salt bridges and their networks

The essential idea behind the analyses of the database of globular proteins was to consider salt bridges not merely as a collection of individual ionic bonds but as networks, and to perform an exhaustive statistical study of the topologies and geometric preferences of different salt-bridge networks. In those networks, charged residues were represented as nodes and connected by an edge when there was an ionic bond (or salt bridge) between them. An ionic bond was considered to be present when a positively charged nitrogen atom of lysine (NZ), arginine (NH1, NH2), or positively charged histidine (HIP: ND1 NE2, both protonated) were found to be within 4.0 Å of a negatively charged oxygen atom of glutamate (OE1, OE2) or aspartate (OD1, OD2).

In order to study the topological and geometric patterns of these networks constituted by ionic bonds, a previously proposed numerical scheme—the motif identifier—was adopted [26], which can be considered a reduced representation of a specific unique network topology. In this numerical representation, a unique network topology is represented by *n* concatenated strings of numbers separated by delimiters, where *n* is the number of nodes in the network. Each string begins with the degree of a node (source node), which is followed by the degrees of the nodes linked to it (neighbouring nodes), in order of their descending degrees. These numeric strings are then collected and concatenated further, in order of descending degrees of the source nodes (Fig. 2). The whole idea and formulation of the numerical topological scheme is based on the assumption that the number of direct neighbors of a node in such a network is always limited (due to steric and electrostatic constraints) to single-digit numbers (i.e., a maximum of 9). This assumption has generally been found to be true for a wide range of protein contact networks [26].

Thus, the motif identifier essentially discriminates between two graphs based on the combined distributions of the degrees of their constituent nodes coupled with the degrees of their neighboring nodes, and it will potentially flag up any variability in these network parameters between the two given graphs (Fig. 2). In principle, the only instance where it will fail to discriminate between two nonidentical networks is in the case of *k*-regular graphs [27] with $k > 2$. Since all the k nodes of a k-regular graph have identical degrees, the degree of any node and/or the degrees of its neighboring nodes cannot account for any variability, even if two *k*-regular graphs of the same size are topologically nonidentical. However, such graphs are irrelevant in the context of salt bridges due to steric and electrostatic constraints.

Propensities of amino acids to form salt bridges

The propensity (Pr(*x*,*s*)) of a charged residue *x* to form a salt bridge was computed by the following expression:

$$\Pr(x,s) = \frac{N(x,s)/N(t,s)}{N(x,d)/N(t,d)} \quad (2)$$

where $N(x,s)$ is the count of residue x found in salt bridges, $N(t,s)$ is the total number of all residues involved in salt bridges, and $N(x,d)$ and $N(t,d)$ are, respectively, the counts of residue x and the total number of residues in the database.

Electrostatic complementarity of charged amino acids

The electrostatic complementarity (E_m) of each charged amino acid residue within the protein interior was calculated according to a previous report [6], and a detailed statistical analysis was performed for each residue. To start with, hydrogen atoms were geometrically fixed to each individual protein chain in the database. Van der Waals surfaces were sampled at 10 dots/Å^2. The surface generation process is discussed in detail elsewhere [25]. The exposure of individual atoms to solvent was estimated by rolling a probe sphere of radius 1.4 Å over the protein atoms [24], and the burial of each residue was estimated by calculating the ratio of the solvent-accessible surface area of the charged amino acid X embedded in the polypeptide chain to that of an identical residue located in a Gly–X–Gly peptide fragment with a fully extended conformation. Partial charges and atomic radii for all protein atoms were assigned in accord with a previously used force field [28], and Asp, Glu, Lys, Arg, doubly protonated histidine (Hip), along with both carboxy and amino terminal groups were considered to be ionized/charged. Crystallographic water molecules and surface-bound ligands were excluded from the calculations and thus modeled as bulk solvent. Ionic radii were assigned to the bound metal ions according to their charges [29].

Delphi (version 6) [30] was used to compute the electrostatic potential of the molecular surface along the polypeptide chain according to a previous report [6]. For each run, a set of dot surface points on which the electrostatic potentials were to be computed was fed to Delphi along with a set of charged atoms contributing to the potential. First, the dot surface points of the individual amino acids (targets) were identified. The electrostatic potential for each residue surface was then

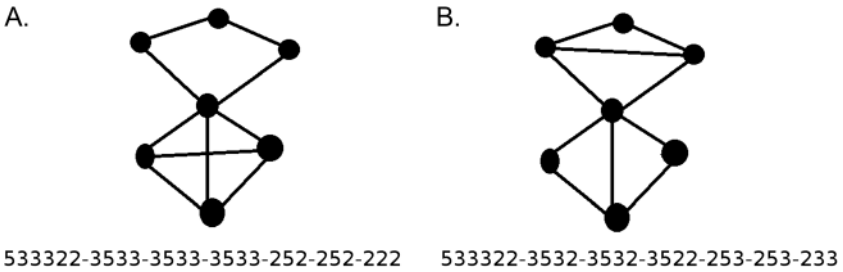

Fig. 2 The motif identifier: accounting for topological variations in graphs. The motif identifier (presented as a concatenated sorted numerical string) is a collection of numeric substrings, each of which is representative of a topological motif pertaining to a node (nodal motif). The first number in each numeric substring is the degree of the corresponding source node, and the other numbers represent the degrees of the direct neighbors of that source node, sorted in order of descending degree. All nodal motifs (numeric substrings) obtained in this way are collected together as elements of an array and further sorted in descending order. Finally, the sorted number strings (nodal motifs) are concatenated using a delimiter ("-"). Two distinct examples are shown to demonstrate the numeric scheme. Both graphs were generated from a common core backbone topology, and then a few links were rewired

calculated twice: due to the atoms of the particular target residue and from the rest of the protein, excluding the selected target amino acid. In both cases, the atoms that did not contribute to the potential were treated as dummy atoms, with only their radii being assigned with zero charge. Thus, each dot surface point on the target residue was tagged with two values of the electrostatic potential. The electrostatic potential complementarity (E_m) of an amino acid residue (in the protein interior) was then defined as the negative of Pearson's correlation coefficient between these two sets of potential values [6].

Again, the potential values corresponding to N dot surface points can be divided into two distinct sets, based on whether the dot point was sourced from main- or side-chain atoms of the target residue, and E_m can be obtained separately for each set. In the current study, however, we were only interested in the E_m obtained from the side-chain dot surface points (E_m^{sc}), since the charge is on the side-chain atoms in the context of ionic bonds.

All images representing protein structures were created in Pymol (The PyMOL Molecular Graphics System, version 1.8, Schrödinger, LLC.) for visual investigation and display.

Results and discussion

Characterizing salt-bridge networks within proteins

A total of 3076 networks were extracted from the globular protein database using the numerical scheme described in the previous section (see "Materials and methods"). The distribution of these networks was found to be dominated by isolated ionic bonds (11-11; 2445 networks, Fig. 3), followed by bifurcated salt bridges consisting of three nodes (211-12-12; 475 networks). The overwhelming majority of the networks with more than three nodes were open linear chains; in other words, trees [26] and their variants. This topological preference is undoubtedly due to the fact that adjacent nodes cannot carry alike charge. A few examples of four-membered closed rings either isolated or fused along an edge [26] were also found. For closed rings, the topological constraints imposed by charge imply that only even numbers of nodes are permitted.

Electrostatic complementarity

Overall, a mild enhancement in E_m was observed for charged residues involved in salt bridges (Fig. 4), with the exception of histidine, which was found to prefer metal coordination sites more than salt bridges [26]. The highest average value of E_m^{sc} was obtained for glutamate (0.68), which also showed the highest increase in $<E_m^{sc}>$ upon inclusion in a salt bridge. Arginine exhibited the highest propensity of any charged residue to form ionic bonds (5.83, versus Glu: 4.77, Asp: 3.92, Lys: 3.43). The participation of histidine in ionic bond networks was by and large negligible (propensity: 0.22).

Compositional and geometric preferences

Several instances of bifurcated salt bridges contributing more to the electrostatic stabilization of proteins than isolated ionic bonds have been reported [10, 13, 14]. With this in mind, bifurcated salt bridges (Fig. 5) were further analyzed for compositional and geometric bias (Table 1). Compositional preferences were clearly distinguishable for arginine-containing salt bridges (75.4% of the whole set), with Glu-Arg-Glu showing the highest occupancy (17%). Similar preferences have been previously observed for arginine–glutamate salt bridges in relation to helix stability [10].

