Salt-bridge Dynamics in Intrinsically Disordered Proteins: A trade-off between electrostatic interactions and structural flexibility

Sankar Basu[1], Parbati Biswas[1]*
[1]Department of Chemistry, University of Delhi
Viswavidyalay Marg
North Campus
Delhi 110007

* Corresponding Author

E-mail:
SB: nemo8130@gmail.com
PB: pbiswas@chemistry.du.ac.in

Abstract

Intrinsically Disordered Proteins (IDPs) are enriched in charged and polar residues; and, therefore, electrostatic interactions play a predominant role in their dynamics. In order to remain multi-functional and exhibit their characteristic binding promiscuity, they need to retain considerable dynamic flexibility. At the same time, they also need to accommodate a large number of oppositely charged residues, which eventually lead to the formation of salt-bridges, imparting local rigidity. The formation of salt-bridges therefore oppose the desired dynamic flexibility. Hence, there appears to be a meticulous trade-off between the two mechanisms which the current study attempts to unravel. With this objective, we identify and analyze salt-bridges, both as isolated as well as composite ionic bond motifs, in the molecular dynamic trajectories of a set of appropriately chosen IDPs. Dynamic structural properties of these salt-bridges like persistence, time evolution of the secondary structural 'order-disorder' transitions, correlated atomic movements, contribution in the overall electrostatic balance of the proteins and other essential features have been studied in necessary detail. The results suggest that the key to maintain such a trade-off over time is the continuous formation and dissolution of salt-bridges with a wide range of persistence. Also, the continuous dynamic interchange of charged-atom-pairs (coming from a variety of oppositely charged side-chains) in the transient ionic bonds supports a model of dynamic flexibility concomitant with the well characterized stochastic conformational switching in these proteins. Furthermore, the analyses of hydrophobic burial profiles present a simple and effective single-parameter-tool (namely, the rGb score) to characterize the instability of the IDPs in solution – which serves as the basis of their high reactivity. The results and conclusions should facilitate the future design of salt-bridges as a mean to further explore the disordered-globular interface in proteins.

Keywords: Intrinsically Disordered Proteins, Salt-bridges, Ionic Bonds, Electrostatic Complementary, Dynamic Flexibility

Introduction

Recent research in Intrinsically Disordered Proteins (IDPs) has brought about a paradigm shift in the central dogma of protein folding, specifically questioning the 'one sequence → one structure → one function' epitome in globular proteins [1]. IDPs lack unique three-dimensional structures under physiological conditions, but are known to carry out a multitude of important biological functions. The intrinsic disorder of a protein may be classified as i) Intrinsically Disordered Protein Regions (IDPRs) where well-defined secondary structures coexist with highly flexible disordered regions and ii) Intrinsically Disordered Proteins (IDPs) which are completely devoid of any distinct tertiary structure. The functional diversity of IDPs/IDPRs may be attributed to their potential promiscuity in binding to different structurally unrelated partners in living cells [2]. IDPs and IDPRs are usually represented as dynamic ensembles of interconvertible conformations [3] instead of well-defined structures that characterizes globular proteins. The structural flexibility of IDPs/IDPRs [4] in uncomplexed form imparts higher conformational entropy [5] due to continuous fluctuations between multiple conformers. Compared to globular proteins, IDPs and IDPRs are highly enriched in charged and polar residues and deficient in hydrophobic residues, which is denoted by a higher mean net charge to mean hydrophobicity ratio [2, 6, 7]. It has been observed in protamines (a class of arginine-rich IDPs with defined cellular functions) that the net charge per residue is one of the discriminating 'order' parameters responsible for the adoption of heterogeneous conformational ensembles in IDPs/IDPRs during the globule-to-coil transition [8]. Also, in contrast to the globular proteins, the hydrophobic residues are randomly scattered along the disordered sequences [4, 9] that do not mediate a hydrophobic collapse [10, 11] accompanied by concomitant water depletion.

Theoretical studies based on designed globular and disordered protein sequences have revealed that there is a transition of the prevalent interactions at the globular-disorder interface across the charge–hydrophobicity boundary [12]. To this end, the globular-disorder (or globule-to-coil) transition may be explained based on a compromise between hydrophobic and electrostatic interactions. The prevalence of polar and charged residues in the disordered proteins potentially trigger electrostatic interactions, both attractive and repulsive, which constitutes the predominant interaction in IDPs/IDPRs [13]. The role of electrostatic interactions are manifest in the binding of IDPs/IDPRs to their targets and in the binding of histone chaperones, [14] which are investigated by combined NMR and Molecular Dynamics (MD) studies [15]. One of the key cellular functions seldom observed in IDPs is their phosphorylation, initiating ultra-sensitive reactions, where the electrostatic interactions play a pivotal role, contributing to ultra-sensitivity [16]. Again, one of the major electrostatic components are the salt bridges, that comprise of ionic bonds between oppositely charged amino acids in close contact. Salt-bridges, both in isolation and as composite ionic bond motifs [17], are known to impart local rigidity in protein structures [18] (analogous to the disulphide bridges [19, 20]) by bringing together distant regions in folded globular proteins, acting as molecular clips at the protein-protein interface and by promoting loop-closures as revealed in computationally modeled partially disordered proteins [17]. However, the structural plasticity of IDPs/IDPRs, rich in prolines and glycines gives rise to high coil forming propensities [4]. Hence, it would be logical to explore the dynamics of salt-bridge formation in partially/completely disordered proteins for the understanding of how IDPRs/IDPs retain structural flexibility at the expense of a large number of charged residues.

The main purpose of this study is to understand the meticulous trade-off between these two seemingly opposite phenomena which may facilitate the design of partially/completely disordered proteins. With this background, we explore the structure and dynamics of these salt-bridges, both individually and also as composite motifs in IDPs/IDPRs along their simulated MD trajectories. We analyze their dynamic persistence, their involvement in mediating secondary structural 'order-

disorder' transitions and other associated structural features in an attempt to rationalize their paradoxical presence in good numbers in the dynamic conformational ensembles of IDPs/IDPRs. We also investigate in detail, the specific contribution of these salt-bridges in depicting the overall electrostatic interaction map of these proteins, given the importance of electrostatic interactions in IDPs/IDPRs. Furthermore, we also attempt to explain the solution-reactivity of IDPs/IDPRs from the analyses of their respective hydrophobic burial profiles.

Materials and Methods

Selection of IDPs

To study the salt-bridge dynamics in IDPs/IDPRs, a set of four disordered proteins were chosen with varying degrees of structural disorder (ranging from 43 to 100%) in their native states, containing oppositely charged amino acids (roughly 1/3rd) ensuring the formation of salt-bridges (**Table 1**). The set consisted of two partially disordered proteins (IDPRs); i) the scaffolding protein GPB from Escherechia virus phix174, PDB ID 1CD3, chain ID: B, ii) the human coagulation factor Xa, PDB ID: 1F0R, chain ID: B and two completely disordered proteins (IDPs), namely α-synuclein (α-syn) and β-amyloid (β-amyl). Both IDPRs, 1CD3 and 1F0R consist of long contiguous stretches of disordered regions characterized by missing electron densities in their respective PDB files. These disordered stretches in IDPRs are located at the N-terminus for 1F0R, while for 1CD3 it is mainly confined to the middle regions. The sequences of the IDPs are obtained from the DISPROT database [21].

Atomic Model Building

The X-ray crystallographic structures of both 1CD3 and 1F0R (resolution: 3.5 Å & 2.1 Å and R-factor: 0.275 & 0.216 respectively) have long contiguous stretches of missing coordinates (i.e., missing electron densities) corresponding to the disordered regions. These missing disordered residues were identified by comparing the SEQRES and ATOM records in their corresponding PDB files and cross-checked with those declared in the REMARK 465 list. The disordered regions for the partially as well as completely disordered proteins were modeled using MODELLER [22]. The modeling of the disordered regions is done in a manner such that it exactly preserves the structure of the ordered part of the protein, i.e., the RMSD (root-mean-square deviation) of the ordered part of the protein is exactly zero. Such methods were followed in earlier works [5, 6].

Molecular Dynamics Simulation

Explicit-water Molecular Dynamics (MD) simulation trajectories for 1CD3, 1F0R and α-syn used in the current calculation were directly obtained from a previous study [6] while, for β-amyl, an identical simulation protocol was followed [6]. The simulations were performed with AMBER 12 program[23] at T=300 K using the ff99SB force field [24, 25] with periodic boundary conditions and TIP3P water model [26]. Energy minimization of each solvated protein was performed twice via the conjugate gradient method to remove unfavorable steric interactions. The energy minimized solvated protein was equilibrated in a NVT ensemble for 100 ps at an initial temperature of 100 K , while the temperature was gradually raised to 300 K at constant volume. This was followed by NPT equilibration for 5 ns at a constant temperature of 300 K and a pressure of 1 bar. A NPT production run of 100 ns with a time step of 2 fs was performed on the equilibrated system of each protein. Berendsen's temperature bath was used to maintain constant temperature with a coupling constant of 2 ps, while constant pressure was regulated by a barostat with a coupling constant of 1 ps.

Trajectories were written at an interval of 2 ps, resulting in 50000 frames (or snapshots) and all analyses were performed on the post-equilibrium 100 ns long trajectories for all four proteins.

Globular Protein Database

To depict and quantify the time evolution of secondary structural transitions in the simulated conformations of IDPs/IDPRs, knowledge-based weights were assigned for each secondary structural class. Thus, a database of high resolution X-ray structures of globular proteins were assembled using the advanced search protocol of PDB [27] with the following culling criteria: (i) resolution not worse than 2 Å (ii) neither working nor free R-factor worse than 0.2 (iii) files containing only uncomplexed proteins without DNA / RNA / DNA-RNA hybrid, (iv) a sequence identity of maximum 30% (v) only wild type proteins including expression tags and (vi) percentage coverage of sequence 90% (Uniprot) [28]. The application of the above culling criteria resulted in 2777 unique chain entries which are mapped to 2692 PDB structures. The resulting database is referred to as GDB and was used as a benchmark to assess and evaluate certain parameters related to IDPs/IDPRs for the first time, to the best of our knowledge.

Secondary Structural Assignments

Secondary structures (helices, strands, coils etc.) were determined from the atomic coordinates by STRIDE [29] and assigned to each amino acid residue in the polypeptide chain based on the available knowledge of both hydrogen bonding pattern and the backbone geometry of existing protein structures in the PDB.

Global Secondary Structural Transition Scores

Scoring functions were designed to capture the time evolution of the secondary structural transitions of the respective protein chain. Two such equivalent scores were designed such that they may be mutually cross-validated. From the secondary structural assignments described above, the fraction of residues assigned to a particular secondary structural class was calculated and used as a statistical weight for secondary structural transitions along the MD trajectories of the IDPs/IDPRs (**Table 2**). The assigned secondary structures were then grouped into two disorder-classes, 'ordered' and 'disordered', based on regularity of the backbone geometry. The 'ordered' class consisted of α-helices, β-strands, 3_{10}-helices and the rare occurrences of Π-helices while the disordered class comprised of turns, coils and bridges. The transition of each residue between its assigned secondary structures at two consecutive time intervals (say at time t and t+Δt) was accordingly classified as either 'disorder-to-order' or vice-versa with a concomitant statistical weight. The transitions that did not result in a change of order-disorder-class were considered as neutral. Based on the above scheme, three scores were defined. The first of the three is the simple ratio (*frt*) of the number of residues undergoing a disorder-to-order transition or vice-versa (N_{trans}) to the total number of residues in the polypeptide chain (N_{tot}), i.e., frt=Ntrans/Ntot. The other two transition scores (gsst1, gsst2) were defined as:

$$gsst1 = \frac{\sum_{i=1}^{N_{trans}} \left| \left(wt_2(i) - wt_1(i) \right) \right| \cdot sign}{N_{trans}} \qquad \text{- (i)}$$

$$gsst2 = \frac{\sum_{i=1}^{N_{trans}} \sqrt{\left(wt_1(i)^2 + wt_2(i)^2\right)} \cdot sign}{N_{trans}} \quad \text{- (ii)}$$

where wt_1 and wt_2 are the secondary structural weights (**Table 2**) of the ith amino acid residue at two consecutive time intervals, t and t+Δt, and the variable 'sign' is set to +1 for disordered → ordered (taken to be the preferred direction) and -1 for the opposite direction (ordered → disordered).

Identifying salt-bridges

In accordance to a recent study [17], ionic bonds within disordered proteins were detected when a positively charged nitrogen atom from the side-chains 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 side-chain oxygen atom of glutamate (OE1, OE2) or aspartate (OD1, OD2).

Salt-bridges Networks

The construction of salt-bridge networks (also known as 'salt-bridge motifs') was extensively discussed in a previous study [17]. Briefly, in these networks, inter-connected charged residues were represented as nodes and the ionic bonds between them as edges. Each network (or motif) was then represented numerically describing its topology. This numerical string called the 'motif identifier' is a concatenated string of numbers (separated by a delimiter '-') where each number-string stands for the topological description of a node in the network; the first digit in the number-string refers to the degree of that node, and the other digits refer to the degrees of its direct neighbors, sorted in a descending order. These concatenated number strings are then represented as elements of an array, further sorted in descending order, and concatenated by a delimiter '-'. The formulation of the motif identifier is based on the assumption that the degrees of such nodes will be restricted to one digit numbers (i.e., a maximum of 9) – which holds true for a wide plethora of contact networks among proteins [33] including ionic bond networks [17]. A pictorial illustration of the numeric scheme can be found in Figure 2 of the original work [17].

Dynamic Persistence of salt-bridges

To calculate the coordinate-based dynamic properties of salt-bridges, simulated conformations were examined at a regular time interval of 50 ps across the entire 100 ns simulation trajectory for each protein. This resulted in a total of 2000 protein conformations assembled from different time frames. The dynamic persistence (*pers*) of a particular salt-bridge was calculated as the ratio of the number of protein conformations containing the salt-bridge with respect to the total number of conformations examined. Persistence was calculated for all ionic bonds that were formed at least once in the entire simulation trajectory. An optimum persistence cut-off of at least 25% (i.e., $pers \geq 0.25$: a salt-bridge found in at least ¼th of all sampled frames) was considered (see section **Persistent salt-bridges** in **Results and Discussion**) to filter out the most relevant 'persistent' salt-bridges. To check the consistency of the results, the entire calculation was repeated at smaller (20 ps) and larger (100 ps) time intervals, as compared to the original time interval of 50 ns. The deviation in the corresponding persistence values were found to be negligible, as reflected in the calculated pairwise root mean square deviations (0.0035: 20 ps vs. 50ps; 0.0051: 100 ps vs. 50 ps) (**Supplementary Table S1**) of the persistent salt-bridges. The choice of the 50 ps time interval was motivated by the range of reported time scales for secondary structural transitions in proteins [31].

Contact Order

Contact Order (CO) of the interacting pairs of ions involved in a salt-bridge was determined as the number of residues that separates any two charged amino acids in the sequence space of the respective amino acids divided by the full length of the polypeptide chain as reported in an earlier study [17]. The analysis of the distribution of CO-values for the salt-bridges are guided by a previous work [17]. In contrast to the established use of CO as an index of the refolding rates in globular proteins [32], this measure was used for the disordered proteins to investigate whether the ionic bonds are able to bring distant parts of the chain in close proximity; and, if so, how persistent are such long-range ionic bonds in IDPRs/IDPs.

