Mechanism underlying conformational effects of the disease-associated Val66Met substitution on the intrinsically disordered region of proBDNF

Ruchi Lohia[1], Reza Salari[1], Grace Brannigan[1,2*],

1 Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA

2 Department of Physics, Rutgers University, Camden, NJ, USA

* grace.brannigan@rutgers.edu(GB)

Abstract

Although the role of electrostatic interactions and mutations that change charge states in intrinsically disordered proteins (IDPs) is well-established, many disease-associated mutations in IDPs are charge-neutral. The Val66Met single nucleotide polymorphism (SNP) encodes a hydrophobic-to-hydrophobic mutation at the midpoint of the prodomain of precursor brain-derived neurotrophic factor (BDNF), one of the earliest SNPs to be associated with neuropsychiatric disorders, for which the underlying molecular mechanism is unknown. Here we report on fully-atomistic temperature replica exchange molecular dynamics simulations of the 90 residue prodomain, for both the V66 and M66 sequence. In addition to secondary structure effects of the mutation, which we interpret in the context of previous data from NMR experiments, we provide residue-level insight into effects on transient tertiary structure, and further decompose the ensembles into several clusters. The Val66Met substitution is shown to replace the most populated cluster, a cluster with low radius of gyration containing a helical break at H65 which comprises about 40% of the V66 ensemble, with a cluster containing a helical break at I67 and a high radius of gyration. This shift significantly reduces transient tertiary contacts in the N-terminal third of the M66 sequence, due to a switch

from cooperation to competition between salt-bridging and hydrogen bonding interactions. We further observe that tertiary contacts are highly correlated to backbone configuration around residue 66 but relatively insensitive to whether the side-chain at residue 66 is Valine or Methionine.

Author Summary

Intrinsically disordered proteins are proteins that have no well-defined structure in at least one functional form. Mutations in one amino acid may still affect their function significantly, especially in subtle ways with cumulative adverse effects on health. Here we report on Molecular Dynamics Simulations of a protein that is critical for neuronal health throughout adulthood (Brain-derived Neurotrophic Factor), particularly the effects of a mutation carried by 25% of the American population, which has been widely studied for its association with aging-related and stress-related disorders, reduced volume of the hippocampus, and variations in episodic memory. We identify a molecular mechanism in which the mutation may change the global conformations of the protein and its ability to bind to receptors.

Introduction

The physiological significance of intrinsically disordered proteins (IDPs), which can explore a wide range of conformational ensembles in their functional form, [1–4] is now well-established. More than 33% of eukaryotic proteins contain disordered regions longer than 30 residues [5], many of which are involved in critical biological functions, including transcriptional regulation and cell signaling [6]. Long intrinsically disordered regions are particularly abundant among cancer and neurodegenerative-associated proteins [7, 8].

IDP amino-acid sequences tend to be low complexity and include numerous charged residues, often in long repeats [1]. In contrast to ordered proteins, in which a complex sequence encodes a well-defined tertiary structure, an IDP sequence determines a heterogeneous conformational ensemble. More than 35% of IDPs reported in DISPROT [9] are strong polyampholytes, and their ensemble properties can be predicted using statistical theories of polyampholytes from polymer physics and global properties of the

sequence, including the fraction of charged residues and the separation of oppositely charged residues (Fig 1a) [10, 11]. This role is consistent with the long-range nature of electrostatic interactions, which can affect coupling between distant residues in an otherwise disordered structure.

Fig 1. Electrostatics in the IDP diagram of states and proBDNF prodomain. a) Diagram of states reported by Ref. [10, 11], based on fraction of positively and negatively-charged residues. As indicated, the V/M66 BDNF prodomain lies on the boundary between the Janus region and the weak polyampholyte/polyelectrolyte regime. b) proBDNF consists of two domains: the prodomain and mature BDNF (mBDNF). c) Net charge per residue (NCPR) for the prodomain, based on a sliding window of 5 residues, showing a positively-charged N-terminal region, a highly negatively-charged mid-sequence region (containing the Val66Met SNP), and a negatively-charged C-terminal region. Parts a) and c) were output by CIDER. [43]

Although IDP sequences are low-complexity and do not encode a well-defined structure, single residue substitutions can still have functional effects that are significant for the organism. More than 20% of disease-associated missense single nucelotide polymorphisms (SNPs) are found in IDPs; [12] although detectable, the relatively subtle functional effects may lead to relatively weak selection pressure, whether positive or negative, allowing the mutation to persist at high frequencies within a population. Numerous structural and simulation studies [13–19] have demonstrated clear effects of single charged-residue insertion, deletion, or substitutions on conformational ensemble and aggregation of IDPs monomers. Single charged residue mutations or post tranlational modifications that change charges will affect the sequence electrostatics predicted to determine ensemble properties simply from statistical physics models, and in short-chains, can also induce qualitative changes by changing the appropriate regime. [10, 13, 20, 21]. Locally, such mutations can modulate residual secondary structure preferences via forming or breaking local salt-bridges or by introducing helix breaking residues. [14, 18, 22]