The angle subtended by the three residues that form the bifurcated salt bridge was computed as follows. Except for lysine, which has a uniquely charged nitrogen atom (NZ), each effective or resultant charge-center was determined as the midpoint of the two degenerate charged O (aspartate, glutamate) or N (arginine, positively charged histidine) atoms. The bifurcation angle (γ) between the two vectors connecting the three

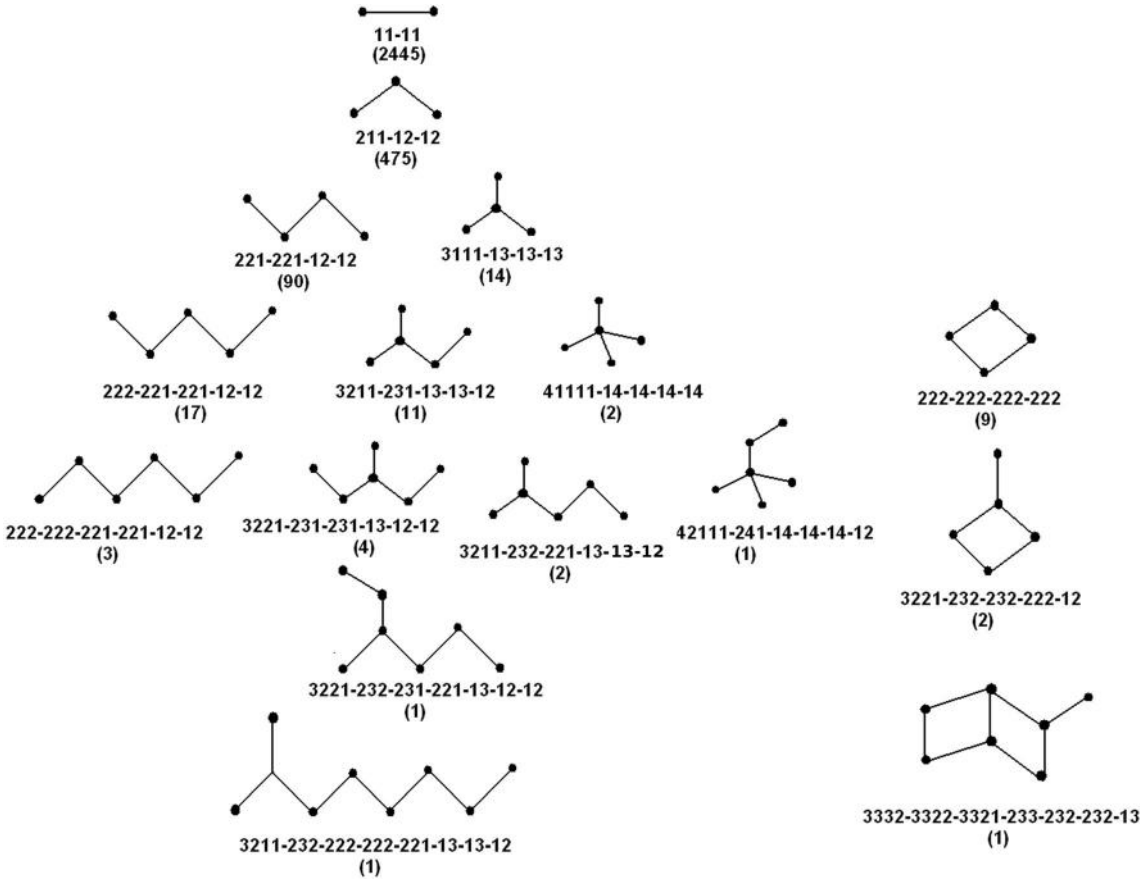

Fig. 3 Statistical distribution of networks of ionic bonds within proteins. Each network topology has a unique motif identifier (numerical string). The number of networks with each topology is given *in parentheses* below the identifier

charge-centers in the bridge was then computed; this angle was found to be obtuse and fairly well constrained (~110° ± 30°), irrespective of the residue composition of the bridge.

Contact order

The contact order (co) of the interacting pairs of ions involved in the salt bridge was calculated as the separation of the two amino acid residues in sequence space divided by the full length of the polypeptide chain. The majority of the interactions were found to be short-range contacts, with ~55% yielding a value of <0.1, which implies that these are ionic bonds between amino acid residues which are within 10% of each other in sequence space with respect to the whole length of the polypeptide chain. The overall distribution decays asymptotically for long-range contacts, as shown by its long tail and a median of 0.079 (Fig. 6). A significant preference for a co value of 0.03—accounting for 42% of the whole population of salt bridges—was observed, which is consistent with previous observations [8] that salt bridges are not only preferentially formed between residues that are close to rather than distant from each other in sequence space, but also at specific

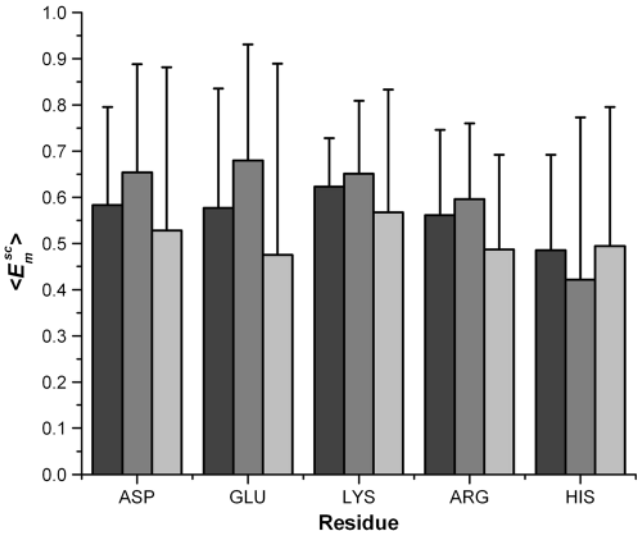

Fig. 4 The involvement of charged residues in salt bridges leads to only a mild enhancement in their E_m^{sc} values. The figure shows the mean E_m^{sc} values (*thick bars*) along with their standard deviations (*thin bars*) for charged residues involved in salt bridges (*mid-gray*), not involved in salt bridges (*light gray*), and pooled together (*dark gray*). Note that histidine (positively charged) shows the opposite trend upon inclusion in a salt bridge to the other residues

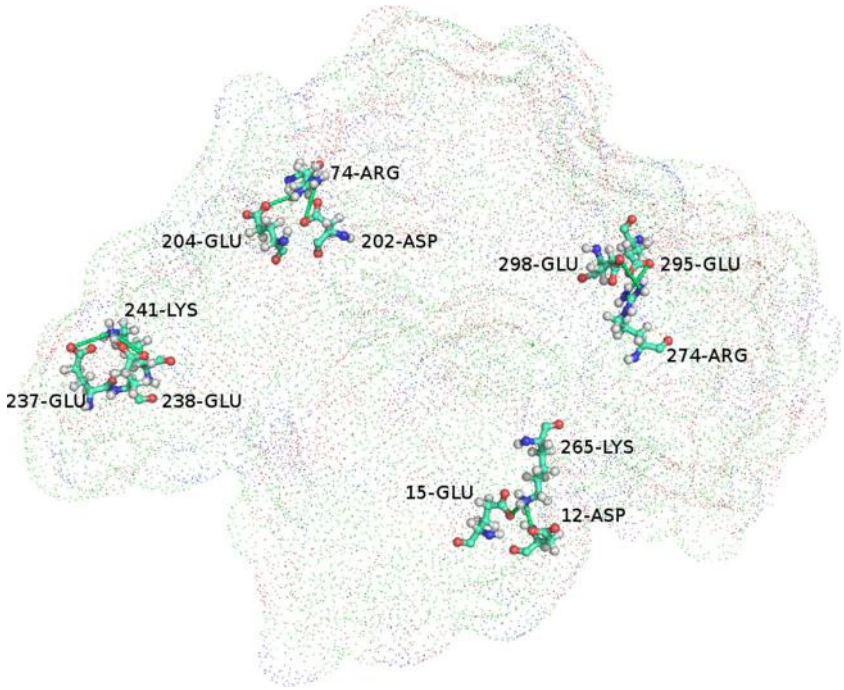

Fig. 5 Bifurcated salt bridges within proteins. All bifurcated salt bridges identified by the present algorithm in the protein 3B8X are shown. The salt bridges are displayed as *yellow sticks* linked to *red (Oxygens: -) or blue (Nitrogen: +) balls* (charged atoms) shown over the molecular surface of the protein (*dots*). The molecular surface was constructed using EDTsurf [47]. Most of the amino acids involved in the salt bridges are partially exposed to the solvent. The bridges shown here include both cationic (three of them) as well as anionic (one) centers and instances of both short-range (three of them) and long-range (one) contacts

small sequence separations. No major preference was observed for ionic bonds involved in bifurcated salt bridges compared to salt bridges in general.

Bifurcated salt bridges in protein–protein/protein–ligand interactions

There is a very fine line between folding and binding of proteins, and the difference between them is even smaller if viewed from a common conceptual platform based on complementarity. More specifically, it was demonstrated in a previous study [6] that folding can actually be envisaged as the docking of interior components onto the natively folded polypeptide chain. From that viewpoint, there is very little or even no difference between the salt bridges involved in the structural stability of a folded protein and those involved in protein–protein interactions (binding) in terms of their physicochemical attributes. However, we were interested in determining the role of bifurcated salt bridges in triggering and stabilizing interprotein associations. In other words, our study necessitated a thorough statistical analysis of bifurcated salt bridges in (i) protein–ligand and (ii) protein–protein interactions. The first calculation was performed in the globular protein database, while another database of native protein–protein complexes containing 1879 high-resolution crystal structures was assembled for the second calculation (see "Materials and methods").