Combined Disorder Status for Persistent salt-bridges

One of the major goals of this work is to depict the role of salt-bridges in the dynamics of conformational interconversions in IDPs/IDPRs. To this end, the each charged residue involved in a salt-bridge, in a given time frame, was categorized as either 'Disordered (D)' or 'Ordered (O)' according to its assigned secondary structure (STRIDE [29]) combined with the classification scheme discussed in the section **Global Secondary Structural Transition Scores**. Hence, the 'combined disorder status' (CDS) of two charged residues involved in a salt-bridge in a given conformation may be classified as either disordered (DD) or ordered (OO) or hybrid (DO / OD). Each possible transition that a salt-bridge may undergo in terms of this CDS between two conformations at consecutive time intervals (e.g., OO → DD, OO → DO etc.) may be assigned a score based on how many of the charged residues involved in the salt bridge pair actually undergo the transition. In accordance, the following scoring scheme (**Table 3**) was adapted. Special care was taken to assign a score resulting from the swapping of hybrid CDS's (i.e., OD → DO and vice-versa).

These scores (or transition weights) for a particular pair of charged residues in a salt bridge were then summed for the conformations at two consecutive time intervals along the simulation trajectory and divided by the total number of such transitions, generating the conservation score for salt-bridge secondary structures (CS_{sec}). Actually, CS_{sec} is a discrete probability measure defined in the range of 0 to 1; where 0 and 1 refers respectively to cases of none and both residues in the pair undergoing changes in their respective disorder-status.

Shape Factor

Shape factor (ρ) is a well known descriptor of shape and compactness for polymers. This parameter (also known as internal structure factor) approaches a value of 1.5 and 0.77 for an ideal chain (also called random coil) and a compact sphere respectively [33–36]. Shape factor may be calculated from the atomic coordinates [37] as the ratio of the radius of gyration (R_g) and the hydrodynamic radius (R_{hyd}), where the two radii, R_g, R_{hyd}, are defined by the following expressions:

$$R_g = \sqrt{\frac{1}{N}\left\langle \sum_{i=1}^{N} \left(r(i) - r_{mean}\right)^2 \right\rangle} \qquad \text{- (iii)}$$

$$\frac{1}{R_{hyd}} = \frac{1}{N^2}\left\langle \sum_{i \neq j} \frac{1}{r(i,j)} \right\rangle \qquad \text{- (iv)}$$

where r(i) is the distance of the ith atom from the centroid of the molecule, denoted as r_{mean} and r(i,j) is the distance between the ith and the jth atoms in the polymer. Thus, R_g may be defined as the root mean square deviation of all atoms from the centroid of a molecule which is then divided by R_{hyd} to nullify the size-effect, so that, ρ may be compared across different sizes of polymers. Shape factor was calculated for different conformations of IDPs/IDPRs to explore the dynamic shape transition of these proteins along the simulation trajectory due to the salt-bridge dynamics and other electrostatic interactions.

Burial of Solvent Exposure

The solvent accessible surface area (ASA) was calculated for each protein atom by NACCESS [38] as in earlier studies [17, 39]. The ASAs were summed up for all atoms for a given residue. The extent of burial (*bur*) or solvent exposure for a specific residue X was calculated as the ratio of the ASA of X located in the protein to the ASA of the same amino acid in a Gly-X-Gly peptide fragment in its fully extended conformation [40]. Residues were distributed according to the extent of burial or solvent exposure with empirical boundaries standardized in earlier works [17, 30, 40]. Precisely, four burial bins were considered that quantifies the degree of solvent exposure: bur1: $0.0 \leq bur \leq 0.05$ (completely buried); bur2: $0.05 < bur \leq 0.15$, bur3: $0.15 \leq bur \leq 0.30$ (partially buried) and bur4: $bur > 0.30$ (completely exposed).

Accessibility Score

The accessibility score (*rGb*) is a knowledge-based measure that compares the hydrophobic burial profile of a given protein with respect to those of the highly resolved globular protein structures. The empirical ranges of this measure have been standardized on globular proteins [41] and also at protein-protein interfaces [42] and they seem to unambiguously agree with each-other. It may be defined as follows.

$$rGb = \frac{1}{N_{res}} \sum_{i=1}^{N} \log_{10}(\mathrm{Pr}_i) \qquad \text{- (v)}$$

where N_{res} is the total number of residues in the polypeptide chain, Pr_i is the propensity of the *ith* amino acid (Ala, Glu, Asp etc.) to acquire a particular degree of solvent exposure. Other details of the score is defined elsewhere [41]. The score is a strong indicator of stability or instability of a protein / peptide in solution while instability may be further extended to explain reactivity.

Electrostatic Complementarity and Score

Electrostatic complementarity (E_m) for amino acid residues in all proteins were calculated precisely following previous studies [17, 43] and a modified version of a log odd probability score (P_{Em}) was computed for the whole protein chain as an overall descriptor of electrostatic interactions adapted from a previous formulation [41]. Briefly, the score can be defined as follows.

$$P(E_m) = \frac{\sum_{i=1}^{N_t} \log_{10} P_i(E_m)}{N_t} ; P_i(E_m) \neq 0 \qquad \text{- (vi)}$$

where N_t is the total number of amino acid residues in a given polypeptide chain and $P_i(E_m)$ is the probability of the ith amino acid residue to acquire an E_m value which falls in a particular bin of

discrete probabilities (P ($x < E_m < (x+0.05)$)) within the defined range of E_m, varying from -1 to +1. Rather than considering only the completely or partially buried residues [41], all four burial bins [41, 43], including the solvent-exposed residues were considered in this work. Also, as opposed to the original study [41] where P_{Em} was calculated only on E_m^{sc} (E_m computed on the side-chain van der Waal's surface of the target amino acid [43]), here, the modified $P(E_m)$ score was calculated for all three E_m measures, namely, E_m^{all} (E_m calculated on the van der Waal's surface of the entire residue of the target amino acid), E_m^{sc} corresponding to its side-chain surface and E_m^{mc} computed on its main-chain surface [43]. These additions have modified the representation of the score, $P(E_m)$ in contrast to P_{Em} ([41]) such that the three different E_m measures can be conveniently denoted as its potential argument (viz., $P(E_m^{all})$, $P(E_m^{sc})$, $P(E_m^{mc})$). All force field parameters (partial charges, van der Waals radii), wherever applicable, were used in consistency with the original work [43].

Dynamic Cross Correlation Map (DCCM)

The Dynamic Cross Correlation Map (DCCM) [44] captures the time correlated positional fluctuations due to the concerted movement of the salt-bridge forming oppositely charged residues. This correlation map is a visual portrayal of the DCC(i,j), defined as follows.

$$DCC(i,j)=\frac{\langle \Delta\vec{r}_i \cdot \Delta\vec{r}_j \rangle_t}{\left(\langle \Delta\vec{r}_i^{\,2} \rangle_t \cdot \langle \Delta\vec{r}_j^{\,2} \rangle_t\right)^{1/2}} \qquad \text{- (vii)}$$

where i and j refer to the reduced single atomic representations (as defined below) of any two charged residues (taken pairwise from a pool of all charged residues) in the molecule. The displacement vector of the ith residue from its mean position at the tth snapshot of the simulated trajectory is denoted by Δr_i (i.e., its representative atom). The map was constructed respectively for both main- and side-chains of each protein. For the main-chain, the centroid of the molecular unit of N_i-C^{α}_i-C_i was considered as the representative of the ith residue while for the side-chain, the charged atoms in the ionic groups were considered. For Lysine, the only charged side-chain atom NZ is considered, while for residues with bifurcated ionic groups involving two degenerate charged atoms of the same type (N/O), the mean position of the two charged atoms were considered (i.e., ARG: (NH1+NH2)/2; GLU: (OE1+OE2)/2; ASP: (OD1+OD2)/2; HIP: (ND1+NE2)/2). For a given protein, all structures obtained from a given simulated trajectory were first superposed on a common template to bring the coordinates on the same frame. After superposition, the time averaged mean positions of all atoms and those of the relevant atomic units were computed, followed by the calculation of the displacement vectors for each representative atom (as in **equation (vii)**) at every sampled (t^{th}) snapshot in the trajectory. 2000 equispaced snapshots were sampled at an interval of 50 ps covering the entire 100 ns trajectory. The dot product of the displacement vectors Δr_i and Δr_j at each (t^{th}) snapshot were then computed, normalized by the product of their individual dot products. This 'normalized dot product' term was then time-averaged over all snapshots in the simulated trajectory giving DCC(i,j). Hence, the values of DCC(i,j) varies between -1 to 1; wherein, -1 and 1, respectively represents perfect anti-correlation and perfect correlation between the movements of the ith and the jth atoms, sustained throughout the entire dynamic trajectory; while a DCC(i,j) of 0 means no correlation. Although the DCC(i,j) matrix consists of entries corresponding to all charged residues in each protein, we are only interested in the oppositely charged pairs forming the persistent salt-bridges.

Cross Correlation

The pair-wise cross-correlation between two discrete time series were enumerated using the 'Xcorr' function of MATLAB wherever applicable. Xcorr returns a cross-correlated sequence (i.e., a

column vector) containing 2N+1 cross-correlation values corresponding to that many time lags (where N is the number of entries in both the discrete time series). The middle most (i.e., the $(N+1)^{th}$) entry in the sequence corresponds to the zero-lag-cross-correlation – which is trivially the point of interest (unless specifically stated otherwise).

Prior to calculating the cross-correlation, all time series were scaled such that all values range between 0 to 1 using the min-max scaling technique: $xt_i=(x_i-x_{min})/(x_{max}-x_{min})$ where x_{min} and x_{max} are the minimum and the maximum values obtained from a given series and x_i and xt_i represent the ith entry in that time series before and after scaling.

In addition, the cross correlation coefficient (CCC_τ) defined as follows was also calculated for the zero time lag.

$$CCC_\tau = \frac{1}{\sigma_x . \sigma_y} E\left(\left(x_t - \mu_x\right).\left(y_{(t+\tau)} - \mu_y\right)\right) \qquad \text{- (viii)}$$

where x_t and y_t are two time series varying with time t, their mean μ_x, μ_y and standard deviations σ_x, σ_y respectively; E denotes the expectation value (time average), the numerator also known as the cross-co-variance and τ stands for the time lag, taken to be zero for calculating CCC_0.

Results and Discussion

IDP sequences are often abundant in polar and charged residues. These charged residues may carry like or opposite charges and may accordingly modulate the overall electrostatic balance of the protein and its binding. Salt-bridges (or ionic bonds) are the major electrostatic components involving the -NH3+ and COO- groups of the charged amino acids. However, formation of dynamically persistent salt-bridges in IDPs, especially those having long range contacts, should, in principle be a deterrent to their intrinsic flexibility, imparting local structural rigidity, analogous to the disulphide bridges in context of globular proteins. Thus there is a subtle balance between the optimum dynamic flexibility in the IDPs/IDPRs and an allowed set of salt-bridges (out of all possible pairwise combinations of residues carrying opposite charges) such that these proteins are able to accommodate the charged residues without compromising their dynamic flexibility. In this work, we investigate this trade-off in both partially and completely disordered proteins by analyzing the time evolution of the salt-bridge motifs (see **Materials and Methods**) along their simulated MD trajectories.

Distribution of salt-bridge motifs in the dynamic trajectories of IDPs

Composite salt-bridge motifs were identified and accumulated (see **Materials and Methods**) along the 100 ns MD simulation trajectories for all four IDPs/IDPRs. 2000 equispaced MD frames sampled at 50 ps were collected and analyzed to obtain a distribution that revealed the preponderance and insignificance the different salt bridge motifs. Essentially we estimate the frequency of occurrence of each salt bridge motif along the simulated trajectories of each protein. The majority of salt-bridge motifs were found to be isolated (motif identifier: 11-11), amounting to 84.8%, 80.2%, 89.9% and 98.8% for 1CD3, 1F0R, α-syn and β-amyl respectively (**Figure 1**) similar to the earlier findings on globular proteins, modeled structures of partially disordered proteins and protein-protein interfaces [17]. In accordance with increasing complexity, the second most frequent motif was found to be bifurcated salt-bridges (211-12-12) accounting for 7.7%, 15.9%, 8.9% and 1.2% in these four proteins respectively. The frequency of these motifs gradually

decreased with increasing size and/or complexity of the networks. Other noticeable low-frequency motifs constituted of (i) four-membered open linear chains or trees (221-221-12-12: 5.8% in 1CD3, 1.6% in 1F0R, 1.1% in α-syn), (ii) four-membered closed rings also known as 4-cycles (222-222-222-222: 1.4% in 1CD3) and (iii) open motifs with four nodes including a central node, also known as 3-stars (3111-13-13-13: 1.8% in 1F0R). The rest of the motifs mostly containing greater than 4 nodes were found in abysmally low frequencies. The partially disordered proteins displayed an overall higher frequency of salt-bridges as revealed by the difference in the total number of motifs (1CD3: 13698, 1F0R: 14055) as compared to the completely disordered ones (α-syn: 8188, β-amyl: 3190). This difference is due to (i) the significantly smaller length of the fourth protein (~1/3rd of the other three) and (ii) the lower fractional content of the charged residues in the IDPs (~28%) as compared to the IDPRs (~34%) **(Table 1)**.

Since dynamically persistent salt-bridges counteract the flexible nature of the IDPs, imparting local rigidity, it may be interesting to find salt-bridges in IDPs/IDPRs, which may be conserved or even grow to form composite salt-bridge networks along their respective simulation trajectories. It is obvious that the salt-bridge motifs undergo transitions via formation and breakage of ionic bonds, which is evident from the analysis of the time evolution profiles (in terms of counts) of each particular motif along the MD trajectories. The length and the fraction of charged residues **(Table 1)** of each protein influences these counts.

The number of isolated salt-bridges in both IDPRs were found to vary between 1 to 12 (mean: 5.8 ±2.1 for 1CD3, 5.7 ±1.8 for 1F0R) while for the IDPs, these numbers vary between 0 to 9 for α-syn (mean: 3.69 ±1.39) and 0 to 5 for β-amyl (mean: 1.58 ±0.92) throughout the 100 ns simulation trajectory (**Figure 2**). For the second most prevalent motif, namely, the bifurcated salt-bridges, the time-averaged counts were 0.53 (±0.63), 1.24 (±0.76), 1.22 (±0.68) and 0.02 (±0.14) for 1CD3, 1F0R, α-syn and β-amyl respectively (**Figure 3**). One interesting feature was the periodic appearance of at least one **(Figure 4)** and at most two (**Supplementary Figure S1**) 4-cycles or closed quadruplets without the diagonal edges (222-222-222-222), found within the 75-100 ns time frames found in all three proteins except β-amyl, which completely lacks the motif. Interestingly, almost all residues forming these closed quadruplets originated from the loops and coils, thereby bringing together the distant disordered regions of α-syn and the IDPRs (1CD3, 1F0R). In fact, this closed ring core topology was found in the distribution of salt-bridge motifs **(Figure 1)**. It may be noted that the charge constraints in salt-bridges only allow for an even number of nodes in a closed ring topology (or cycle) to be found within proteins [17].

Persistent salt-bridges

Another pertinent problem involves the dynamic persistence of the salt-bridges – whether they are sustainable enough in the long run. To this end, the topological details of each salt-bridge motif were totally ignored and the complete distribution of the salt-bridge motifs were reconsidered as a non-redundant set of discrete ionic bonds. The dynamic persistence (*pers*) of a salt-bridge is defined as the fraction of the time-frames in which the salt-bridge / ionic bond appears with respect to the total number of time-frames sampled in the trajectory. Initially, all salt-bridges were counted that occurred at least once in the sampled time frames. These numbers were found to be 69, 102, 75 and 15 for 1CD3, 1F0R, α-syn and β-amyl respectively. To assess whether the ionic bonds in these salt bridges were formed by chance, as a probabilistic consequence of the simulation protocol, their persistence values were examined. Hence, only those salt-bridges, which were formed in at least 1/4th ($pers \geq 0.25$) of the sampled time frames were screened for further analyses, accounting for slightly more than 10% of the ionic bonds. The choice of this cut-off was based on the overall frequency distribution of the persistence values (**Figure 5A**) with a large fraction of insignificant

contacts that have likely occurred by chance, detected only in a few frames (~67% of the contacts having a persistence of 5% or less).