For IDPs with a relatively low fraction of charged residues, typical of the Janus region of the state diagram proposed by Das and Pappu [10, 11] (Fig 1a), more subtle differences among neutral amino-acids play an increasingly important role in determining the ensemble. More than 15% of disease-associated IDP polymorphisms are substitutions between two charge-neutral residues. [12] The extent to which such substitutions in IDPs can affect non-local aspects of the conformational ensemble is

uncertain; such substitution directly affects short-range interactions, and structure-based coupling between distant residues in IDPs is expected to be weak. Nonetheless, correlations between secondary structure of distant residues has been frequently observed in IDPs [14,23]; for example, several cancer mutations in transactivation domain of tumor suppressor p53 can lead to helicity changes in residues sequentially far away from the mutation sites [14].

In structured proteins, contacts between residues distant along the sequence are reflected in the tertiary structure, but developing a framework for describing the analogous property in IDPs has not been straightforward. Among traditional structural biology techniques, NMR has been most useful for characterizing IDPs, but is frequently limited to residual secondary structure (Ref. [7,24] and references therein). Molecular dynamics (MD) simulations have played a significant role in understanding IDP structure and dynamics [25–30], but face limitations on chain length similar to those incurred in simulations of protein folding; most unbiased simulations have been performed in implicit solvent and/or involve chains too short to meaningfully sample contacts between residues far apart on the peptide chain. Studies of aggregation among multiple shorter monomeric IDPs [31,32] have provided some of the most useful frameworks for considering tertiary contacts between residues which are distantly connected along the peptide backbone. Point mutations are also known to affect these contacts via differential salt-bridge and hydrogen-bonding formations, with mutations that change charge states affecting conformational ensemble via altered salt-bridge networks. [31]

Many SNPs in IDPs are associated with neurological, aging-associated neurodegenerative, or psychiatric disorders; despite an exponential increase in the amount of available genetic data, identifying the genetic origins of such disorders has proven remarkably challenging, with few variants identified as replicable predictors of disease. One of the earliest identified variants is the Val66Met SNP (rs6265) in the pro-domain region of Brain-derived Neurotrophic Factor (BDNF), [33] a signaling protein that retains a critical role in neurogenesis and synaptogenesis throughout adulthood(Fig 1b). [34] It has been implicated in maintenance of the hippocampus and the mechanism underlying action of numerous antidepressants, [35,36] including rapidly acting low-dose ketamine. [37] An extensive library of genome-wide association (and

even earlier) studies have repeatedly identified the Val66Met SNP as reducing hippocampal volume and episodic memory, as well as predicting increased susceptibility to neuropsychiatric disorders including schizophrenia, bipolar, and unipolar depression, but associations have been inconsistent and population dependent. [33, 37–40]

Difficulties in obtaining unambiguous disease associations at the proBDNF Val66Met SNP using GWAS are paralleled by challenges in characterizing its effects on the properties of the BDNF prodomain using structural techniques. A crystal structure of a homologous neurotrophic factor in complex with a shared receptor, revealed a well-defined volume corresponding to the prodomain, but which lacked resolvable density. [41]

It was subsequently revealed that the cleaved prodomains (~ 90 residues) are found in monomeric states *in vivo*, and the M66 (but not V66) form binds to SorCS2 (sortilin-related VPS10p domain containing receptor 2), leading to axonal growth cone retraction. [42] NMR measurements on the prodomain confirmed significant intrinsic disorder for both forms, with differential secondary structure preference around residue 66. [42]. It was not possible to gain any insight into the BDNF prodomain tertiary "structure", with uncertainty in interpretation of NMR signal obscuring whether secondary structure is affected far from the SNP, but additional NMR experiments implicated residue 66 in binding of M66 prodomain to SorCS2. [42]

In this work, we report on unbiased fully-atomistic replica-exchange MD simulations of the 90 residue BDNF prodomain in explicit solvent, for both V66 and M66 forms. This sequence falls at the boundary of the Janus and globular domains in the diagram proposed by Das and Pappu. [10, 11] We measure secondary structure tendencies consistent with NMR data, and show that differential helical tendencies are consistent with increased entropic cost of the valine side-chain in a helix conformation. We develop a tractable framework for interpreting transient tertiary contacts, allowing us to identify about 12 regular pairing sites along the BDNF prodomain sequence, which are stretches of the protein several residues long with an elevated tendency to form β bridges. Most significantly, we find that the backbone conformation local to residue 66 can predict such pairings. The charge-neutral Val66Met SNP is observed to affect the loose tertiary "structure" of BDNF prodomain by altering the local secondary structure of the adjacent highly charged region, which in turn affects long-range interactions between distant

residues.