For the first calculation, only proteins from the globular protein database that contained nonprotein (hetero) atoms apart from those in the crystallographic water molecules or buffer components were considered. The subset of structures that had one or more bifurcated salt bridges in contact with any of those heteroatoms were then identified. A maximum atomic separation of 4 Å between any heavy atom coming from the bifurcated salt bridge and any of the nonprotein

Table 1 Composition and geometry of the bifurcation angle (γ)

Composition	Count	Percentage	<γ>
GLU–ARG–GLU	81	17.05	120.6 (28.2)
ASP–ARG–GLU	52	10.95	121.1 (26.1)
GLU–ARG–ASP	45	9.50	110.8 (29.1)
ASP–ARG–ASP	43	9.05	112.3 (27.3)
ARG–GLU–ARG	35	7.37	97.4 (37.4)
ARG–ASP–ARG	28	5.89	92.9 (30.7)
GLU–LYS–GLU	26	5.47	116.0 (21.7)
LYS–GLU–ARG	26	5.47	102.1 (27.7)
ARG–GLU–LYS	26	5.47	100.4 (40.7)
ASP–LYS–ASP	26	5.47	93.6 (26.1)
LYS–GLU–LYS	22	4.63	113.8 (26.0)
GLU–LYS–ASP	18	3.79	109.0 (14.3)
ASP–LYS–GLU	13	2.74	115.3 (22.9)
ARG–ASP–LYS	13	2.74	93.8 (33.9)
LYS–ASP–ARG	8	1.68	95.4 (20.9)
LYS–ASP–LYS	7	1.47	76.8 (8.4)
LYS–ASP–HIS	2	0.42	126.1 (15.4)
GLU–HIS–GLU	1	0.21	165.4
ARG–GLU–HIS	1	0.21	70.9
GLU–HIS–ASP	1	0.21	70.1
ASP–HIS–ASP	1	0.21	169.6

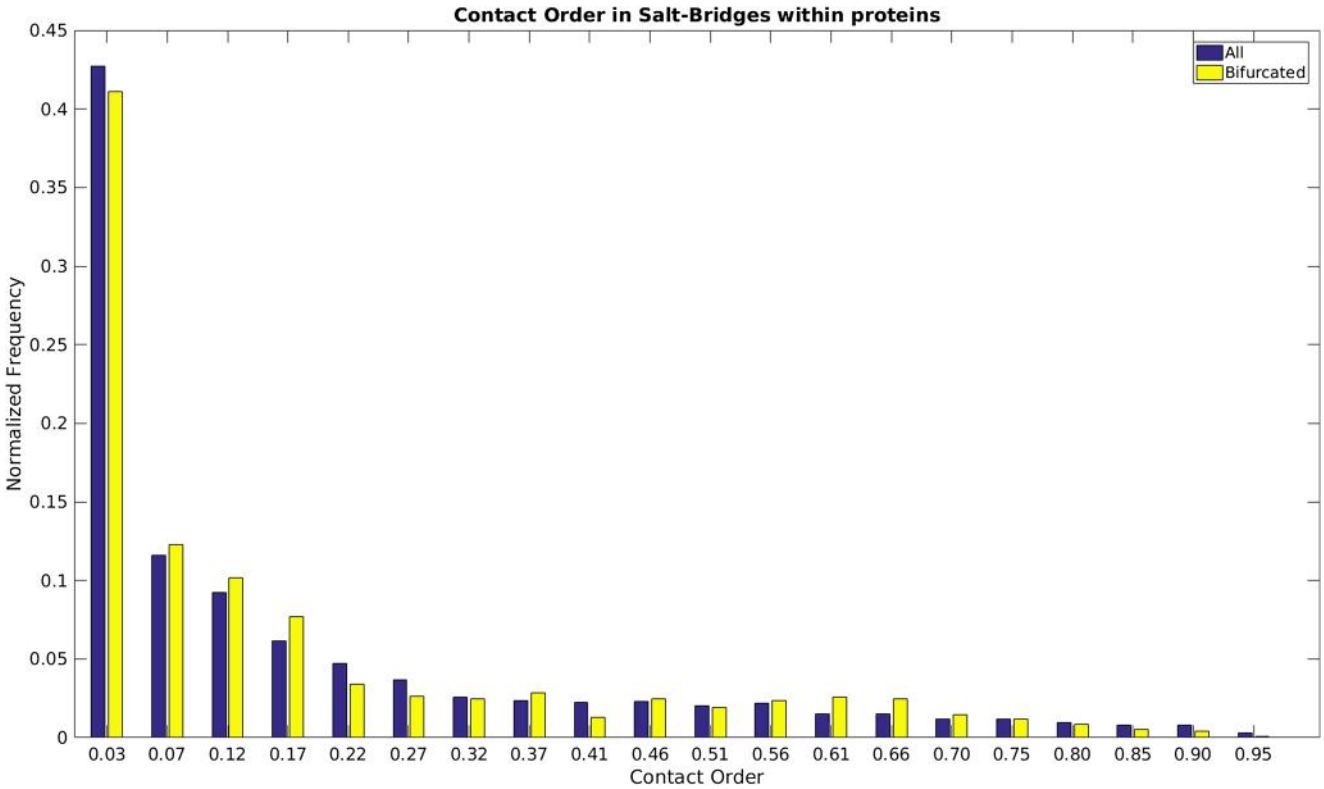

Fig. 6 Contact order in salt bridges within proteins. The bar diagram plots contact order values (see "Materials and methods") for inter-residue charged atomic contacts found in salt-bridge networks. As is evident from the plot, most contacts are close-range rather than distant contacts, and there was no particular difference in contact order between bifurcated salt bridges and salt bridges in general

atoms was imposed to filter the contacts. This stringent contact criterion was set to retain consistency with the adopted definition of salt bridges and to remove crystallographic artifacts from pseudo contacts. A total of 69 structures in which one or more bifurcated salt bridges was directly involved in ligand binding were obtained, which included 92 salt-bridge–ligand interactions (Table S1 in the "Electronic supplementary material," ESM). The ligands ranged from isolated metal ions to bulky prosthetic groups and from essential cofactors to coenzymes and enzyme substrates (Fig. 7). For the latter (coenzymes and enzyme substrates), the salt-bridge residues mapped to the active sites of the proteins.

An interesting example is the binding of isocitrate dehydrogenase (PDB ID: 1LWD) to its substrate isocitrate, wherein guanidium groups from the two arginines are involved in the formation of two ionic bonds with two carboxylic acid groups from the isocitrate (Fig. 7b). In effect, the topology of the full interactome changes to a closed quadrangle (222-222-222-222) from an open triple/angle (211-12-12: a bifurcated salt bridge) if the isocitrate is considered to be part of the extended protein–substrate salt-bridge network. Note that the bifurcated open triple requires just one of its members (the middle node in the network) to be involved in two simultaneous interactions, whereas the quadrangle requires every member to be involved in several simultaneous interactions in order to ensure that it is closed. The example is also notable because not only is the initial salt bridge bifurcated (or fork-like), but all the key interacting chemical groups are bifurcated too.

The second calculation involving protein–protein interactions appears to provide clearer evidence of the role played by bifurcated salt bridges in protein function. Indeed, some of the compositional and geometric features of protein–protein interfaces are quite odd and contrast markedly with those of protein interiors. For example, isolated nonpolar residues enclosed by polar or charged amino acids [31] are often found at protein–protein interfaces, in contrast to the hydrophobic clusters found regularly within protein interiors. These contrasting features suggested that it would be interesting to investigate differences in the distribution and functional specificity of salt bridges at the interface relative to the interior. Only salt bridges that involved both molecular partners (i.e., the two interacting polypeptide chains) were considered in this investigation; in other words, in all of the interfacial salt bridges extracted, both of the interacting polypeptide chains contributed at least one salt-bridge-forming residue. Again, besides isolated salt bridges, bifurcated salt bridges were found to be the most common type of salt bridge and to be prevalent at interfaces. Out of a total of 1879 native PPI complexes, 1088 contained one or more interfacial salt bridges, while 211 complexes were found to contain one or more bifurcated interfacial salt bridges (Table S2 in the ESM). There was a total of 283 bifurcated salt bridges (211-12-12), making them the second most common type after isolated salt bridges (11-11; 2022 in total).