This persistence cutoff drastically reduced the number of salt-bridges to 10, 13, 5 and 2 for 1CD3, 1F0R, α-syn and β-amyl respectively by discarding the insignificant transient salt-bridges. Given the overwhelming majority of these short-lived contacts, the fluctuating dynamics of these ionic bonds may influence the local flexibility of these proteins. The calculated persistence values spanned a range from 25% to 96% with instances of dynamically stable (say, the 80%-persistent) salt-bridges in all four proteins with values hitting almost every possible 10% bin-interval in the range (**Figure 5B**). The overall distribution also portray that the salt-bridges span different time scales ranging from instantaneous, brief, moderate and long-lived. Thus, the formation and dissolution of salt-bridges appear continuously throughout the entire simulation trajectory. This dynamics of salt-bridges supports a physical model for dynamic flexibility, which also appeared to be true visually (**Supplementary Video S1**).

The other noticeable feature, especially relevant for the ephemeral salt-bridges was the transient formation and breakage of ionic bonds associated with a particular charged atom (say, NZ of 63-LYS in 1CD3) and various oppositely charged atoms originating from different residues throughout the main-chain. This dynamic interchange of charged-atom-pairs in fluctuating ionic bonds structurally supports different alternative conformations at different temporal patches along the MD trajectory, consistent with the notion of the conformational flexibility [45] in IDPs/IDPRs. An illustrative example is shown in **Figure 6** where the ε-NH3+ group of 28-Lys in β-amyl has been found in contact alternatively with four different negatively charged amino acids (i.e., with their side-chain –COO- groups) at different temporal patches along the trajectory: 23-Asp (14 - 24 ns), 1-Asp (61-62 ns, 70-75 ns), 22-Glu (63 – 69 ns, 78 – 89 ns) and 3-Glu (97 – 99 ns). IDPs/IDPRs are well known for their binding promiscuity, [2, 46] while here we find another level of promiscuity in them, namely, in ionic bond formation. This also supports the concept of dynamic flexibility that makes them structurally adjustable and adaptable to different conformations upon binding to different molecular partners.

In parallel, it was also important to know, whether these ionic bonds were formed between residues that were close or distant in the amino acid sequence space. To that end, contact orders (see **Materials and Methods**) were calculated for all salt-bridges with at least 25% persistence. Most contacts (~43%) were short-ranged (CO<0.1), however, there were quite a few (~40%) middle-range contacts (0.1<CO<0.3) as well as some (~17%) long-range persistent ionic bonds (CO>0.3) too. In particular, the IDPRs (1CD3, 1F0R) contained quite a few middle and long range contacts (**Table 4**) and, interestingly, majority of such contacts involved an arginine (100% for 1CD3, 67% for 1F0R). Visual inspection in Pymol revealed that most contacts were susceptible to form local (short-range contacts) and non-local turns or loop-closures (long-range contacts) while some of them linked between the coils and remnants of regular secondary structural regions in partially disordered proteins, e.g., helical turns in 1CD3, beta strands in 1F0R (**Figure 7**). For IDPs, persistent salt-bridges were exclusively found between the coils imparting some dynamic structural constraints which appeared to act against the elongated coil structure (as obtained in the initial MODELLER model) along the simulation trajectory. In α-syn, a bifurcated salt-bridge (23-LYS~28-GLU-34-LYS) was also found to be dynamically persistent even though it comprised of local contacts. In β-amyl, only two persistent salt-bridges were found, and both of them were on the same side of a short-helix (located towards the C-terminus of the protein and consisted of 9 residues,), which seem to gradually evolve during the course of the simulation (first appearing at around 28 ns) and found to be frequent and stable thereafter (found in 75% of the frames from 28 to 100 ns). These salt-bridges were thus structurally non-interrupting for the short helix. It is also noteworthy

that except for an outwardly facing Lysine (28-Lys), the helix is otherwise devoid of any polar or charged residues; rather it consists of hydrophobic residues (Ala, Met, Leu, Ile) and Glycines only. The other interesting fact was that all charged residues (including those that does not get involved in dynamically persistent ionic bonds) were exclusively found towards the N-terminus of this short-helix, leaving it free and unperturbed (**Figure 8**). The secondary structures in the snapshots were depicted by STRIDE [29] and confirmed visually by Pymol and VMD [47].

Global Secondary Structural Transition

The time evolution of the global secondary structural transitions for the entire protein chain were estimated by two equivalent transition scores (gsst1, gsst2) described in the **Materials and Methods**. Since there are empirical bounds on both scores based on their formulation and weights (see **Materials and Methods**), it would be appropriate to understand their defined numerical ranges. As can be seen in **Supplementary Figure S2** gsst1 gives a flat plane, since this score is based on just the absolute difference of two weights (wt_1 and wt_2) whereas gsst2, being a resultant vector modulus or 2D Euclidean distance, returns a rugged bi-modal surface; rugged because of the discrete sampling of the weights. Both scores will give 'zero' for a neutral transition (see **Materials and Methods**).

Along MD trajectories of each protein, the transition scores (gsst1, gsst2) were calculated between every two consecutive (i^{th} and $(i+1)^{th}$) snapshots sampled at time t and t+Δt respectively. Taking consecutive snapshots makes the time evolution profiles smooth and continuous capturing even the smallest temporal fluctuations. The fraction of the residues undergoing the disorder-to-order secondary structural transitions (frt) was also calculated for 2000 conformations corresponding to successive equispaced time frames sampled at an interval of 50 ps. The time averages were computed for each measure with their standard deviations (**Supplementary Table S2**) from their corresponding time trajectory. There were 3% to 7% average structural transitions between the consecutive snapshots as reflected in <frt>. It is natural to expect that the time averages (<gsst1>, <gsst2>) should be close to zero for both transition scores for all proteins because the fluctuations are computed in a pairwise manner between the consecutive time frames. The standard deviations in both <gsst1> and <gsst2> were of similar order of magnitude(s) for all four proteins.

The theoretical equivalence of the two transition scores (gsst1, gsst2) has been discussed in detail. Hence, it was natural to examine their agreement revealed from actual data (i.e., the two discrete time series). This was done by calculating the normalized-cross-correlation sequences between the corresponding time series. To that end, both series were first normalized using the min-max scaling technique (see **Materials and Methods**). This linear transformation scales all values from 0 to 1 such that the cross-correlation values across different profiles may be compared. In this way, it is also convenient to set a random baseline corresponding to the min-max scaling of the series. This random baseline serves as a necessary reference to assess the significance of the cross-correlation values obtained. To standardize such a random baseline, 10,000 pairs of random number sequences of lengths equal to the number of sampled conformations at time intervals of 50 ps (i.e., 2000) were generated and their maximum and zero-lag cross-correlation values were stored. The random numbers are trivially defined between 0 to 1. The mean (μ_r) and standard deviations (σ_r) obtained in the maximum and zero-lag cross-correlation values enumerated from the random sequences were 506.06 ±9.20, 499.85 ±9.80 respectively, closely resembling each other. Hence, we set a reference to decipher the significance of the cross-correlation values.

The magnitudes of the mutual cross-correlation calculated between the two discrete normalized series (gsst1, gsst2) at zero lag are in very good agreement among all four proteins (1CD3: 596.81,

1F0R: 645.43, α-syn: 622.7, β-amyl: 615.43). These zero-lag cross-correlation values were mapped to the corresponding maxima obtained in the cross-correlated sequences. The values were well above the (μ_r +3σ_r) threshold indicative of significant correlations. Another noticeable feature was the abrupt and highly localized peak of the cross-correlated sequences obtained at zero lag, (**Supplementary Figure S3**) - which gives an impression of an auto-correlated sequence as quantified by two transition scores (gsst1 or gsst2). This shows the coordination of these two formulated measures in actual data-space.

Next, attempts were made to reveal if there is any potential dynamic correlation between either of the two transition scores (gsst1, gsst2) and the number of salt-bridges (N_{sbg}) formed at each time interval. All relevant quantities, i.e., gsst1, gsst2, N_{sbg} were first normalized using the min-max scaling technique described above. Hence, all cross-correlations were computed on values scaled between 0 to 1. All four proteins showed comparable correlations (**Supplementary Figure S4, S5**) in both scores (gsst1 vs. N_{sbg}, gsst2 vs. N_{sbg}) with their respective maxima asymptotically approaching their corresponding zero-lag cross-correlation values (**Table 5**). However the correlations may be insignificant due to the maximum cross-correlation values falling well below the random baseline (μ_r). This was further confirmed by the corresponding zero-lag-cross-correlation coefficient values (CCC_0, defined between -1 to 1, see **Materials and Methods**), approaching zero for all four proteins (1CD3: 0.02, 0.02; 1F0R: -0.02, -0.02; α-syn: -0.03, -0.02, β-amyl: 0.01; 0.02 for gsst1, gsst2 respectively). In contrast, the CCC_0 values between gsst1 and gsst2 were found to be 0.91, 0.93, 0.93 and 0.92 for 1CD3, 1F0R, α-syn and β-amyl respectively.

Thus the global transition scores are rather limited to capture the effect of salt-bridge dynamics in depicting the disorder-order transitions in IDPs/IDPRs, demanding for additional local measures of secondary structural transitions which are direct measures of the formation and breakage of salt-bridges.

Salt-bridge mediated Secondary Structural Transitions

Since the global transition scores were not informative in terms of deciphering the contribution of salt-bridge dynamics regarding the time evolution of the secondary structural transitions, a measure involving the combined disorder status (CDS) of persistent salt-bridges is used, which is elaborated in the **Materials and Methods.** This parameter, namely, the conservation score for salt-bridge secondary structures (CS_{sec}) takes into account both the number as well as the type of transition (see **Materials and Methods**). CS_{sec} will give zero for no transitions (i.e., 100% conservation) and will increase with increasing number of transitions. Also, the individual weights for each possible transition (e.g., OO → DD, OO → DO etc.) were standardized in such a manner that the maximum value of CS_{sec} is one, corresponding to a persistent salt-bridge of the non-hybrid type (i.e., DD/OO) undergoing transition in both of its residue-disorder-status between each pair of consecutive time intervals along the simulation trajectory (e.g., OO → DD → OO → DD … and so on) (see **Materials and Methods**). The relative occupancy of a predominant secondary structural motif (PSSM) was also calculated (e.g., Coil~Turn, α-Helix~α-Helix etc.) for each persistent salt-bridge as a fraction of the most frequent secondary structural motif relative to all motifs found in all time frames (see **Materials and Methods**). For a poor conservation score (corresponding to a salt-bridge more or less conserved in its secondary structural status along time), this PSSM value must be mapped to its major secondary structural motif.

For the partially disordered proteins, the conservation scores (CS_{sec}) for the persistent salt-bridges spanned across the entire range from 0 to 1 (**Table 6, Supplementary Table S3**). Though, most of

the persistent salt-bridges were highly conserved in terms of their 'combined-disorder-status', there were some highly variable (98-ASP~100-ARG: 0.80, 22-GLU~64-ARG: 0.77 in 1CD3; 82-LYS~97-GLU: 0.99 in 1F0R) salt-bridges as well. All highly variable salt-bridges were found to contain regular secondary structural elements. Also, these high-CS_{sec} salt-bridges correspond to low- to middle-range contacts that spans across a broad range of persistence values (0.37 to 0.81). No correlation was found between the *pers* and CS_{sec} and hence they appear to be independent dynamic properties of these proteins. The predominant secondary structural motif for the salt-bridges in 1CD3 involved both structured as well as unstructured coiled regions, whereas, those for 1F0R were mainly unstructured.

For the completely disordered proteins, all persistent salt-bridges were highly conserved in their 'combined-disorder-status' except one particular salt-bridge: 28-GLU~34-LYS (CS_{sec}: 0.76) where the predominant secondary structural motif involves a rare 3_{10}-Helix (PSSM: 0.754). This is expected since these proteins rarely involve ordered regions along the simulation trajectory and due to the fact that the above discussed (see section: **Persistent salt-bridges**) short-helix formed in β-amyl (**Figure 8**) is mostly devoid of any charged residues (but for the terminal Lysine: 28-Lys) thereby eliminating its scope to participate in a salt-bridge.

Dynamic bending in IDP chains

A protein, being a hetero-polymer, can not grow linearly for an indefinite period in space; it has to take a turn at some point [31, 48]. This was the key idea behind the proposition of the cotranslational folding model in proteins [49, 50], later proven experimentally [50] and supported by plausible genetic mechanisms (viz., pause codons) [51]. The model states that a protein chain begins to fold as soon as its N-terminus is exposed to the folding site from the translocation complex [49] – which is generally valid for all proteins undergoing ribosomal translation. In the current work, however, we were restricted to use atomic models, partially or completely built by MODELER, as the starting templates for the MD simulations (because of the lack of their experimental structures). These modeled structures resemble elongated random coils marked by the lack of any tertiary structures as revealed visually by Pymol. However, when the time-evolved simulated structures were displayed visually (in VMD), substantial dynamic bending with respect to the elongated initial template were evident as the electrostatic interactions overwhelmingly contributed to the persistent salt-bridges, bringing together distant regions of the protein chains (**Supplementary Video S1**). However, the curved structures changed continuously within a non-rigid (yet non-random) statistical ensemble of conformations, characterized as stochastic conformational switching in IDPs [45]. One way to quantify this bending or reshaping is to calculate the shape factors [37] which is the ratio of the radius of gyration and the hydrodynamic radius of the molecule (see **Materials and Methods**). Shape factor is an indicator of the shape and compactness of polymers accounting for their dynamic shapes induced by the change of chemical environments [52]. As a benchmark, shape factors of globular proteins (ρ_{glob}) were calculated from the database GDB (see **Materials and Methods**). The ρ_{glob} values were found to be extremely stable, (characterized by a standard deviation of ~1/18th of their mean value), restricted within a narrow bin of 1.0 ±0.054 (**Figure 9**). For calculating the shape factor in IDPs/IDPRs, the modeled structure was considered as the 0th snapshot to provide a reference value followed by 2000 post-equilibrium snapshots (sampled at 50 ps interval). The value of the shape factor for the template structures of the disordered proteins varied between 1.3 to 1.6, close to the theoretical values of ρ for ideal chains, calculated as 1.5 [33]. The lowest ρ value (1.23) was recorded for 1CD3 – a finding which is consistent with the fact that it has the highest fraction (~66.7%) of ordered secondary structure (**Table 1**) among the four disordered proteins. These initial ρ values were found

to fluctuate in the same range for a brief duration before attaining an equilibrium value of ~1.0 ±0.054.

Degree of Solvent Exposure: Approaching Porous Globules

Interior packing is not a primary factor involved in the dynamics of IDPs/IDPRs, as they adopt expanded conformations and remain sufficiently exposed to the solvent, while undergoing continuous dynamic bending. So to speak, the packing densities of disordered proteins are expected to be remarkably different from those of the globular ones, which are similar to that of the crystalline solids [53]. There are different measures of atomic packing with different degree of details [30, 37, 40, 54, 55], however, for each and all such measures, a primary condition to retain optimum packing is to have a substantial fraction of amino acid residues deeply or partially buried within the protein interior. Thus the fraction of buried and solvent exposed residues in these IDPs were examined along the simulation trajectory. The residues were then distributed in four previously standardized burial bins [40], and the fraction of the residues present in each bin was calculated. The calculation was performed on 500 snapshots sampled at 100 ps intervals for all four IDPs. The time-averaged fraction of residues present in the four burial bins (bin1: completely buried, bin2, 3: partially buried / partially exposed, bin4: completely exposed; see **Materials and Methods**) were 0.03 (±0.02), 0.07 (±0.03), 0.15 (±0.05) and 0.75 (±0.09) respectively for all four proteins. Clearly, the bulk share (~75-80%) of residues remain solvent exposed dynamically – which was also confirmed by the actual time evolution plots (**Figure 10**).