Results and Discussion

Effects of Val66Met on secondary structure

Sensitivity of predictions to algorithm

d2D was used to reanalyze the chemical shift data from [42] as described in methods, and indicated significant PPII tendency in BDNF prodomain for both V66 and M66 forms (Fig 2a). Both TALOS+ and d2D indicated significantly higher β propensity in V66 at regions $0'$ and $+1'$ relative to M66 (Fig 2b). Unlike TALOS+ predictions, d2D also predicts higher β at region $+4'$ for V66. This may reflect the simplest scenario involving secondary structure coupling: reducing β propensity at one residue automatically reduces β propensity for any β bridging partners of that residue. (In the MD simulations, we do observe the $0'$ and/or $+1'$ pairing with $+4'$ in the V66 form, an interaction likely stabilized by the negative charge in the $+1'$ region interacting favorably with the positive charge of the $+4'$ region in the BXH cluster, which is likely to be even more populated at the colder temperature used for the NMR data.)

Fig 2. Secondary structure predictions from MD and NMR observables. a) Predictions based on NMR chemical shifts from [42] at 280K, reanalyzed using d2D as described in methods, with MD predictions colored according to replica temperature. The background of the plots are colored according to residue type: blue-basic, red-acidic, green-polar, white-hydrophobic. Putative "pairing regions" corresponding to peaks in β structure for V66 are marked as solid circles. b) Difference between V66 and M66 d2D secondary structure predictions according to NMR (black lines, at 280K) and MD data (blue lines, at 300K). c) Cartoon representation of prodomain sequence with residue type colored as in a and b, and d) β structure propensities from d2D and VMD for both V66 and M66.

As shown in Fig 2d, some conformations will be identified as indicating β structure by d2D but not by STRIDE; in general, these are conformations in which the backbone is consistent with a β conformation, but has no partnering strand. Conversely, for some residues STRIDE identified β structure while d2D did not; this is most striking at the neutral stretch of residues around $+4'$, where d2D predicts β structure from NMR chemical shifts and STRIDE predicts β structure from MD coordinates, but d2D does *n*ot identify significant β structure from the MD-generated chemical shifts.

Predictions from MD trajectories *vs* NMR data

Consistent with intrinsic disorder, helix and β propensity for each residue was low for both MD at 300K and NMR data at 280K (<30%). In general, agreement is very good, with discrepancies indicating slightly reduced secondary structure in the MD simulations, consistent with simple expectations based on the higher temperature used. Discrepancies are unlikely to simply reflect poor convergence of the MD data, since trends for different temperatures were smooth, and all replicas were able to diffuse in the temperature range 300K to 420K (S1 Fig).

Helix formation around 0′

The Val66Met mutation reversed the effect of increasing temperature on observed helicity at residues local to the mutation. For regions -1′,0′,+1′, helical tendencies for M66 and V66 increased and decreased, respectively, as temperature increased from 300K to 420K. This is consistent with an increased entropic cost of helical formation for the valine sidechain, which can access only one of three possible side-chain conformations. [63] For the BDNF prodomain sequence in particular, there is a potential steric clash between a side-chain substituent at the y-position and a nearby carbonyl oxygen, located on the preceding turn of the alpha-helix.

Several regions of both sequences (0′,-1′,+1′, +6′ and the region between +3′ and +4′) can form helices with at least 8 consecutive residues at 300K (Fig 3a). When compared with V66, M66 is 50% more likely to form consecutive helical fragments of 9 or more residues at 300K (Fig 3b). Extended helices at 0′, which is also the midpoint of the sequence, contrast the hinge-like midpoint associated with residual disorder at 0′

Fig 3. Temperature dependence of helical length around residue 66. a) Distribution of helix length for each residue. Regions 0′,-1′,+1′, +6′ and the region between +3′ and +4′ can form long stretch of helical fragments in both V66 and M66. M66 shows cooperative helix formation at residue 66. (b) Distribution of lengths for helices containing residue 66, for a range of temperatures.

For helices formed at residue 66 (0′), helix length is highly dependent on temperature regardless of amino-acid side-chain. While the formation of larger helices (characterized by 12 or more residues) increases at high temperature for M66, we observe an opposite trend in V66, with a higher propensity of forming smaller helices

(with eight or fewer residues) as temperature increases (Fig 3b,S2 Fig). This is also consistent with a reduced entropic cost of helix formation for the Methionine side-chain relative to the Valine side-chain. As shown in S2 Figb long helices are observed in multiple replicas for the M66 sequence.