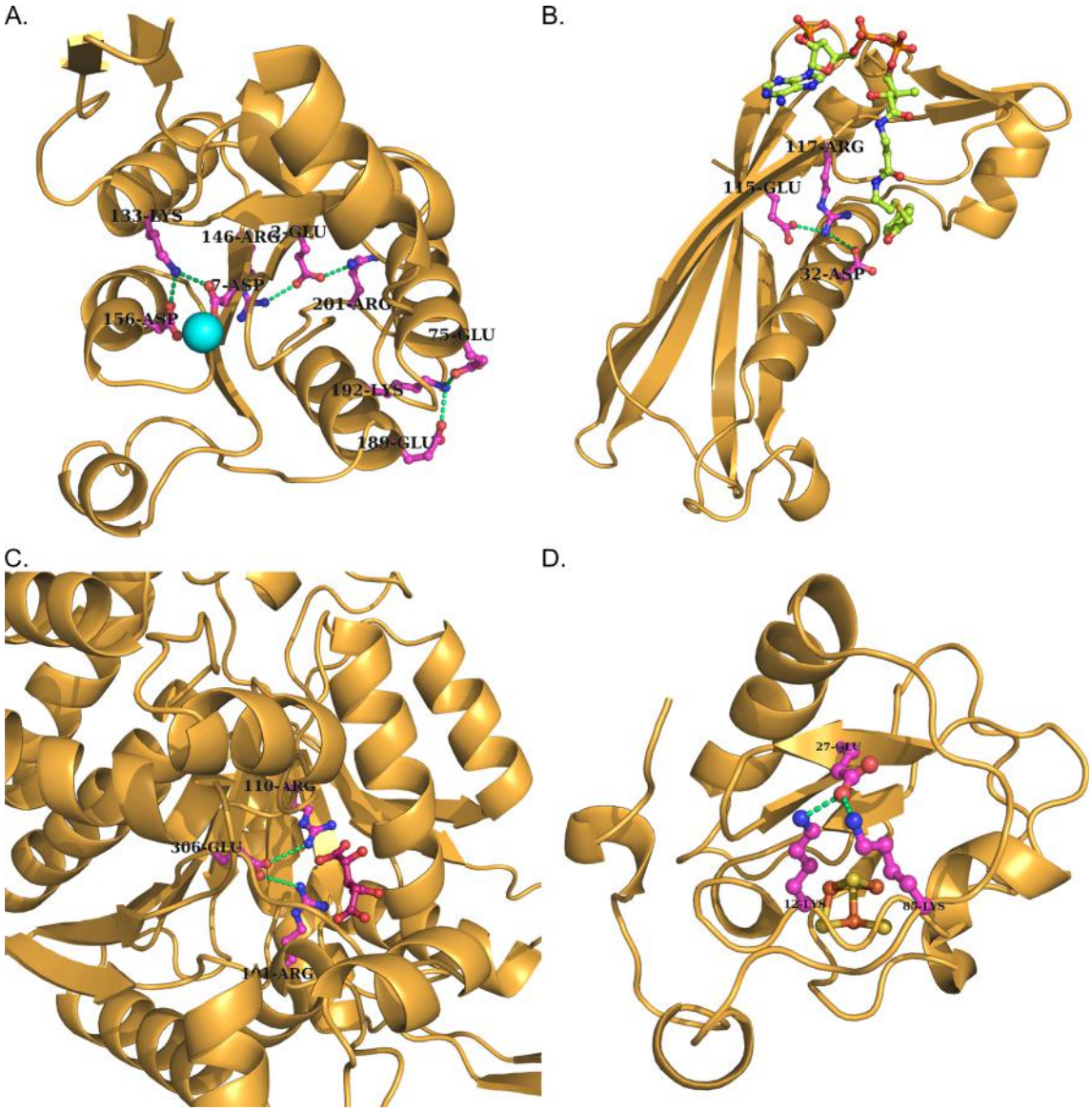

Fig. 7a–d Bifurcated salt bridges anchoring ligands within proteins. **a** Homoserine kinase from *Pseudomonas aeruginosa* (1RKU) has three bifurcated salt bridges, all of which are involved in the packing of secondary-structural elements (helices and sheets). Moreover, one of them (ASP7–LYS133–ASP156) also coordinates to a Mg^{2+} ion via the ε-NH_3^+ group of the lysine. **b** Isocitrate dehydrogenase (1LWD) from porcine heart mitochondria has a centrally located bifurcated salt bridge (ARG101–GLU306–ARG110) which binds the reaction substrate (isocitrate) via two ionic bonds of the guanidium groups of two arginines with two carboxylic acid groups from the isocitrate. This is a unique example of the formation of a closed quadrangular protein–substrate salt bridge from a bifurcated salt bridge. **c** A hydrolase from *Pseudomonas sp.* (1LO7) has a central bifurcated salt bridge (ASP32–ARG117–GLU115) which anchors to the substrate coenzyme/extended prosthetic group 4-hydroxybenzoyl CoA via an ionic bond mediated by an aspartate side chain. **d** Ferridoxin (1PC4) from *Azotobacter vinelandii* has a peripheral bifurcated salt bridge (LYS85–GLU27–LYS12) that guides the protein to its cofactor, the iron–sulfur cluster Fe_3S_3

In several instances, there were actually multiple salt bridges at the protein–protein interface, exhibiting various bridge separation distances. In some cases, for example, the interfacial salt bridges occurred in such close proximity (Fig. 8a) that they hinted at a potential cooperative effect promoting ioinic interactions at the interface; in other cases, the multiple salt bridges were distantly spaced and thus spanned a large contour of the interface (Fig. 8b), as if to cover more of the surface area buried upon complexation. The interfacial bifurcated salt bridges acted as molecular clips between the interacting protein surfaces.

Salt-bridge-mediated interactions between ordered and disordered regions in partially disordered proteins

Intrinsically disordered proteins have been studied intensively in protein science, especially over the last decade. A significant fraction of that research effort has been directed into

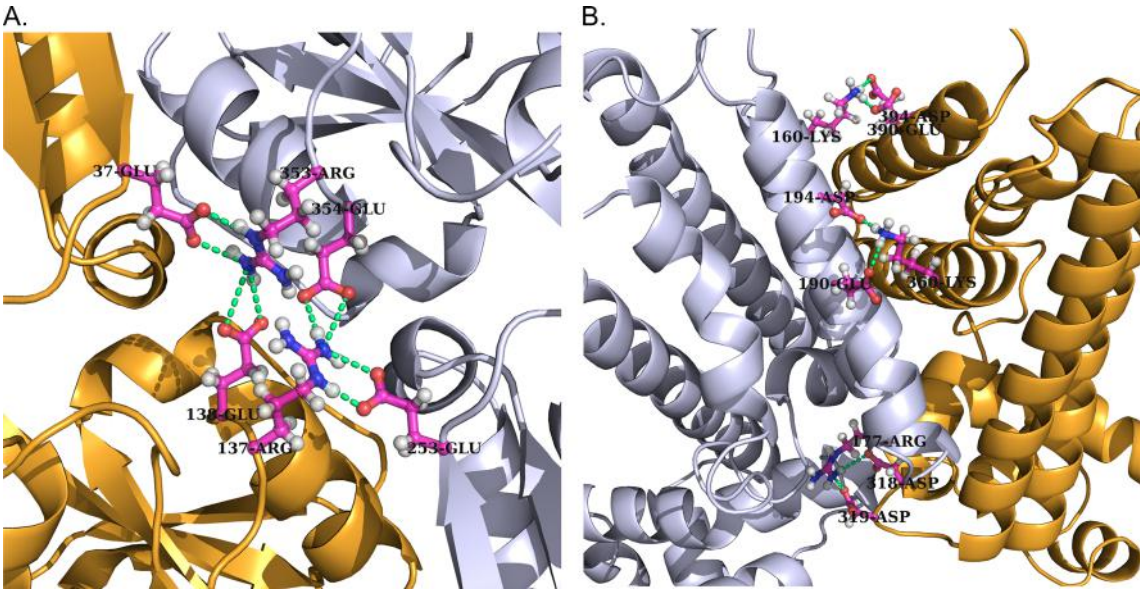

Fig. 8a–b Bifurcated salt bridges acting as molecular clips at protein–protein interfaces. **a** Multiple bifurcated salt bridges (GLU253-B–ARG137-A–GLU354-B; GLU138-A–ARG353-B–GLU37-A) that are centrally located at the interface and act cooperatively to clip together the interface at the effector binding domain of a Benm variant (2H9B). **b** Multiple distantly separated bifurcated salt bridges (GLU390-B–LYS160-A–ASP394-B; ASP319-B–ARG177-A–ASP318-B; ASP194-A–LYS360-B–GLU190-A) clip together the interface across a large surface area of a transcriptional regulator—a TetR family protein complex from *Cytophaga hutchinsonii* (3EUP)

deciphering a possible disorder code and unraveling the physical chemistry associated with the disorder-to-order transition that the "protean" residues undergo upon binding [18, 32, 33]. Machine learning algorithms have been extensively used to obtain sequence-based predictions of both the disordered and protean residues [34, 35]. It is well established that the disordered regions are rich in charged and polar amino acids in contrast to having a low hydrophobic content [36], which may provide mechanistic insights into the origin of their disorder. It has also been suggested that their low hydrophobic content disfavors self-folding [37] as it potentially decreases the number of possible two-body contacts [38].

Unlike in the nucleation–condensation model of the hydrophobic packing [26] and folding [39, 40] of globular proteins, the aforementioned features practically rule out the possibility that hydrophobic interactions playing an equivalent determining role in the partial folding of the IDPRs; instead, they emphasize the contributions of ionic bonds and atomic contacts mediated by charged and polar residues. To test this hypothesis, proteins containing >50% structural disorder [18] were assembled and the missing disordered regions were modeled using MODELER (see "Materials and methods"). In terms of the secondary-structural content (as calculated by Stride), there was an overwhelming preference for coils (or disordered loops) in these modeled structures, as expected. The fraction of coils ranged from 12.3% to 96%, with an average of 61.75% (± 18.17%); the average fraction of turns was 16.02% (± 10.33%).