When the burial profiles of the IDPs/IDPRs are viewed together with their corresponding shape factor profiles – it is revealed that despite dynamic bending (as elaborated in the earlier section), the molecules still remain substantially exposed to the solvent. These features closely resemble that of the disordered collapsed globules [56, 57] often encountered in redesigned proteins with over- or under-packed cores [57]. Due to packing defects, these redesigned globules become enough porous and eventually result in the complete opening up of the structures, thereby allowing water molecules to penetrate into the protein core. The water molecules entrapped this way within the collapsed globules have further been characterized as potentially important for controling crucial IDP-events like binding induced folding and amyloid aggregation, by virtue of their unique kinetic features [56]. Also, since, optimum tertiary packing can not occur in these molecules with such high degree of solvent exposure (due to the lack of enough nearest neighboring surface points to pack against), these results also effectively rule out 'packing' to be a major component in the overall structural feature in the dynamic ensemble of IDPs/IDPRs.

To further characterize the hydrophobic burial profile of the amino acid residues in IDPs/IDPRs (with reference to the globular proteins), the accessibility score (*rGb*: see **Materials and Methods**) was also estimated for the time-evolved ensemble of structures. This parameter (***rGb***) estimates the distribution of different amino acid residues (i.e., hydrophobicities) with respect to their burial in a given structure and calibrates that structure with reference to the same estimate derived from natively folded globular proteins. However, the distribution of *rGb* is quite broad, characterized by a $\mu=1.5\sigma$[1] width both in globular proteins ($\mu=0.055 \pm 0.022$) [41] as well as in protein-protein complexes ($\mu=0.059 \pm 0.022$) [42] accounting for a variety of possible distribution patterns of hydrophobic burial as a function of shape and size of proteins. The time-averaged statistics obtained for *rGb* in IDPs/IDPRs registered significantly lower values as compared to that of the globular proteins (**Figure 11**). The *rGb* value is slightly above zero for IDPRs (1CD3, 1F0R), while it is negative for the IDPs (α-syn, β-amyl). All values are out of range with respect to the globular protein reference and the value for β-amyl is exceptionally low. This means that in IDPs/IDPRs, the

1 μ: mean, σ: standard deviation

trend of hydrophobic burial is actually opposite to that in globular proteins, where, most hydrophobic residues remain exposed to the solvent. While, the charged residues may, in fact, sometimes get partially buried, surrounded by enough protein atoms due to the formation of ionic bonds – more plausible for residues involved in the composite salt-bridge motifs due to their close proximity. Indeed, ~21.7% of the charged residues forming persistent salt-bridges (*pers* ≥ 0.25) were found to be at least partially buried (time-averaged burial ≤ 0.3). Conversely, 77.6% of the hydrophobic residues (viz., Ala, Val, Leu, Ile, Met, Phe, Tyr, Trp) were found to be completely exposed to the solvent (time-averaged burial > 0.30). This reversed trend in the hydrophobic burial in IDPs/IDPRs (with respect to folded globules) may contribute to their high solution-reactivity, binding promiscuity [2] and the high entropic cost required to stabilize them [58]. All these features are in accordance with the 'coupled binding and folding' model [59–61] (also known as 'binding induced folding' [56]) of the IDPs/IDPRs.

Impact of salt-bridges on the overall electrostatic balance in IDPs

Though atomic packing (or steric overlap) has been studied in IDPs, it is more relevant in the context of IDPRs [37] and has been assessed in terms of point atomic contacts, rather than surface fits as in globular proteins [30, 40]. Point atomic contacts are naturally limited in portraying the true picture of atomic packing due to their inherent inability to depict the geometric match between the interacting side-chains [30]. As the unstructured regions are overwhelmingly prevalent, packing does not seem to be the major interaction required to maintain the dynamic flexibility of IDPs/IDPRs. IDPs/IDPRs does not allow hydrophobic collapse like that of folded globules. Rather, given the predominance of charged and polar residues, the global electrostatic balance is expected to be a major determinant in the formation of salt-bridges. Therefore, it is necessary to investigate the role of salt-bridges in the global electrostatic balance of IDPs. In this regard, the Complementarity Plot (CP) [41, 43], an established structure validation tool originally proposed for globular proteins was used. CP analyzes the combined effect of packing and electrostatics together for amino acid side-chains embedded in proteins. This tool is built with the provision of analyzing the individual components, i.e., packing and electrostatics independently. The primary goal of this study was to analyze the role of salt-bridges in the context of the global electrostatic balance of IDPs/IDPRs. Thus the observation window was based on the calculated electrostatic complementarity (E_m) of the amino acid residues of the disordered polypeptide chains along their MD trajectories. Hundred equidistant snapshots were collected at an interval of 100 ps to represent the dynamic ensemble of each protein. The software suite for CP (http://www.saha.ac.in/biop/www/sarama.html) returns three different measures of electrostatic complementarities, E_m^{all}, E_m^{sc} and E_m^{mc} (see **Materials and Methods**) computed on 'all', 'side-chain' and 'main-chain' van der Waal's dot surface points of the 'target' amino acid residue respectively. The original plots in CP uses E_m^{sc} in accordance to the corresponding shape complementarity measure S_m^{sc}. For the current purpose, we were particularly interested in E_m^{all} since the overall impact of the salt-bridges on the full residue surfaces was of prime interest. The log odd probability score ($P(E_m)$) (see **Materials and Methods**) summed over amino acid residues distributed in different burial bins of the entire poly-peptide chains gives an estimate of its overall electrostatic balance. Only buried and partially buried residues that are present in the first three burial bins were considered in the original study [41], however, for this work, the fourth burial bin corresponding to the solvent exposed residues (*bur* > 0.30) were also considered (see **Materials and Methods**). This fourth burial bin contains the largest fraction of the amino acid residues (~80% or more) consistently throughout the entire MD trajectory in all proteins (**Figure 10**).

The 'complementarity' calculation was repeated twice: (i) with all atoms of the protein contributing to the electrostatic complementarity (set 1) and (ii) with salt-bridges being neutralized (set 2). For

set 2, salt-bridges were identified *in situ* in each MD snapshot, and, the side-chains involved in the salt-bridges were assigned partial charges of the corresponding uncharged amino acids. This methodology was in accordance with the one used earlier to quantify the relative contribution of salt-bridges to electrostatic complementarity (EC) at protein-protein interfaces [62].

Mild variations were noted in the three P_{Em} measures calculated on main-chain, side-chain and the van der Waal's surfaces of the whole residue. This indicates that the electrostatic balance is evenly retained at different surface patches throughout the entire residue surface without having any bias towards a particular patch (say, the side-chain dot surface points). For all three P_{Em} measures ($P(E_m^{all})$, $P(E_m^{sc})$, $P(E_m^{mc})$), a consistent trend (**Figure 12**, **Supplementary Figure S6, S7**) was observed throughout the MD simulation trajectories where the scores in set 2 were always lesser than or equal to their corresponding scores in set 1. This characteristic trend was found for all four proteins and in no case an exception was observed. The impact was more pronounced in IDPRs rather than IDPs (**Figure 12**), which is in accordance with the prevalence of persistent salt-bridges in the former. It was also interesting to find out that the time evolved all-atom-$P(E_m)$ values (set 1) for all four proteins were consistently above the suggested empirical threshold of -1.8 (standardized for globular proteins [41]) throughout the entire MD trajectory. For proteins in general, a $P(E_m)$ value falling below this threshold corresponds to structures with unbalanced electric fields throughout its van der Waal's surface representing a state of electrostatic dissonance caused by sub-optimal distributions of atomic partial charges in the molecule. The fact that $P(E_m)$ for the IDPs/IDPRs were consistently above this empirical threshold throughout the trajectory (**Figure 12**, **Supplementary Figure S6, S7**) reveals the meticulous balance of electric fields coming from different parts of the chains in their respective time-evolved confomational ensembles. In other words, IDPs/IDPRs seem to maintain their global electrostatic balance with time while continuously fluctuating between different conformations. Although, there was an average of ~5% reduction in the $P(E_m)$ values when salt-bridges were neutralized (**Supplementary Figure S8**), noticeably, the reduced values did not drop below the $P(E_m)$ threshold of -1.8. This means that even without the contribution of the salt-bridges, IDPs/IDPRs manage to retain their global electrostatic balance. However, the ionic bonds unequivocally contribute to the overall balance. Similar results were also obtained at protein-protein interfaces [62] and within folded globular proteins [17] where the electrostatic complementarity was augmented by the contribution of salt-bridges, though it did not reduce drastically even after neutralizing the salt-bridges.

Correlated movements of oppositely charged residues forming persistent salt-bridges

Dynamic Cross Correlation Maps (see **Materials and Methods**) were constructed to examine the correlated movements between the oppositely charged residues forming persistent salt-bridges. In consistency with all other calculations, a persistence cut-off of 25% was considered. Traditionally DCCM has been used to portray protein-domain movements between dimers involving whole main-chain trajectories [63, 64]. In contrast, our focus was on the dynamic coordination of individual pair of (charged) residues. A majority of the (i,j) pairs recorded high DCC(i,j) values close to 1 (**Table 7**) revealing dynamically sustained correlated movements between them. In more detail, 73% and 70% of (i,j) pairs gave rise to a DCC(i,j) value greater than 0.70 in main- and side-chain atoms respectively (**Figure 13, Supplementary Figure S9**). These correlated pairs spanned across ionic bonds of different contact orders, viz., short range contacts (e.g., 3-GLU ~ 5-ARG in β-amyl: **Table 7**), middle range (e.g., 10-ARG ~ 27-GLU in 1F0R) and long range contacts (2-GLU ~ 108-ARG in 1CD3) with persistence values ranging from 0.257 (viz., 20-GLU ~ 58-LYS in α-syn) to 0.963 (viz., 98-ASP ~ 100-ARG in 1CD3). Only weak correlations were obtained between DCC and persistence (0.33, 0.40 for main- and side-chain DCC respectively) and also between DCC and contact order (-0.40, -0.35). The negative sign (anti-correlation) in the later case indicates that short-range contacts

generally have a tendency to show a higher degree of correlated movement – which may be anticipated intuitively. In fact, in 85% (main-chain) and 80% (side-chain) of the cases involving short-range contacts (say, CO ≤ 0.2), a DCC(i,j) value of ≥ 0.70 was observed. However, as pointed out earlier, long and middle range contacts also attributed to high DCC, while, conversely, low / moderate DCC values were also obtained, though rarely for reasonably short-range contacts (e.g., 108-ARG ~ 133-GLU, 82-LYS ~ 97-GLU in 1F0R). It must be noted that 92% of all pairs exceeded a DCC(i,j) value of 0.7 (83% exceeded 0.8) out of the pull of all salt-bridges that have a persistence of 0.5 (i.e., the salt-bridge occurred half the time) or more. Taking into account both the above criteria (i.e., CO ≤ 0.2 and *pers* ≥ 0.5 simultaneously) 100% of all residue pairs forming persistent short-range ionic bonds gave rise to a DCC of 0.70 or more (in both main- and side-chain). Taking all these together, it appears that short-range and persistent ionic bonds generally have a higher tendency to dynamically sustain correlated movements, though, the correlation may not be restricted to any particular sequence separation (i.e., contact order) and/or dynamic spatial separation (i.e., persistence) of the pair of oppositely charged residues. The DCC(i,j) values computed for main-chain and side-chain atomic units (see **Materials and Methods**) were also in good agreement, giving a pairwise Pearson's correlation of 0.97 over thirty persistent salt-bridges (P-value < 0.00001) occuring in all four proteins. This means that the correlated movements generally occur throughout both residues and not restricted to a particular atomic segment. Noticeably, there was only a single instance of weak anti-correlation obtained in the DCC values (77-ARG ~ 105-GLU in 1CD3) continuously throughout both in main- and side-chain atoms.

Conclusion

The main objective of the current study was to understand the basis of a meticulous trade-off between accommodating oppositely charged residues and retaining the desired dynamic flexibility in IDPs/IDPRs. Accordingly, the IDP sequences were chosen such that there are enough oppositely charged residues. The results of this study confirm that the key to maintain such a trade-off lies in the continuous time evolution of salt-bridges with a wide range of persistence. The salt-bridges in IDPs/IDPRs span across the full scale (0 to 1) of persistence values, ranging from extremely short-lived ionic bonds to those that persist consistently throughout the entire simulation trajectory. Moreover, for the transient ionic bonds, a high degree of promiscuity is exhibited by charged groups (say, –NH3+), in terms of pairing to different oppositely charged groups (say, –COO-) originating from a variety of amino-acid residues. Study of the contact order in salt-bridges also revealed the occurrence of short-range, middle-range as well as long-range contacts with no distinct correlation with their corresponding persistence values. In other words, the salt-bridges composed of either closely spaced residues or residues that are distant in the sequence space may correspond to high persistence (say, *pers* ≥ 0.8). All these parameters collectively portray a model of dynamic flexibility which was also confirmed visually. Such a model can accommodate oppositely charged residues by allowing the formation of salt-bridges with varying degree of persistence and different ranges of contact orders, and, thereby, map structurally to an ensemble of distant conformations rather than minimally fluctuating about a single global-minima structure as in globular proteins. The ionic bond interactions also appear to be pivotal for the dynamic bending in the IDPs/IDPRs and contribute substantially to their global electrostatic balance. Furthermore, the analyses of hydrophobic burial profiles present a simple and effective single-parameter-tool, the *rGb* score to characterize the potential instability of IDPs/IDPRs in solution which may account for their high reactivity. Interestingly, the charged residues involved in persistent salt-bridges appear to move in a correlated manner throughout the entire dynamic trajectory. This trend was more prominent for short-range contacts with higher persistence. In addition, the study also identified and characterized the dynamics of the composite salt-bridge motifs in the conformational ensembles of the IDPs/IDPRs which mapped the secondary structural transitions to the salt-bridge dynamics. This

work thus constitutes an important knowledge-base in protein design. So to speak, the results and conclusions of the current study should provoke future studies to explore the disordered-globular interface in proteins and their evolutionary relationship based on salt-bridge design, leading to a mutually reversible switch.

Acknowledgment and Funding

The work was supported by the Department of Science and Technology – Science and Engineering Research Board (DST-SERB research grant **PDF/2015/001079/LS).**

Figure Legends

Figure 1. Salt-bridge motifs found in simulated conformational ensembles in IDPs. Each motif (network) is represented as a ball (node: charged residue) and stick (edge: ionic bond) model and tagged by the numeric motif identifier (see **Materials and Methods**). The numbers in parentheses stand for the cumulative counts of the corresponding salt-bridge motif in the whole 100 ns trajectories for 1CD3, 1F0R, α-syn and β-amyl respectively.

Figure 2. Frequency of the isolated salt-bridge motif in IDPs along time. The number of isolated salt-bridges formed in each snapshot (MD frame) is plotted as a time series for the four proteins: 1CD3, 1F0R, α-syn and β-amyl. The MD frames sampled at a consistent interval of 50 ps enabling 2000 such frames to cover the whole of 100 ns trajectory.

Figure 3. Frequency of the bifurcated salt-bridge motif in IDPs along time. The number of bifurcated salt-bridges formed in each snapshot (MD frame) is plotted as a time series for the four proteins: 1CD3, 1F0R, α-syn and β-amyl. Sampling of the MD frames is identical to the description given in **Figure 2**.