It is obvious that the modeled structures represent only part of the conformational space comprising the whole plethora of possible conformational ensembles for the disordered regions. However, selecting the best structural representative would have involved time-averaged structures subjected to large-scale molecular dynamic simulations, which was outside the scope of the current study. That said, it was important to identify the extreme cases involving largely asymmetric protein models with completely unstructured long, wobbling loops that were present even in the best of the top ten models, and then to eliminate them from the atomic-contact-based calculation in order to remove statistical artifacts. The long, wobbling, disordered loops in these models occupied a huge physical volume and had practically no chance of participating in internal packing, as confirmed by visual inspection in Pymol (Fig. S1 in the ESM). Since the objective was to study the role of salt-bridge-mediated atomic contacts in the partial folding of these proteins, it was necessary to remove these outliers. The pseudo-centroid of the protein molecule was calculated from the set of experimentally solved atoms, and the distance of each completely exposed atom from this pseudo-centroid (see "Materials and methods") was calculated. The root mean square deviation of these distances (igA) was used as an indicator of globular asymmetry. Applying this parameter as an ad-hoc cutoff allowed largely asymmetric models to be discarded. The same parameter when calculated from the database of globular proteins (see "Materials and methods") gave a modest value of 4.92 (± 2.2), reflecting their approximately spherical shapes, apart from a few elongated proteins. In dramatic contrast, the average igA value obtained for the disordered proteins was 29.35 (± 30.34), with a maximum of 171.37, reflecting significant asymmetry and no sign of even

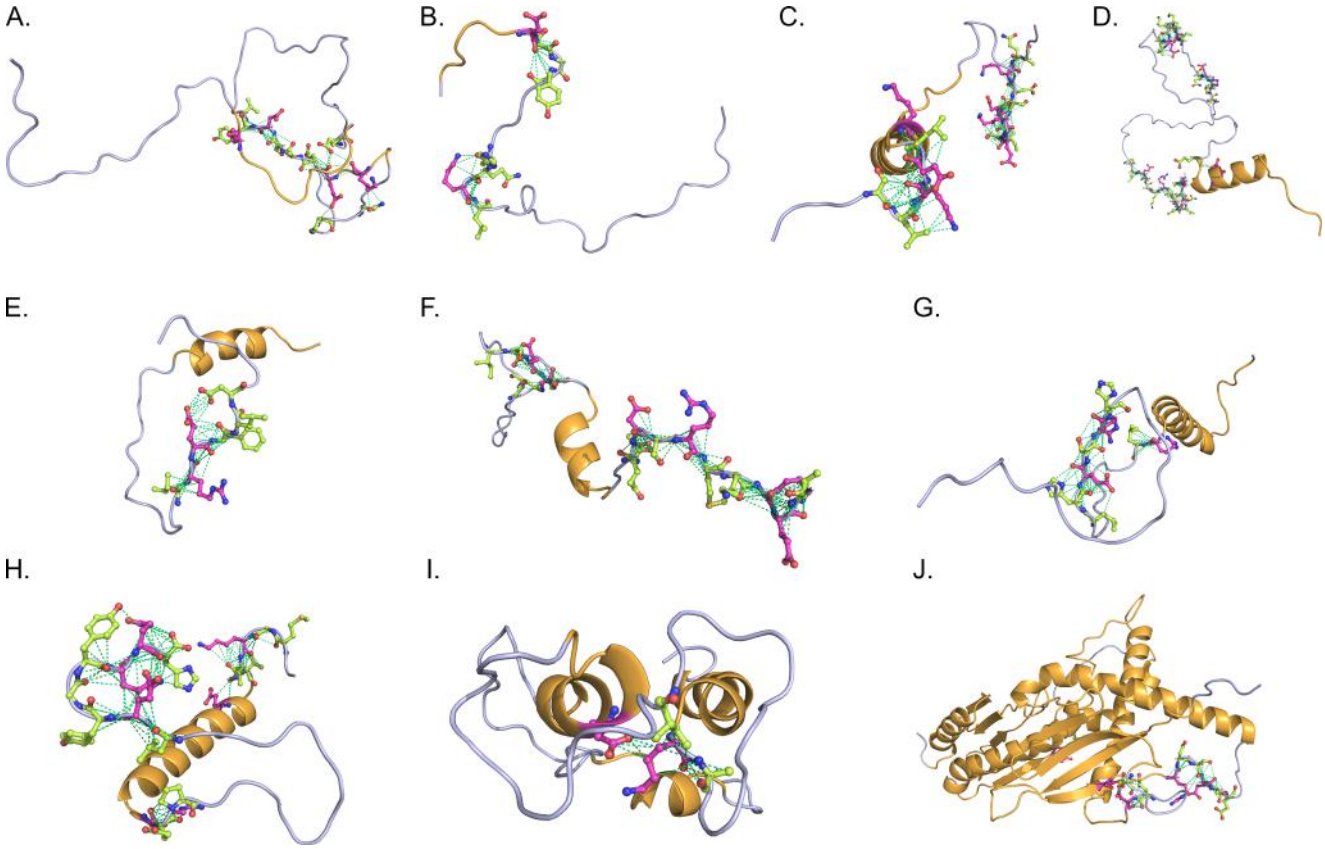

Fig. 9a–j Salt-bridge-mediated contacts in partially disordered proteins. The figure presents selected examples of salt-bridge-mediated interactions representing a variety of interaction modes between residues located in disordered loops (*light blue*) and salt-bridge residues located in structured regions (*bright orange*) in modeled structures containing IDPRs. The different modes of interaction are discussed in more detail in the main text. The PDB IDs are as follows: **a** 3KND, **b** 3IFN, **c** 3SJH, **d** 3MN7, **e** 2JZ3, **f** 4A1G, **g** 4G91, **h** 1H8B, **i** 1FH1, **j** 4H62

partial folding, given the volume constraints in a crowded living cell. We did not want the calculation to be contaminated by such unrealistic atomic models. However, setting an appropriate cutoff is still a largely empirical and context-dependent, as the largely unknown system of IDPs contrasts significantly with that of the globular protein system. Thus, it appeared necessary to perform visual inspection and to use an appropriate statistical rationale. One approach is to retain only models for which the average igA value is no more than $\mu + 3\sigma$ with respect to the values obtained from the globular proteins (μ: mean, σ: standard deviation), concomitantly allowing about a twofold increase in the associated standard deviation. A cutoff (in igA) of 20 Å gave rise to an average igA of 11.5 (± 4.1), which was found to be the closest match to the above filtering criterion. When applied, this reduced the number of structures to 62 (a recovery of ~ 57%). Note that even this cutoff is fairly relaxed considering that the objective was not to overlook any candidate that could give rise to sufficient disorder–order atomic contacts. This method (i.e., using a range of cutoffs for igA) was used previously [6] to successfully standardize and delineate the boundaries between different grid size bins and thereby effectively fix the box dimensions (cubic lattice) for globular proteins when performing continuum electrostatic calculations for those proteins by the Poisson–Boltzmann method using Delphi [30]. The validity of the filtering criterion was also cross-validated using the accessibility score (see "Materials and methods"), which was raised from 0.0018 (± 0.042) for the unfiltered set to 0.0033 (± 0.048) for the filtered set, leading to an approximately twofold increase in the average score, with standard deviations falling in the same range. Therefore, the filtered set of disordered protein models did appear to be better in terms of solution stability and hence more realistic.

To test whether salt-bridge-mediated interactions were influential in the partial folding of these IDPRs, salt bridges were identified via the same consistent criteria (see "Materials and methods") from each whole-protein-molecule in the aforementioned filtered set. Other atomic contacts were then identified between any pair of heavy atoms within 4 Å of each other, where one heavy atom was contributed by a residue located in the modeled disordered region and the other by a charged residue involved in a salt bridge irrespective of its location on the protein. A total of 180 salt bridges were identified, contributing to 4630 salt-bridge-mediated non-salt-bridge atomic contacts that led to 383 inter-residue interactions. Again, isolated ionic bonds were