Figure 4. The closed quadruplet salt-bridge motif in IDPs. The 4-cycle motifs encountered in (A) 1F0R (B) α-syn within 75-100 ns MD frames. The ionic bonds have been displayed as yellow dashed lines between residues represented as ball and stick models on top of the corresponding main-chains portrayed as cartoon. The residues forming the closed quadruplets are 38-Lys, 133-Glu, 21-Glu, 108-Arg (1F0R) and 21-Lys, 20-Glu, 57-Glu, 58-Lys (α-syn).

Figure 5. Dynamic persistence of salt-bridges in IDPs. (A) Normalized frequency distribution of persistence values (*pers)* computed for all salt-bridges in all four IDPs/IDPRs. The frequency bars corresponding to salt-bridges having a persistence value greater than equal to 0.25 have been drwan in yellow while those with *pers* < 0.25 drawn in blue. (B) The number of persistent salt-bridges (*pers* ≥ 0.25) plotted as coloured bars against their corresponding persistence values for the four IDPs/IDPRs, using different colours as given in the index-box embedded in the figure.

Figure 6. Example of multiple ionic bonds formed by the same charged residue at different temporal patches in IDPs. The figure portrayers the different transient ionic bonds (displayed as the yellow dashed lines) involving the amino acid 28-Lys and different oppositely charged residues (1-Asp, 3-Glu, 22-Glu, 23-Asp) at various temporal patches (see **Main Text**) in the MD trajectory of β-amyl. The concerned charged residues have been drawn as ball and sticks on the background of

the corresponding main-chains, drawn in cartoon with different colours. The different main-chain conformations were superposed on a common template prior to their display.

Figure 7. Persistent salt-bridges mapped on the time-evolved IDP structures. All the persistent salt-bridges (*pers* ≥ 0.25) for each IDP, portrayed as yellow dashed lines between corresponding charged residues displayed as ball and sticks, have been mapped on top of a representative snapshot chosen randomly from the MD trajectory.

Figure 8. Example of a 'disorder-to-order' secondary structural transition in IDPs, as revealed in beta-amyl. The figure highlights (B) the short-helix (coloured magenta) formed during the MD-simulation run in β-amyl (see **Main-Text**) and (A) the corresponding segment (same color) found as a disordered coil in the starting atomic model used as the initial template in the MD. Charged residues have been portrayed as ball and sticks in both.

Figure 9. Time-evolved shape factor profiles of IDPs with respect to the globular protein reference range. Shape factor values for each individual IDP plotted against simulation time. As a reference, the mean shape factor value (μ) calculated on globular proteins (ρ_{glob}) is displayed as the continuous horizontal line drawn at x~1.0 ranged by the two dashed lines above and below mapping to (μ+σ) and (μ-σ) respectively, where σ stands for the standard deviation in ρ_{glob}.

Figure 10. Time-evolved burial profiles in IDPs/IDPRs. Fractional content of amino acid residues (frac bur) falling in the four burial bins (bur1-bur4: see **Main Text**) plotted against simulation time for the four IDPs/IDPRs.

Figure 11. Time-averaged accessibility scores raised by the IDPs with respect to the globular protein reference range. The accessibility score (*rGb*) averaged over simulation time plotted as thick magenta bars and the corresponding standard deviations plotted as error bars individually for the four IDPs. The mean and standard deviation for the same parameter calculated on a data-base of globular proteins is also displayed in a similar manner as a reference.

Figure 12. Impact of salt-bridges on the overall electrostatic balance in IDPs as revealed by the time-evolved $P(E_m)$ score. The log-odd probability score $P(E_m)$ computed on the distribution of electrostatic complimentarity (E_m) raised by amino acid residues as a function of their burial plotted against simulation time for the four IDPs. This particular figure portrays the profiles for $P(E_m^{all})$computed on the whole-residue van der Waal's surface. The Prussian blue curves represent the $P(E_m^{all})$ calculated considering the contribution of all atoms while the orange-red dashed curves represent those calculated when the salt-bridges were neutralized (see **Main Text**).

Figure 13. Dynamic Cross Correlation Map (DCCM) composed of the pair of oppositely charged residues forming persistent salt-bridges alone (calculated on their main-chain atoms). Entries apart from the relevant ones (i.e., the salt-bridge forming pair of oppositely charged residues) in the matrices are zero-padded. Amino acid residues are represented (along the horizontal and vertical axes) by their residue sequence numbers (in an ascending order) concatenated with their one letter abbreviation codes. Values are plotted according to the color-bars in each sub-plot. Note that only 1CD3 has a negative entry and thus the corresponding color-bar is different than in the other three.

Tables

Table 1. IDPs/IDPRs used in the Molecular Dynamic study.

Abbreviated Name	Sequence Length (amino acids)	Source Database	Location of the disordered region	Fraction Disordered (%)	Fraction Charged (%)	Fraction positively Charged (%)	Fraction negatively Charged (%)
1CD3	120	PDB	middle	43.3	33.4	16.7	16.7
1F0R	134	PDB	N-terminus	61.2	34.3	13.4	20.9
α-syn	140	DISPROT	Whole sequence	100	28.6	11.4	17.2
β-amyl	42	DISPROT	Whole sequence	100	28.6	14.3	14.3

Table 2. Statistical Weights for different secondary structural classes calculated on the globular protein database, GDB. The secondary structural classes categorized into two main disorder class (see **Main Text**).

Class	Secondary Structure	Weight
Order	α-Helix	0.33255
	β-Strand	0.21848
	3_{10}-Helix	0.04424
	Π-Helix	0.00016
Disorder	Turn	0.20834
	Coil	0.18380
	Bridge	0.01242

Table 3. A checker board representation of the scheme adapted to score salt-bridge mediated disorder-to-order transitions. The 'pair' refers to the pair of charged residues forming the salt-bridge. CDS refers to Combined Disordered Status (see **Main Text**)

CDS of the pair at time **t** ↓	CDS of the pair at time **t+Δt** →			
	DD	OO	OD	DO
DD	0	1	0.5	0.5
OO	1	0	0.5	0.5
OD	0.5	0	0.5	0.25
DO	0.5	0.5	0.25	0

Table 4. Persistence values (*pers*) and Contact Orders (CO) of the persistent salt-bridges in the four IDPs/IDPRs.

Salt-bridge		***pers***	**CO**
1CD3			
98-ASP	100-ARG	0.963	0.017

2-GLU	108-ARG	0.819	0.883
64-ARG	68-GLU	0.797	0.033
26-ARG	28-GLU	0.580	0.017
15-GLU	50-ARG	0.536	0.292
22-GLU	76-ARG	0.449	0.450
22-GLU	64-ARG	0.368	0.350
61-ARG	65-ASP	0.322	0.033
27-ASP	77-ARG	0.250	0.417
77-ARG	105-GLU	0.250	0.233
1F0R			
90-ASP	117-LYS	0.818	0.202
28-ASP	31-LYS	0.803	0.022
10-ARG	27-GLU	0.726	0.127
10-ARG	30-ASP	0.643	0.149
14-GLU	23-ARG	0.448	0.067
21-GLU	38-LYS	0.430	0.127
82-LYS	97-GLU	0.401	0.112
108-ARG	133-GLU	0.378	0.187
15-GLU	23-ARG	0.335	0.060
41-ASP	55-LYS	0.321	0.105
96-HIP	98-GLU	0.311	0.015
58-ASP	108-ARG	0.285	0.373
62-GLU	81-ARG	0.267	0.142
α-syn			
96-LYS	104-GLU	0.937	0.057
21-LYS	57-GLU	0.509	0.257
23-LYS	28-GLU	0.398	0.036
28-GLU	34-LYS	0.379	0.043
20-GLU	58-LYS	0.257	0.271
β-amyl			
3-GLU	5-ARG	0.878	0.048
11-GLU	13-HIP	0.266	0.048

Table 5. Normalized Cross Correlation (CC) of the global secondary structural transition scores (gsst's) with number of salt-bridges (N_{sbg}). Along with the maximum Cross Correlation (CC_{max}), zero lag Cross Correlation values (CC_{zero_lag}) are also tabulated while the lag difference of CC_{max} with respect to Cc_{zero_lag} is given as 'Lag at CC_{max}'.

Transition	**CC measures**	**1CD3**	**1F0R**	**α-syn**	**β-amyl**

Scores					
gsst1	CC_{zero_lag}	412.60	454.95	385.63	317.88
	CC_{max}	413.75	457.11	388.81	319.42
	Lag at CC_{max}	3	-5	6	-17
gsst2	CC_{zero_lag}	430.04	455.91	386.21	317.85
	CC_{max}	431.11	457.79	388.36	317.93
	Lag at CC_{max}	2	-5	6	-3

Table 6. Conservation Score (CS_{sec}) for local secondary structural transition mediated by the persistent salt-bridges in the four IDPs/IDPRs. The corresponding Predominant Secondary Structural Motif (PSSM) and their relative occupancy (O_{PSSM}) in the whole trajectory are also tabulated.

Salt-bridge		***pers***	**CS_{sec}**	**PSSM**	**O_{PSSM}**
1CD3					
98-ASP	100-ARG	0.963	0.00050	Turn~Turn	0.861
2-GLU	108-ARG	0.819	0.79502	Coil~β-Strand	0.792
64-ARG	68-GLU	0.797	0.46561	α-Helix~α-Helix	0.346
26-ARG	28-GLU	0.580	0.08942	Coil~Coil	0.447
15-GLU	50-ARG	0.536	0.00100	Coil~Turn	0.333
22-GLU	76-ARG	0.449	0.07816	Coil~Turn	0.503
22-GLU	64-ARG	0.368	0.77351	Coil~α-Helix	0.538
61-ARG	65-ASP	0.322	0.35218	α-Helix~α-Helix	0.430
27-ASP	77-ARG	0.250	0.34692	Coil~Turn	0.341
77-ARG	105-GLU	0.250	0.02726	Turn~Coil	0.322
1F0R					
90-ASP	117-LYS	0.818	0.49099	3_{10}-Helix~Turn	0.453
28-ASP	31-LYS	0.803	0.26114	Turn~Turn	0.735
10-ARG	27-GLU	0.726	0.24912	Coil~Turn	0.555
10-ARG	30-ASP	0.643	0.13045	Coil~Turn	0.600
14-GLU	23-ARG	0.448	0.02203	Turn~Turn	0.770
21-GLU	38-LYS	0.430	0.00651	Turn~Coil	0.806
82-LYS	97-GLU	0.401	0.99449	Turn~β-Strand	0.741
108-ARG	133-GLU	0.378	0.00501	Turn~Coil	0.767
15-GLU	23-ARG	0.335	0.02203	Turn~Turn	0.908
41-ASP	55-LYS	0.321	0.00100	Turn~Coil	0.523

96-HIE	98-GLU	0.311	0.01502	β-Strand~β-Strand	0.983
58-ASP	108-ARG	0.285	0.01502	Turn~Turn	0.568
62-GLU	81-ARG	0.267	0.38107	Turn~Coil	0.485
α-syn					
96-LYS	104-GLU	0.937	0.00100	Turn~Turn	0.544
21-LYS	57-GLU	0.509	0.05558	Turn~Coil	0.666
23-LYS	28-GLU	0.398	0.00250	Turn~Turn	0.565
28-GLU	34-LYS	0.379	0.76114	Turn~3_{10}-Helix	0.754
20-GLU	58-LYS	0.257	0.04907	Turn~Coil	0.675
β-amyl					
3-GLU	5-ARG	0.878	0.07468	Coil~Coil	0.423
11-GLU	13-HIE	0.266	0.17135	Coil~Turn	0.698

Table 7. Dynamic Cross Correlation (DCC) of persistent salt-bridge forming alike charged residues. DCC_{mc} and DCC_{sc} refers to the DCC values computed for main- and side-chain atoms respectively.

Salt-bridge		***pers***	**DCC_{mc}**	**DCC_{sc}**
1CD3				
98-ASP	100-ARG	0.963	0.800	0.932
2-GLU	108-ARG	0.819	0.761	0.797
64-ARG	68-GLU	0.797	0.927	0.900
26-ARG	28-GLU	0.580	0.978	0.948
15-GLU	50-ARG	0.536	0.831	0.913
22-GLU	76-ARG	0.449	0.271	0.338
22-GLU	64-ARG	0.368	0.184	0.039
61-ARG	65-ASP	0.322	0.915	0.737
27-ASP	77-ARG	0.250	0.105	0.316
77-ARG	105-GLU	0.250	-0.156	-0.242
1F0R				
90-ASP	117-LYS	0.818	0.518	0.528
28-ASP	31-LYS	0.803	0.987	0.994
10-ARG	27-GLU	0.726	0.971	0.972
10-ARG	30-ASP	0.643	0.946	0.961
14-GLU	23-ARG	0.448	0.968	0.981
21-GLU	38-LYS	0.430	0.919	0.920
82-LYS	97-GLU	0.401	0.457	0.577
108-ARG	133-GLU	0.378	0.347	0.418
15-GLU	23-ARG	0.335	0.971	0.981
41-ASP	55-LYS	0.321	0.828	0.826

96-HIE	98-GLU	0.311	0.812	0.653
58-ASP	108-ARG	0.285	0.941	0.833
62-GLU	81-ARG	0.267	0.372	0.339
α-syn				
96-LYS	104-GLU	0.937	0.965	0.986
21-LYS	57-GLU	0.509	0.931	0.951
23-LYS	28-GLU	0.398	0.952	0.950
28-GLU	34-LYS	0.379	0.974	0.969
20-GLU	58-LYS	0.257	0.940	0.924
β-amyl				
3-GLU	5-ARG	0.878	0.965	0.965
11-GLU	13-HIP	0.266	0.950	0.943

References

1. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331 . doi: 10.1006/jmbi.1999.3110

2. Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta 1834:932–951 . doi: 10.1016/j.bbapap.2012.12.008

3. Nath A, Sammalkorpi M, DeWitt DC, et al (2012) The Conformational Ensembles of α-Synuclein and Tau: Combining Single-Molecule FRET and Simulations. Biophys J 103:1940–1949 . doi: 10.1016/j.bpj.2012.09.032

4. Cheng S, Cetinkaya M, Gräter F (2010) How Sequence Determines Elasticity of Disordered Proteins. Biophys J 99:3863–3869 . doi: 10.1016/j.bpj.2010.10.011

5. Baruah A, Rani P, Biswas P (2015) Conformational Entropy of Intrinsically Disordered Proteins from Amino Acid Triads. Sci Rep 5: . doi: 10.1038/srep11740

6. Rani P, Biswas P (2015) Local Structure and Dynamics of Hydration Water in Intrinsically Disordered Proteins. J Phys Chem B 119:10858–10867 . doi: 10.1021/jp511961c

7. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246 . doi: 10.1146/annurev.biophys.37.032807.125924

8. 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

9. 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

10. Gutin AM, Abkevich VI, Shakhnovich EI (1995) Is burst hydrophobic collapse necessary for protein folding? Biochemistry (Mosc) 34:3066–3076

11. Zhou R, Huang X, Margulis CJ, Berne BJ (2004) Hydrophobic collapse in multidomain protein folding. Science 305:1605–1609 . doi: 10.1126/science.1101176

12. Baruah A, Biswas P (2016) Globular–disorder transition in proteins: a compromise between hydrophobic and electrostatic interactions? Phys Chem Chem Phys 18:23207–23214 . doi: 10.1039/C6CP03185D

13. Wong ETC (2012) Electrostatics in intrinsically disordered proteins. University of British Columbia

14. Liu C, Wang T, Bai Y, Wang J (2017) Electrostatic forces govern the binding mechanism of intrinsically disordered histone chaperones. PLOS ONE 12:e0178405 . doi: 10.1371/journal.pone.0178405

15. Xue Y, Yuwen T, Zhu F, Skrynnikov NR (2014) Role of electrostatic interactions in binding of peptides and intrinsically disordered proteins to their folded targets. 1. NMR and MD characterization of the complex between the c-Crk N-SH3 domain and the peptide Sos. Biochemistry (Mosc) 53:6473–6495 . doi: 10.1021/bi500904f

16. Borg M, Mittag T, Pawson T, et al (2007) Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. Proc Natl Acad Sci 104:9650–9655 . doi: 10.1073/pnas.0702580104

17. Basu S, Mukharjee D (2017) Salt-bridge networks within globular and disordered proteins: characterizing trends for designable interactions. J Mol Model 23:206 . doi: 10.1007/s00894-017-3376-y

18. 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

19. Betz SF (1993) Disulfide bonds and the stability of globular proteins. Protein Sci Publ Protein Soc 2:1551–1558

20. Thornton JM (1981) Disulphide bridges in globular proteins. J Mol Biol 151:261–287 . doi: 10.1016/0022-2836(81)90515-5

21. Sickmeier M, Hamilton JA, LeGall T, et al (2007) DisProt: the Database of Disordered Proteins. Nucleic Acids Res 35:D786–D793 . doi: 10.1093/nar/gkl893

22. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative Protein Structure Modeling Using Modeller. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al 0 5:Unit-5.6 . doi: 10.1002/0471250953.bi0506s15

23. The Amber Molecular Dynamics Package. http://ambermd.org/. Accessed 16 Sep 2017

24. Hornak V, Abel R, Okur A, et al (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725 . doi: 10.1002/prot.21123

25. Best RB, Hummer G (2009) Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. J Phys Chem B 113:9004–9015 . doi: 10.1021/jp901540t

26. Jorgensen WL, Chandrasekhar J, Madura JD, et al (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935 . doi: 10.1063/1.445869

27. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242 . doi: 10.1093/nar/28.1.235

28. Suzek BE, Huang H, McGarvey P, et al (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23:1282–1288 . doi: 10.1093/bioinformatics/btm098

29. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res 32:W500–W502 . doi: 10.1093/nar/gkh429

30. 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

31. Kobayashi Y (2016) Statistical mechanics of protein structural transitions: Insights from the island model. Biophys Physicobiology 13:251–262 . doi: 10.2142/biophysico.13.0_251

32. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277:985–994 . doi: 10.1006/jmbi.1998.1645

33. Teraoka I (2002) Polymer Solutions: An Introduction to Physical Properties. John Wiley & Sons

34. Burchard W (1999) Solution Properties of Branched Macromolecules. In: Branched Polymers II. Springer, Berlin, Heidelberg, pp 113–194

35. Burchard W, Schmidt M, Stockmayer WH (1980) Information on Polydispersity and Branching from Combined Quasi-Elastic and Intergrated Scattering. Macromolecules 13:1265–1272 . doi: 10.1021/ma60077a045

36. (1983) Light Scattering from Polymers. Springer Berlin Heidelberg, Berlin, Heidelberg

37. Rawat N, Biswas P (2012) Hydrophobic Moments, Shape, and Packing in Disordered Proteins. J Phys Chem B 116:6326–6335 . doi: 10.1021/jp3016529

38. Hubbard S, Thornton J (1993) NACCESS. Computer Program, Department of Biochemistry and Molecular Biology, University College London. - Open Access Library. http://www.oalib.com/references/5299711. Accessed 1 Mar 2017

39. Basu S, Wallner B (2016) Finding correct protein–protein docking models using ProQDock. Bioinformatics 32:i262–i270 . doi: 10.1093/bioinformatics/btw257

40. 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

41. 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

42. 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

43. 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

44. Swaminathan S, Harte WE, Beveridge DL (1991) Investigation of domain structure in protein via molecular dynamics simulation: Application to HIV-1 protease dimer. J Am Chem Soc 113:2717–2721

45. Choi UB, McCann JJ, Weninger KR, Bowen ME (2011) Beyond the Random Coil: Stochastic Conformational Switching in Intrinsically Disordered Proteins. Struct Lond Engl 1993 19:566–576 . doi: 10.1016/j.str.2011.01.011

46. 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

47. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38, 27–28

48. Gupta SK, Kumar A (1987) Nonlinear Step Growth Polymerization. In: Reaction Engineering of Step Growth Polymerization. Springer, Boston, MA, pp 93–128

49. Basharov MA (2000) Cotranslational Folding of Proteins. Biochem Mosc 65:1380–1384 . doi: 10.1023/A:1002800822475

50. Fedorov AN, Baldwin TO (1997) Cotranslational Protein Folding. J Biol Chem 272:32715–32718 . doi: 10.1074/jbc.272.52.32715

51. Makhoul CH, Trifonov EN (2002) Distribution of Rare Triplets Along mRNA and Their Relation to Protein Folding. J Biomol Struct Dyn 20:413–420 . doi: 10.1080/07391102.2002.10506859

52. Rai GJ, Kumar A, Biswas P (2015) Dynamics of dendrimers with excluded volume: A comparison with experiments and simulations. J Rheol 60:111–120 . doi: 10.1122/1.4937378

53. Richards FM (1974) The interpretation of protein structures: Total volume, group volume distributions and packing density. J Mol Biol 82:1–14 . doi: 10.1016/0022-2836(74)90570-1

54. Sheffler W, Baker D (2009) RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein Sci Publ Protein Soc 18:229–239 . doi: 10.1002/pro.8

55. Word JM, Lovell SC, LaBean TH, et al (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms11Edited by J. Thornton. J Mol Biol 285:1711–1733 . doi: 10.1006/jmbi.1998.2400

56. Arya S, Mukhopadhyay S (2014) Ordered water within the collapsed globules of an amyloidogenic intrinsically disordered protein. J Phys Chem B 118:9191–9198 . doi: 10.1021/jp504076a

57. Dahiyat BI, Mayo SL (1997) Probing the role of packing specificity in protein design. Proc Natl Acad Sci 94:10172–10177

58. Law SM, Gagnon JK, Mapp AK, Brooks CL (2014) Prepaying the entropic cost for allosteric regulation in KIX. Proc Natl Acad Sci U S A 111:12067–12072 . doi: 10.1073/pnas.1405831111

59. Espinoza-Fonseca LM (2009) Thermodynamic Aspects of Coupled Binding and Folding of an Intrinsically Disordered Protein: A Computational Alanine Scanning Study. Biochemistry (Mosc) 48:11332–11334 . doi: 10.1021/bi901705z

60. Wang Y, Chu X, Longhi S, et al (2013) Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. Proc Natl Acad Sci 110:E3743–E3752 . doi: 10.1073/pnas.1308381110

61. Shammas SL, Crabtree MD, Dahal L, et al (2016) Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. J Biol Chem 291:6689–6695 . doi: 10.1074/jbc.R115.692715

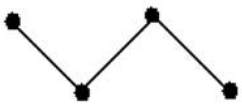
62. McCoy AJ, Chandana Epa V, Colman PM (1997) Electrostatic complementarity at protein/protein interfaces. J Mol Biol 268:570–584 . doi: 10.1006/jmbi.1997.0987

63. Agarwal PK, Geist A, Gorin A (2004) Protein dynamics and enzymatic catalysis: investigating the peptidyl-prolyl cis-trans isomerization activity of cyclophilin A. Biochemistry (Mosc) 43:10605–10618 . doi: 10.1021/bi0495228

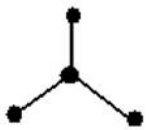
64. Kasahara K, Fukuda I, Nakamura H (2014) A Novel Approach of Dynamic Cross Correlation Analysis on Molecular Dynamics Simulations and Its Application to Ets1 Dimer–DNA Complex. PLOS ONE 9:e112419 . doi: 10.1371/journal.pone.0112419

11-11
(11617, 11213, 7360, 3152)

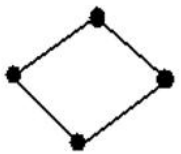
211-12-12
(1050, 2236, 736, 38)



221-221-12-12
(799, 224, 89, 0)

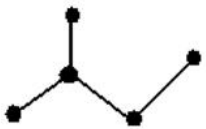


3111-13-13-13
(9, 250, 0, 0)

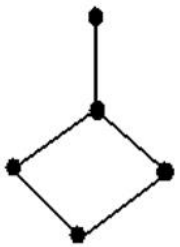


222-222-222-222
(191, 9, 2, 0)

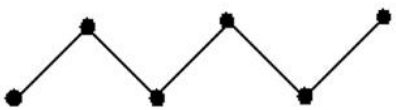
222-221-221-12-12
(26, 21, 1, 0)



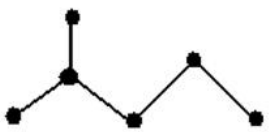
3211-231-13-13-12
(3, 27, 0, 0)



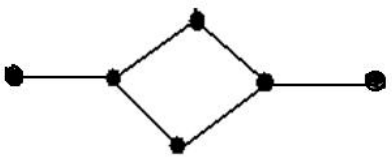
3221-232-232-222-13
(1, 2, 0, 0)



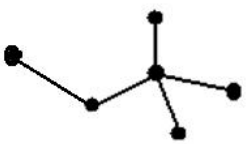
222-222-221-221-12-12
(1, 0, 0, 0)



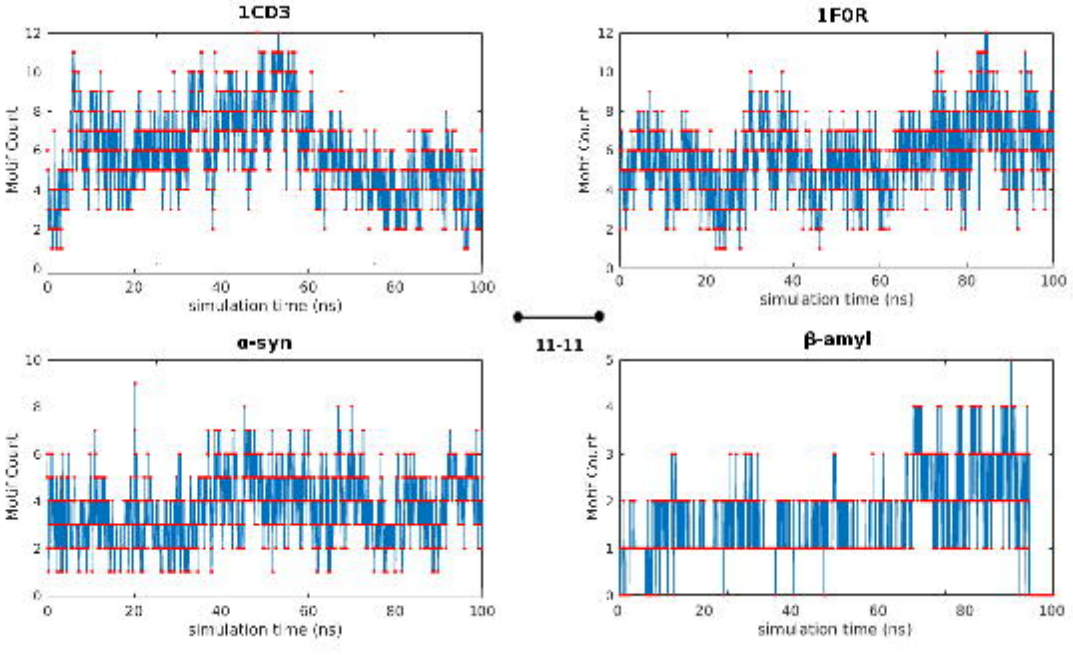
3211-232-221-13-13-12
(0, 2, 0, 0)

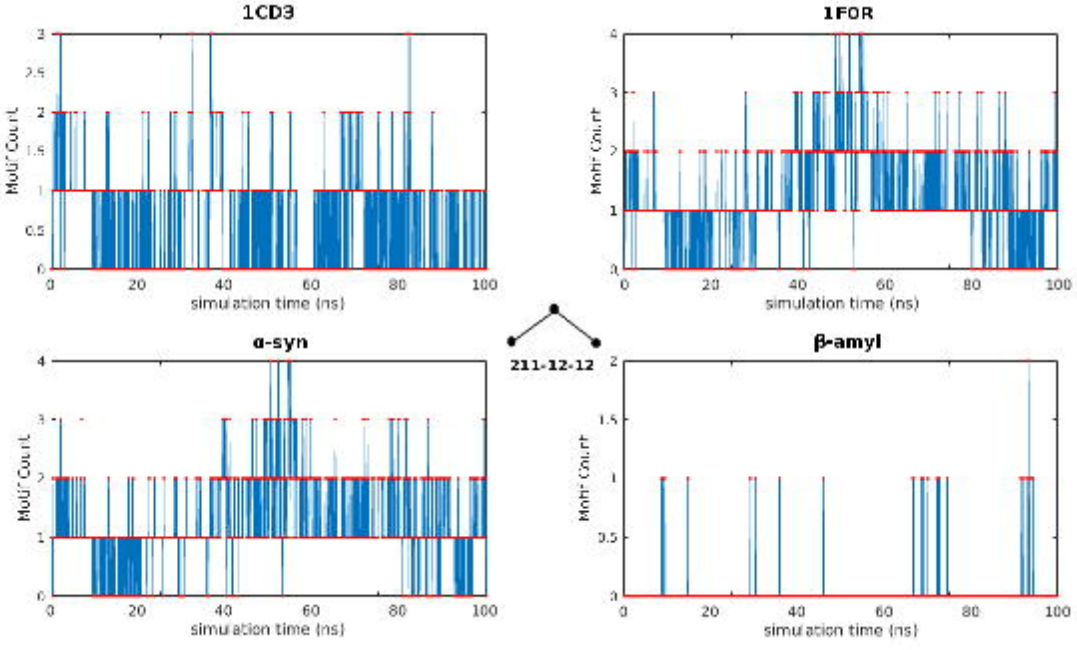


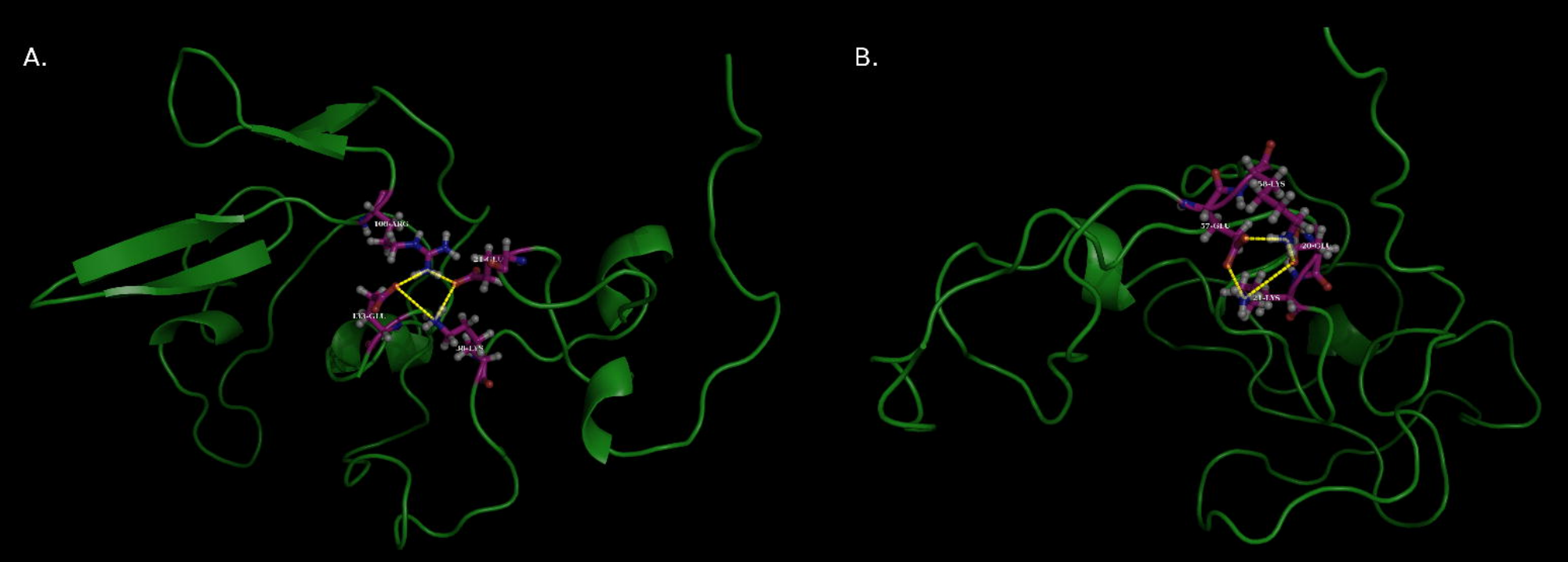
3221-3221-233-233-13-13
(0, 1, 0, 0)