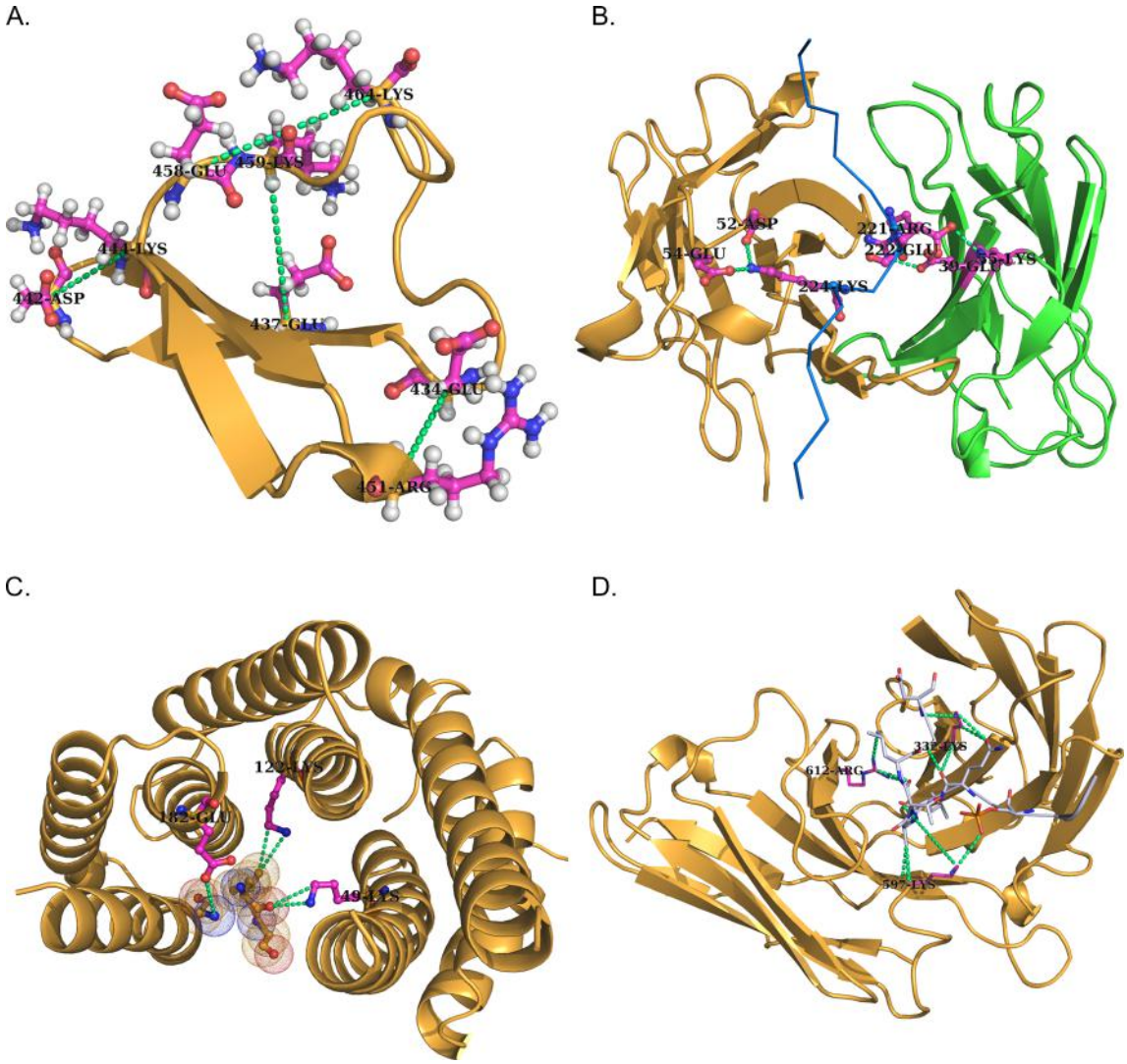

Fig. 10a–d Role of salt bridges and salt-bridge-mediated interactions in the formation of the tau fold and the stabilities of different tau peptides upon binding. **a** The NMR structure of the 202-amino-acid-long fragment of the tau protein in its free form (PDB ID: 2N4R). The two orthogonally oriented salt bridges together form a T-shaped substructure in the free form of tau which appears to be instrumental in maintaining the fold. **b–d** Different tau peptides in their bound forms (PDB ID: **b** 4TQE, **c** 5DMG, **d** 5BTV). In **b**, the *marine-colored zigzag line* (shown as a *ribbon* in Pymol) represents the most well studied 15 residue long tau peptide. Note that one bifurcated and two isolated salt bridges stitch the peptide to the Fab fragment of the anti-tau antibody (consisting of light and heavy chains). **c** and **d** present cases where other tau peptides (different to the one in **a**) are stabilized upon binding to their partners; this not only directly involves salt bridges but it also promotes salt-bridge-mediated interactions. A more detailed discussion of this is given in the main text

the most common contact type (11-11; 108 ionic bonds), followed by bifurcated salt bridges (211-12-12; 12 bridges). The salt-bridge-mediated contacts should, by definition, involve both polar (or charged) and nonpolar atoms from the main or side chains of the interacting residues. The distribution of contacts showed a large preference for main-chain atoms, which constituted ~80% of the atoms involved in contacts (main-chain to main-chain contacts: 2962, main-chain to side-chain contacts: 1441, side-chain to side-chain contacts: 227). This is most likely due to the fact that the (few) charged side-chain atoms (in $-NH^+$ and $-COO^-$ groups) are already occupied in salt bridges and are therefore less available to form other contacts compared to the main-chain atoms. The contacts mediated by salt-bridge-forming charged residues do indeed appear to be subservient in the partial folding of these quasi-stable proteins. Several prototype interaction modes were identified (Fig. 9), ranging from loop bridging (Fig. 9a) to loop bending and turning (Fig. 9b–d, g), helix-loop clipping (Fig. 9c, e–h), partial stabilization of loops through the promotion of a series of contacts (Fig. 9c, f, g), and the linking of helices via mediating contacts in helix-loop-helix and other related secondary-structural motifs (Fig. 9i–j). The results should guide the further exploration of the modus operandi of these partially disordered proteins in future studies.

Role of salt-bridge-mediated interactions in the stabilization of disordered protein regions upon binding

We also had a close look at a few of the most popular IDPs/IDPRs for which experimental structures (of either their free or bound forms) are available. The aim was to investigate whether their structural stabilization in the bound form or their lack of stabilization in the free form had any influence on the salt-bridge-mediated interactions. Due to their intrinsic disorder, IDPs generally do not form stable and diffractable crystals in their free forms. Even in their bound forms, only fragments of the whole protein are available. These fragments representing the disorder-to-order transitioning residues upon binding are termed "protean" (mutable) fragments, and a growing field of computational tools for sequence-based prediction of these disorder-to-order transitioning binding regions are being developed [34, 35].

Case studies of tau

One of the most popular IDPs is undoubtedly the tau protein, which is one of the primary causes of the neurofibrillary degeneration that occurs in Alzheimer's disease. Apart from a solution (NMR) structure of its free form (2N4R), a few crystal structures are available for this protein when it is bound to different cellular partners and diagnostic markers. Most of these, however, cover only a small fraction (3–15 residues) of the whole protein. Among the available complexes, the current analyses opted for 4TQE, 5DMG, and 5BTV. In two of the complexes, different peptide fragments of the tau protein (tau peptides) are bound to anti-tau antibodies; in the third complex, a different peptide fragment of the same protein is bound to the epithelial cell marker protein 1 (stratifin), forming the sigma complex. The study also included the NMR structure (the first and the lowest-energy model) of the standalone molecule in the analyses. Detailed structural investigation from the perspective of salt-bridge-mediated stabilization revealed some really interesting features.

In its free form (i.e., the 202-residue-long fragment), tau is partially disordered: it is roughly 50% β-sheet and ~50% coils, as revealed by NMR. Its structure was found to have four isolated salt bridges (11-11), all of which play pivotal roles in its folding as they are either involved in loop- closure or in linking together distant secondary-structural elements in space (Fig. 10a). Two of the salt bridges were located at opposite ends of a horizontal pseudo-axis and stabilized (intra- and interstrand) turns, whereas the other two salt bridges were closely spaced, aligned orthogonally to one another, and formed a T-shaped pair. The relative orientation of the stem of the "T" was vertical with respect to the aforementioned horizontal pseudo-axis, meaning that the stem provided mechanical support to the overall structure. In effect, the bar of the "T" resembles a flyover.

In the first of the complexes, namely 4TQE, a lysine (K224-Tau) situated centrally in the thread-like 15-amino-acid-long tau peptide is stitched meticulously by two negatively charged residues (D52-H, E54-H) from the heavy chain (H) of the anti-tau antibody to form an interfacial bifurcated salt bridge (Fig. 10b). In addition, there are two more isolated salt bridges (K55-L–E222-Tau, E39-L–R221-Tau) at the interface which are contributed by the light chain (L) and stitch the disordered peptide thread even further along the deep groove of the antibody's Fab fragment. It is important to note that the charged residues involved in the salt bridges are closely spaced and located centrally on the peptide, thereby enhancing the strength and stability of the binding. Thus, the role of ionic bond stabilization in this disorder-to-order transition cannot be overlooked.

In the other two examples (5BTV and 5DMG), charged residues appear to be crucial to the binding and stabilization of the peptides. It is important to note that all three tau peptides studied here were quite different. Ionic bonds as well as noncovalent N···C contacts involving charged residues from both the tau peptide and its molecular partner were found to be crucial to its stabilization (Fig. 10c–d). The intermolecular ionic bonds, in effect, formed extended networks of conjugated salt bridges.