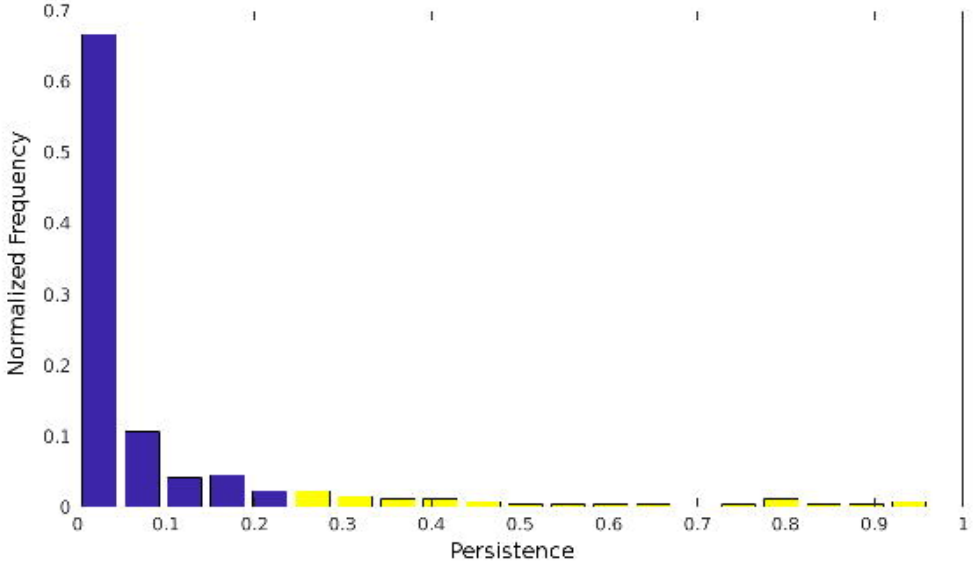
42111-241-14-14-14-12
(0, 2, 0, 0)


1CD3
1F0R
α-syn
11-11
β-amyl
Motif Count
simulation time (ns)

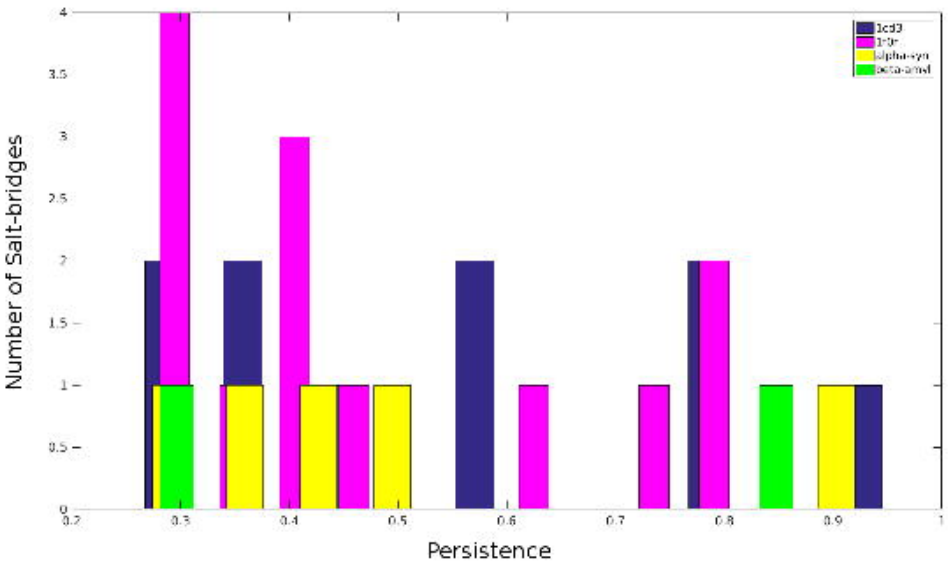

1CD3
1FOR
α-syn
β-amyl
211-12-12
simulation time (ns)

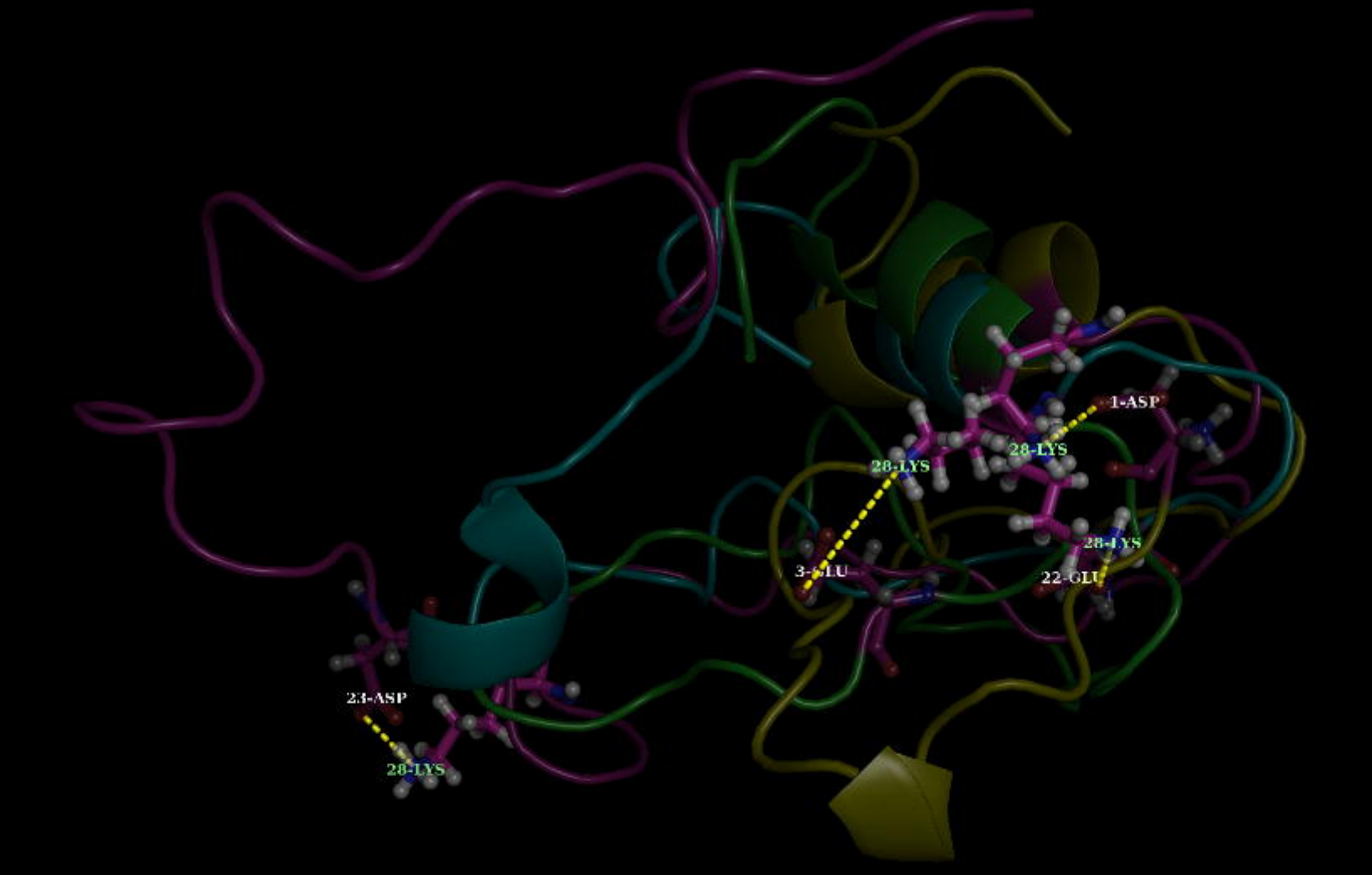

A.
100-ARG
214-LU
133-GLU
98-LYS
B.
58-LYS
57-GLU
20-GLU
21-LYS

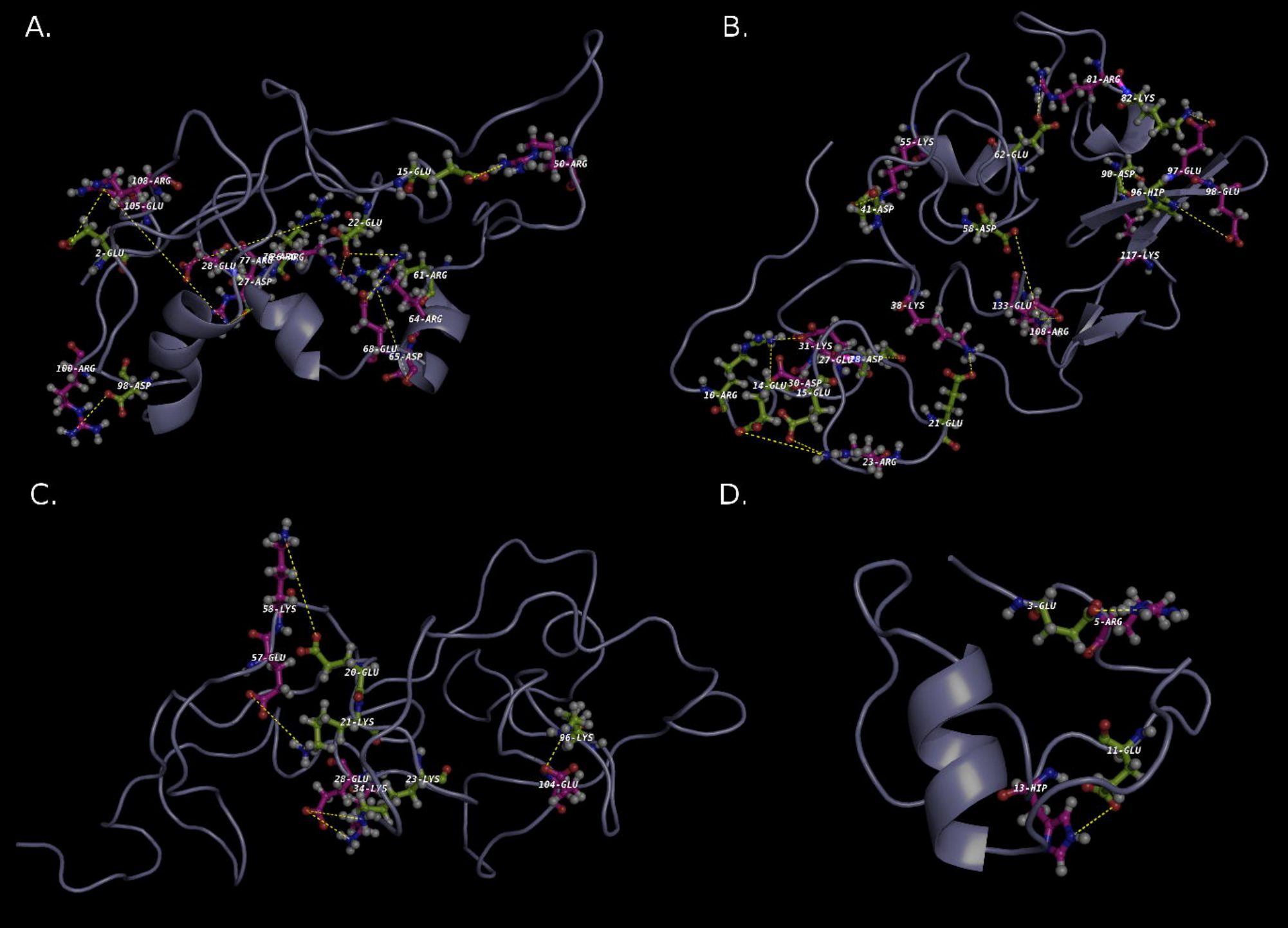
A.

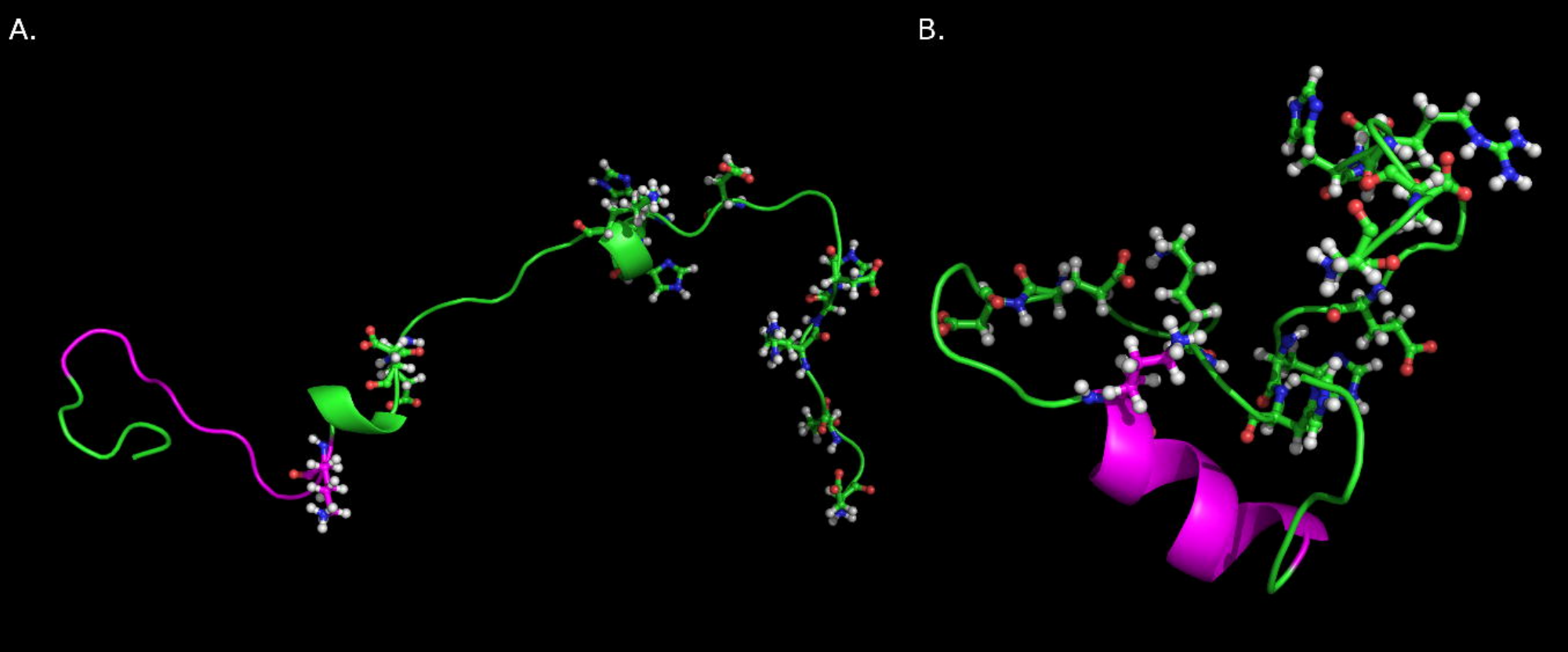

Normalized Frequency
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Persistence

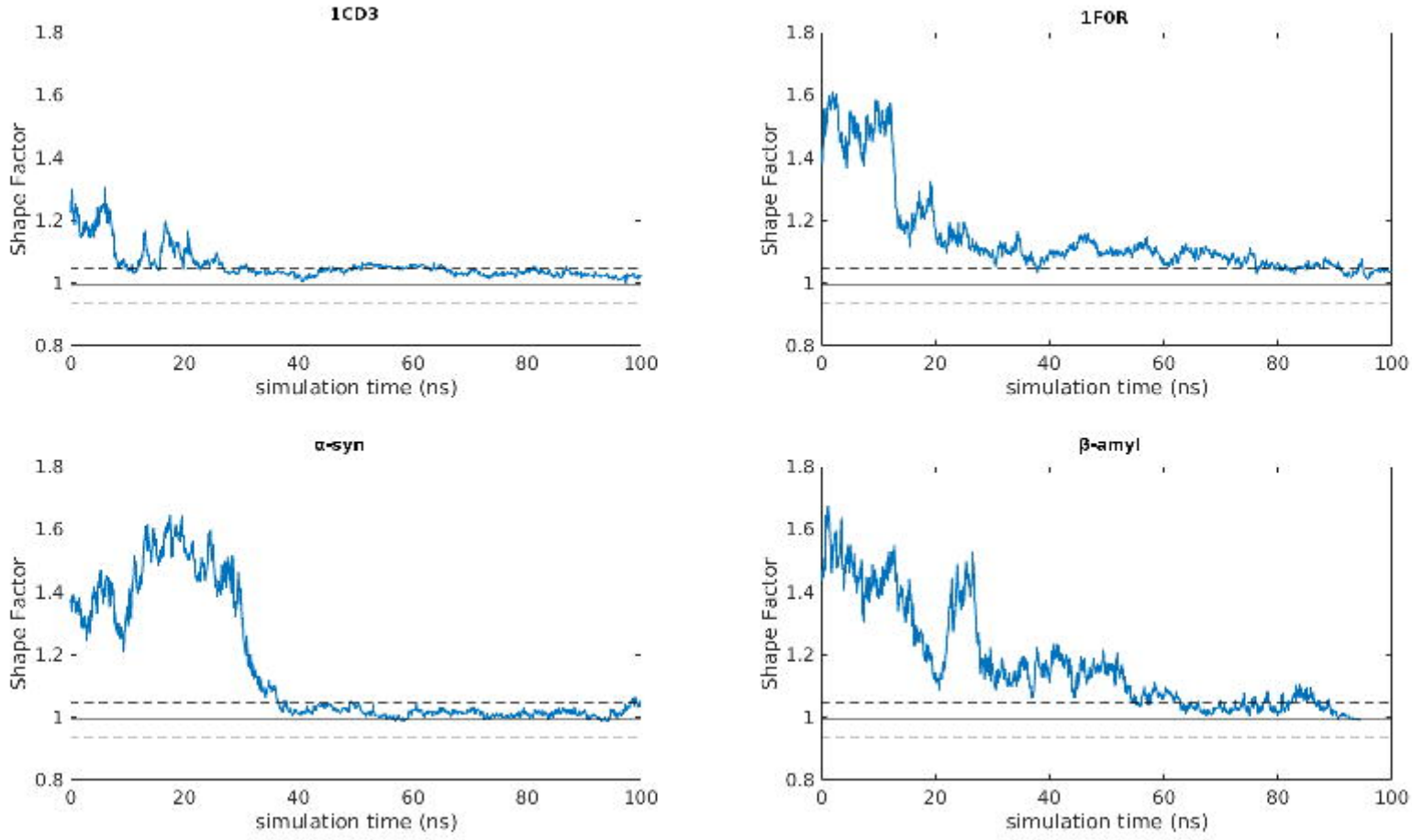
B.

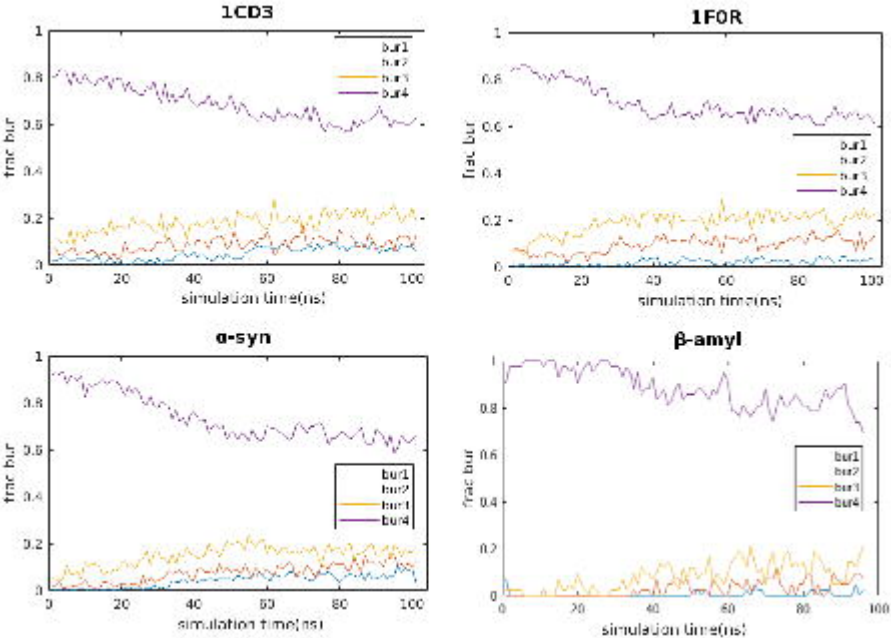

Number of Salt-bridges
4
3.5
3
2.5
2
1.5
1
0.5
0
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Persistence

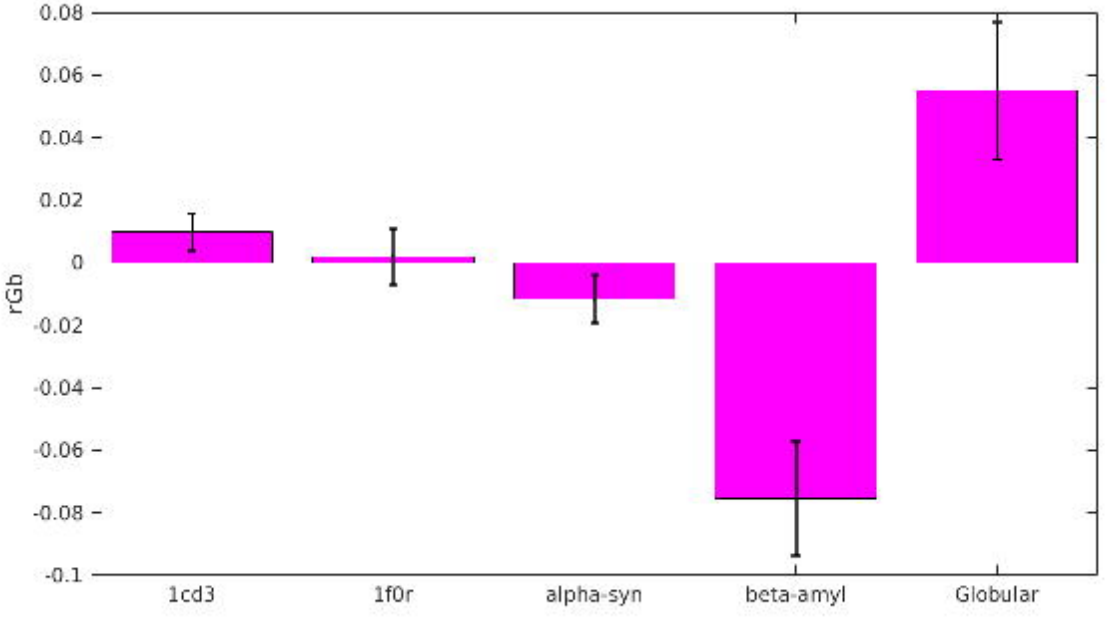

1-ASP
28-LYS
28-LYS
28-LYS
3-GLU
22-GLU
23-ASP
28-LYS

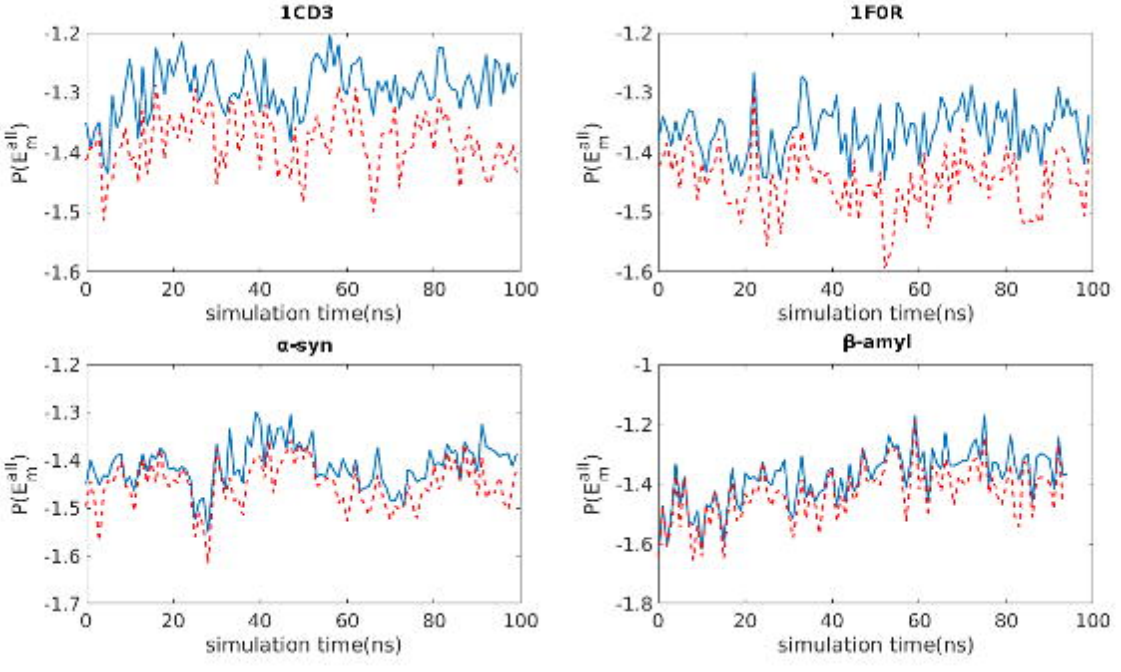

A.
108-ARG
105-GLU
2-GLU
28-GLU
77-ARG
27-ASP
22-GLU
15-GLU
50-ARG
61-ARG
64-ARG
68-GLU
65-ASP
100-ARG
98-ASP
B.
81-ARG
82-LYS
55-LYS
62-GLU
90-ASP
97-GLU
96-HIP
98-GLU
41-ASP
58-ASP
117-LYS
38-LYS
133-GLU
108-ARG
31-LYS
27-GLU
28-ASP
14-GLU
30-ASP
15-GLU
10-ARG
21-GLU
23-ARG
C.
58-LYS
57-GLU
20-GLU
21-LYS
96-LYS
28-GLU
34-LYS
23-LYS
104-GLU
D.
3-GLU
5-ARG
11-GLU
13-HIP

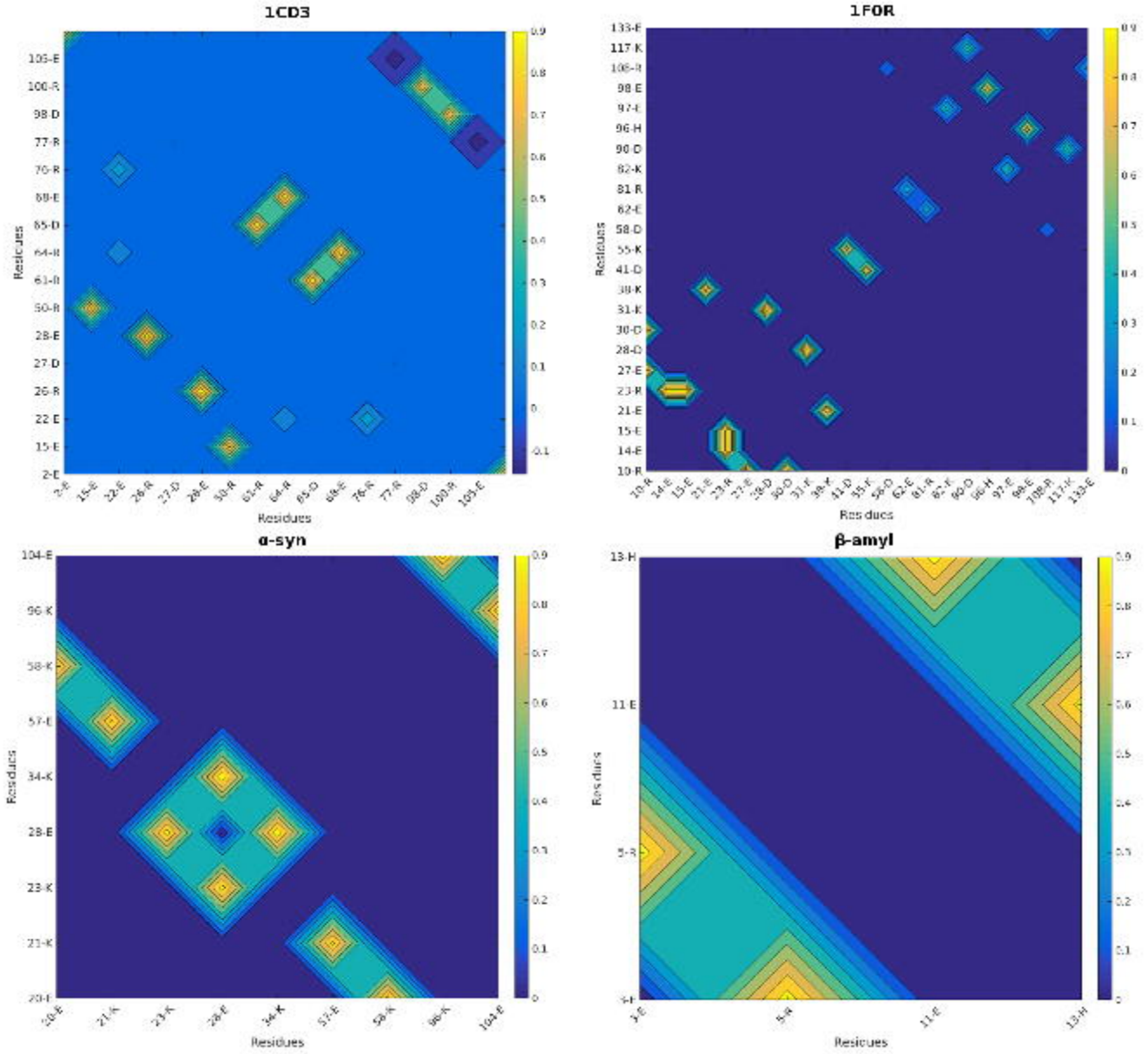

A.
B.


1CD3
1F0R
α-syn
β-amyl
Shape Factor
simulation time (ns)
1.8
1.6
1.4
1.2
1
0.8
0
20
40
60
80
100


1CD3
1FOR
α-syn
β-amyl
frac bur
simulation time(ns)
bur1
bur2
bur3
bur4


0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
rGb
1cd3
1f0r
alpha-syn
beta-amyl
Globular


1CD3
1FOR
α-syn
β-amyl
P(E_m^all)
simulation time(ns)
-1.2
-1.3
-1.4
-1.5
-1.6
-1.7
-1.8
-1
0
20
40
60
80
100


1CD3
Residues
Residues
1FOR
Residues
Residues
α-syn
Residues
Residues
β-amyl
Residues
Residues