Case study of crambin

Another case study was performed, focusing on the plant protein crambin (from *Crambe hispanica*). The structure of crambin was solved at an ultrahigh resolution of 0.54 Å (PDB ID: 1EJG). This structure presents an unique case of a protein molecule having multiple-occupancy atoms throughout the whole chain, with each of its isoform being individually disordered (http://deposit.rcsb.org/format-faq-v1.html). This protein was later solved at an even higher resolution (PDB ID: 3NIR, 0.48 Å), and both of these ultrahigh-resolution structures proved useful in a study of protein helices reconciled with three-centered hydrogen bonds [41]. Interestingly, the whole protein (1EJG) gave rise to just a single isolated salt bridge (ARG17–GLU23), which connected the largest helix to a long irregular loop. Most remarkably, the two charged residues (ARG17, GLU23) involved in this single isolated salt bridge were also involved in another 113 atomic contacts with 21 other residues, such that this salt bridge was critical to the integrity of the entire protein fold (Fig. 11a). These atomic contacts were calculated using the same cutoff of 4.0 Å used consistently throughout the whole study. In other words, if any heavy atom from the salt-bridge-forming residues was found to be within 4.0 Å of any other heavy atom from any other part of the protein, those heavy atoms were considered to be in contact. The atomic contacts

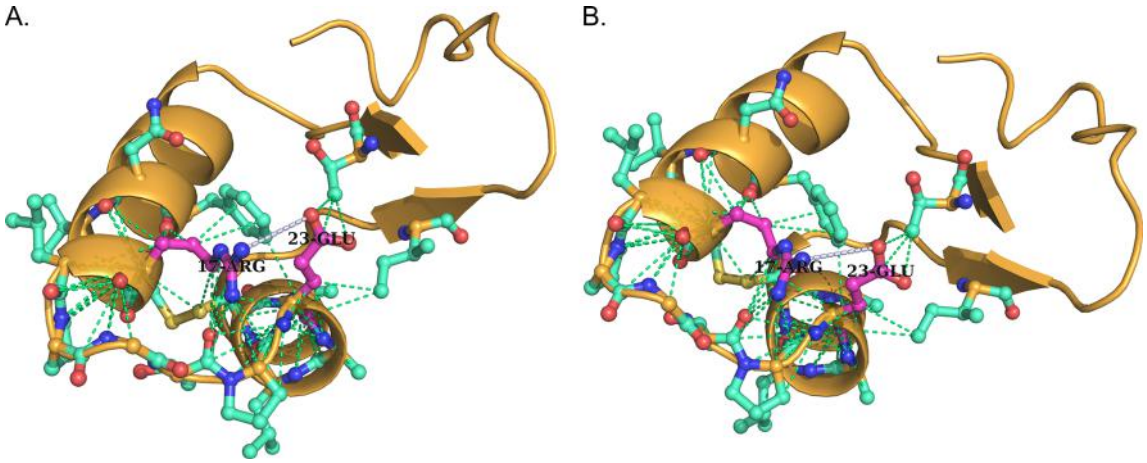

Fig. 11a–b Structural integration and stabilization of crambin, involving salt-bridge-mediated interactions. **a** and **b** present two ultrahigh-resolution structures of the protein crambin (PDB ID: 1EJG and 3NIR, respectively), solved at slightly different resolutions but showing identical results. Both structures were used in the study in order to cross-validate each other. As elaborated in detail in the main text, the structures represent a unique case of a single salt bridge (ARG17–GLU23) involved in more than 100 atomic contacts with over 20 other residues, such that this salt bridge is critical to the structural integrity of the entire protein fold

were evenly distributed among nonpolar (C: 56) and polar (O, N, S: 57) atoms. The salt bridge's role in keeping the fold together was especially prominent at the more structured half of the fold containing the two helices. Therefore, this "isolated" salt bridge does not appear to be isolated in a broader structural sense, given that it is crucial to maintaining the entire fold. The results obtained from the structure 1EJR were cross-validated with those from the higher-resolution structure, 3NIR (Fig. 11b).

Identifying trends towards designable salt-bridge interactions

Protein design can still be considered an emerging field that addresses the different structure–function attributes of native and alternative folds of proteins, and thus involves both fundamental and applied research. It ranges from explorations of alternating packing modes in protein interiors (i.e., hydrophobic core design [42]) to attempts to decipher the inverse

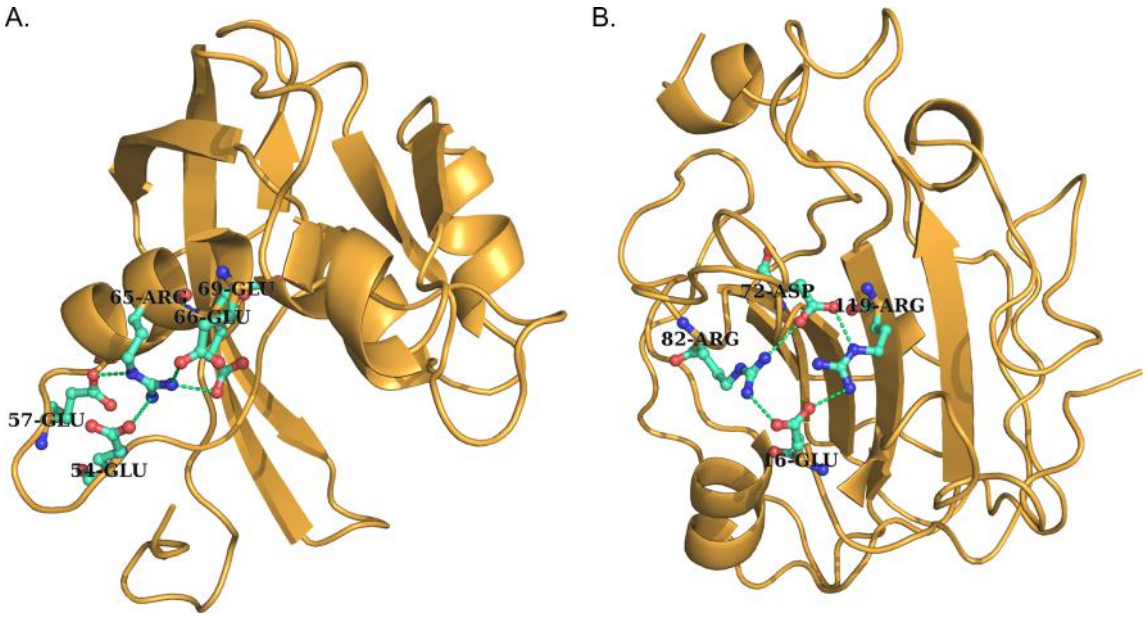

Fig. 12a–b Composite salt-bridge networks stabilizing secondary-structural associations within globular proteins. The composite figure shows two classic examples of the intermotif association of secondary structural elements stabilized by composite salt-bridge networks. **a** Helix–loop stabilization mediated by a five-membered branched tree, namely the 4-star (motif identifier: 41111-14-14-14-14) found in diphosphoinositol polyphosphate phosphohydrolase 1 of human (PDB ID: 2FVV). **b** Promoting the closed association of a short helix with two loops and a strand in a closed ring topology, namely a quadrangle or 4-cycle (motif identifier: 222-222-222-222), as found in rat phosphatidylethanolamine-binding protein (PDB ID: 2IQY)

protein folding problem (full sequence design [43]), searches for structurally distinct evolutionary protein variants [44], increasing the thermal stability of proteins (for example by incorporating disulfide bridges [45] and triplet cliques [26]), the introduction of new functions and catalytic activities into existing proteins (by incorporating metal-coordinating motifs, salt bridges [8], and other designed polar clusters) [44], and the identification of trends in protein evolution from intrinsically disordered ancestors [46]. Among the various topics addressed in this field, the design of salt bridges is definitely one of the toughest problems to tackle, particularly due to the difficulty involved in handling the side-chain conformational space for elongated charged residues (Lys, Arg, Glu) [8] with a wide spectrum of possible allowed side-chain rotamers. In other words, while the residues involved in a deliberately incorporated salt bridge in a designed protein model may approach each other in a certain mode (or as a particular combination of rotamers) *in silico*, that does not guarantee that those rotamers will indeed be selected in the real folded protein. Some other combination of rotamers in the actual designed protein may not favor the formation of the salt bridge designed *in silico*. One plausible way to handle this issue may be an idea borrowed from the related problem of hydrophobic core design; specifically, alternatively packed cores within protein interiors. In alternatively packed hydrophobic cores, different combinations of hydrophobic amino acid residues with different rotamers can satisfy certain global packing constraints, thus leading to the same fold [42]. In network terminology, these global packing constraints may relate to clustering, reach (path lengths) [26], partitioning, and similar global network properties. Note that a specific three-dimensional fold is a collective feature that is not critically dependent on the specifics of each individual interaction that contributes to the fold. This could suggest a design strategy for salt bridges that aims to incorporate new functions into existing proteins, given the fact that we now have characterized the basic motifs and rules of thumb for the formation of different composite salt-bridge networks (Fig. 3). The same principles of hydrophobic core design may also be applicable to the design of composite salt-bridge networks: a combination of rotamers selected for a pair of charged residues may be replaced with a different pair of rotamers that offer a greater opportunity for the rescue of the ensemble of ionic interactions in the designed protein. In other words, this would potentially reduce the risk involved in the design protocol. A closer look at the different network topologies (Fig.3) identified for salt bridges by mapping their abstract graph representations to the real three-dimensional structural context demonstrated their involvement in intermotif interactions (Fig. 12) between a collection of secondary-structural elements (helices, strands, and loops) that are crucial to fold integrity. Molecular dynamics simulation studies have further elucidated the role of the configurational entropy of salt-bridge networks in the thermostability of peptides [9]; the current study complements those by elucidating salt-bridge motifs, encouraging further research along that direction.

Conclusions

The study reported in the present paper analyzed the variety and extent of different modes of association (networks) for salt bridges on or within folded proteins and at protein–protein interfaces. The bifurcated salt bridge appears to be a special and prevalent salt-bridge motif. Apart from serving as a major component of the meticulous electrostatic balance of a folded protein, detailed functional characterization of salt bridges shows that they are key participants in the binding of ligands, cofactors, metal ions, and prosthetic groups, and that they facilitate a variety of protein–protein interactions. One very interesting functional motif involving bifurcated interfacial salt bridges could be envisaged as molecular clips; these clips are involved in stitching together large surface contours at interacting protein–protein interfaces. Salt-bridge-mediated interactions seem to play a pivotal role in the partial folding of proteins containing large numbers of disordered regions. The results obtained in this work should aid the conceptualization of how these proteins manage to retain the necessary amount of disorder, even in their functionally active bound forms, and it supports the proposed notion of geometrically specific designable interactions involving salt bridges.

Acknowledgements We express our heartfelt gratitude to Prof. Rahul Banerjee, Saha Institute of Nuclear Physics, Kolkata, India, for his uninterrupted motivations. We also express our sincere thanks to Prof. Anjan Kumar Dasgupta (Department of Biochemistry, University of Calcutta), for his constant support and encouragement of the work. The work was supported by the Department of Science and Technology—Science and Engineering Research Board (DST-SERB research grant PDF/2015/001079/LS).

References

1. Janin J, Chothia C (1990) The structure of protein–protein recognition sites. J. Biol. Chem. 265:16027–16030
2. Jones S, Thornton JM (1995) Protein–protein interactions: a review of protein dimer structures. Prog. Biophys. Mol. Biol. 63:31–65
3. McCoy AJ, Chandana Epa V, Colman PM (1997) Electrostatic complementarity at protein/protein interfaces. J. Mol. Biol. 268: 570–584. doi:10.1006/jmbi.1997.0987
4. Gilson MK, Honig B (1988) Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins 4:7–18. doi:10.1002/prot.340040104
5. Li C, Li L, Petukh M, Alexov E (2013) Progress in developing Poisson–Boltzmann equation solvers. Mol Based Math Biol 1:42–62. doi:10.2478/mlbmb-2013-0002
6. Basu S, Bhattacharyya D, Banerjee R (2012) Self-complementarity within proteins: bridging the gap between binding and folding. Biophys. J. 102:2605–2614. doi:10.1016/j.bpj.2012.04.029

7. Musafia B, Buchner V, Arad D (1995) Complex salt bridges in proteins: statistical analysis of structure and function. J. Mol. Biol. 254:761–770. doi:10.1006/jmbi.1995.0653
8. Donald JE, Kulp DW, DeGrado WF (2011) Salt bridges: geometrically specific, designable interactions. Proteins 79:898–915. doi:10.1002/prot.22927
9. Missimer JH, Steinmetz MO, Baron R, et al (2007) Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. Protein Sci Publ Protein Soc 16:1349–1359. doi:10.1110/ps.062542907
10. Walker KD, Causgrove TP (2009) Contribution of arginine-glutamate salt bridges to helix stability. J. Mol. Model. 15:1213–1219. doi:10.1007/s00894-009-0482-5
11. Honig B, Yang AS (1995) Free energy balance in protein folding. Adv. Protein Chem. 46:27–58
12. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280:1–9. doi:10.1006/jmbi.1998.1843
13. Torshin IY, Weber IT, Harrison RW (2002) Geometric criteria of hydrogen bonds in proteins and identification of 'bifurcated' hydrogen bonds. Protein Eng. 15:359–363. doi:10.1093/protein/15.5.359
14. Di Primo C, Deprez E, Sligar SG, Hui Bon Hoa G (1997) Origin of the photoacoustic signal in cytochrome P-450 $_{cam}$: role of the Arg186−Asp251−Lys178 bifurcated salt bridge. Biochemistry (Mosc) 36:112–118. doi:10.1021/bi961508a
15. Basu S, Bhattacharyya D, Banerjee R (2014) Applications of complementarity plot in error detection and structure validation of proteins. Indian J. Biochem. Biophys. 51:188–200
16. Basu S, Bhattacharyya D, Wallner B (2014) SARAMAint: the complementarity plot for protein–protein interface. J Bioinforma. Intell. Control. 3:309–314. doi:10.1166/jbic.2014.1103
17. Basu S, Wallner B (2016) Finding correct protein–protein docking models using ProQDock. Bioinformatics 32:i262–i270. doi:10.1093/bioinformatics/btw257
18. Baruah A, Rani P, Biswas P (2015) Conformational entropy of intrinsically disordered proteins from amino acid triads. Sci. Rep. 5:11740. doi:10.1038/srep11740
19. Wang G, Dunbrack RL (2003) PISCES: a protein sequence culling server. Bioinformatics 19:1589–1591. doi:10.1093/bioinformatics/btg224
20. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics 5:5.6. doi: 10.1002/0471250953.bi0506s15
21. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/. Accessed 26 Feb 2017
22. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23(4):566–579. https://www.ncbi.nlm.nih.gov/pubmed/8749853. Accessed 28 Feb 2017
23. Hubbard S, Thornton J (1993) NACCESS. Department of Biochemistry and Molecular Biology, University College, London. http://www.oalib.com/references/5299711. Accessed 1 Mar 2017
24. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. 55:379–400
25. Banerjee R, Sen M, Bhattacharya D, Saha P (2003) The jigsaw puzzle model: search for conformational specificity in protein interiors. J. Mol. Biol. 333:211–226
26. Basu S, Bhattacharyya D, Banerjee R (2011) Mapping the distribution of packing topologies within protein interiors shows predominant preference for specific packing motifs. BMC Bioinformatics 12:195. doi:10.1186/1471-2105-12-195
27. Harary F (1969) Graph theory. Addison–Wesley, Reading
28. Cornell WD, Cieplak P, Bayly CI, et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117:5179–5197. doi:10.1021/ja00124a002
29. Shannon RD (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. Acta Crystallogr. A 32:751–767. doi:10.1107/S0567739476001551
30. Li L, Li C, Sarkar S, et al (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys. 5:9. doi:10.1186/2046-1682-5-9
31. Jones S, Thornton JM (1996) Principles of protein–protein interactions. Proc. Natl. Acad. Sci. U.S.A. 93:13–20. http://www.pnas.org/content/93/1/13. Accessed 20 Jan 2017
32. Theillet F-X, Kalmar L, Tompa P, et al (2013) The alphabet of intrinsic disorder. Intrinsically Disord. Proteins 1:e24360. doi:10.4161/idp.24360
33. Fukuchi S, Sakamoto S, Nobe Y, et al (2012) IDEAL: Intrinsically Disordered proteins with Extensive Annotations and Literature. Nucleic Acids Res. 40:D507-511 doi: 10.1093/nar/gkr884
34. Jones DT, Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31(6):857–863. http://bioinformatics.oxfordjournals.org/content/31/6/857.full. Accessed 28 Apr 2016
35. Basu S, Söderquist F, Wallner B (2017) Proteus: a random forest classifier to predict disorder-to-order transitioning binding regions in intrinsically disordered proteins. J. Comput. Aided Mol. Des. 1–14 doi:10.1007/s10822-017-0020-y
36. Mao AH, Crick SL, Vitalis A, et al (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U. S. A. 107:8183–8188. doi:10.1073/pnas.0911107107
37. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41:415–427
38. Schlessinger A, Punta M, Rost B (2007) Natively unstructured regions in proteins identified from contact predictions. Bioinforma. Oxf. Engl. 23:2376–2384. doi:10.1093/bioinformatics/btm349
39. Fersht AR (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc. Natl. Acad. Sci. U. S. A. 92:10869–10873
40. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J. Mol. Biol. 254:260–288. doi:10.1006/jmbi.1995.0616
41. Kuster DJ, Liu C, Fang Z, et al (2015) High-resolution crystal structures of protein helices reconciled with three-centered hydrogen bonds and multipole electrostatics. PLoS One 10:e0123146. doi:10.1371/journal.pone.0123146
42. Munson M, Balasubramanian S, Fleming KG, et al (1996) What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Protein Sci. Publ. Protein Soc. 5:1584–1593

43. Yue K, Dill KA (1992) Inverse protein folding problem: designing polymer sequences. Proc. Natl. Acad. Sci. U. S. A. 89:4163–4167
44. Jacobs T, Williams B, Williams T, et al (2016) Design of structurally distinct proteins using strategies inspired by evolution. Science 352:687–690. doi:10.1126/science.aad8036
45. Betz SF (1993) Disulfide bonds and the stability of globular proteins. Protein Sci. Publ. Protein Soc. 2:1551–1558
46. Zhu H, Sepulveda E, Hartmann MD, et al (2016) Origin of a folded repeat protein from an intrinsically disordered ancestor. eLife 5: e16761. doi:10.7554/eLife.16761
47. Xu D, Zhang Y (2009) Generating triangulated macromolecular surfaces by Euclidean distance transform. PLoS One 4:e8140. doi:10.1371/journal.pone.0008140