Clustering based on three midpoint residues predicts long-range tertiary contacts

Widely used clustering methods (RSMD and dpCA clustering) yielded a prohibitively large number of clusters for the BDNF prodomain, due to its intrinsic disorder and the sequence length. Clustering on the basis of backbone conformation of the three neutral residues at the sequence midpoint was found to divide the ensemble into five meaningful clusters which could also predict tertiary contacts far from the midpoint. This clustering process is described further in Methods, and the three residues include residue V/M 66 as well as its immediate neighboring residues. The five clusters fell into two general categories: locally ordered (BBB, HHH), and locally disordered (BXH, XHB, HBB)(S4 Fig, S5 Fig).

Sequence-dependence of cluster population

Cluster populations, particularly locally disordered populations, were sensitive to the side-chain of residue 66 (Fig 4b). At 300K, the most populated clusters were locally disordered for both sequences, but had a reversed order of dihedral conformations; BXH (40%) and XHB (35%) dominated the V66 and M66 ensembles, respectively. For the purpose of clustering, the"B" conformation is broader than the usual β region of a Ramachandran plot, but it is incompatible with a helical backbone. A conformation assigned to cluster BXH thus requires a break in any helices at residue H65, while conformations assigned to cluster XHB require breaks at residue I67. We show below that the former predicts a particularly compact conformation with a well-defined set of tertiary contacts balanced between the N-terminal and C-terminal side of the prodomain, while the latter predicts an expanded conformation with an N-terminal side that has few backbone contacts within that side of the sequence, and a C-terminal side that has many.

Fig 4. Temperature dependence of cluster populations for V66 and M66. Conformations were assigned to clusters as described in methods.

Conventional contact maps provided little useful insight into tertiary structure for either forms of the protein, which was consistent with the intrinsic disorder and frequent transient interactions among neutral side-chains. Identification of contact residues via backbone hydrogen bonding or side chain salt-bridging yielded several persistent, weak long-range interactions at 300K for both sequences, which can be represented along a single axis (Fig 5). These diagrams for the entire ensemble indicate far more tertiary contacts between the sequence midpoint and the '+' (C-terminal) domain of the sequence for M66, relative to V66.

Fig 5. Linear networks of transient tertiary contacts a) Hydrogen-bonding and b) salt-bridging pairs are shown for the entire ensemble and then decomposed by cluster at low temperatures (300K-317K). The backbone tertiary-contact network is made for V66 and M66, with each residue serving as a node in the network, as described in Methods. Residues at the pairing regions are colored black. Backbone interactions serve as edges between individual network nodes; the thickness of the edge corresponds to the strength of the hydrogen bond, and the transparency of the edge increases as its frequency increases. If residue 66 or its nearby residues (A51-P79) are involved in hydrogen-bond formation then the edge is drawn above the node; otherwise, it is drawn at the bottom of the node.

The origin of this difference was determined using the clustering process outlined previously. Diagrams of each the ordered clusters (BBB and HHH) reveal similar patterns for V and M within each cluster; for these clusters, tertiary contacts are much more sensitive to the backbone conformation at residues 65-67 than to whether residue 66 has a side-chain of V or M. It is challenging to test the same claim for disordered clusters, because the disordered clusters are each sparsely populated in either V66 or M66 at 300K; e.g. XHB and BXH do not contribute significantly to the V66 and M66 ensembles respectively, but are the most populated cluster for M66 and V66 respectively.

Backbone conformation at protein midpoint predicts radius of gyration

The average radius of gyration R_g indicates an ensemble of more compact conformations for the V66 sequence at 300K; $R_g = 1.35 \pm .01$ nm and $1.39 \pm .01$ nm for V66 and M66 respectively. We observe a reversal of the temperature dependence in R_g with the Val66Met mutation; with increasing temperature, R_g increases for V66 and decreases for

M66 (Fig 6a, S3 Fig). The distribution of R_g indicated three main contributions; as described in Methods, the R_g curve for V66 and M66 was fit with a triple gaussian distribution with means of $\sim 1.30nm$ (μ_c-collapsed), $\sim 1.36nm$ (μ_i-intermediate) and $\sim 1.46nm$ (μ_e-expanded) (Fig 6b). At 300K, for V: $A_i \sim A_c >> A_e$ while for M: $A_i > A_c \sim A_e$ (Fig 6c).

Fig 6. Contributions of each cluster to ensemble of compact, intermediate, and expanded conformations. a) Radius of gyration distribution for entire ensemble is shown for a range of replica temperatures; with the M66 (but not V66) curve revealing at least one high radius of gyration conformation at lower temperatures. b) Distributions were fit with three gaussian curves of means $\sim 1.30nm$ (μ_c-collapsed), $\sim 1.36nm$ (μ_i-intermediate) and $\sim 1.46nm$ (μ_e-expanded). c) Fit parameters for amplitude of each Gaussian distribution at 300K, revealing a high contribution from the collapsed and expanded states in V66 and M66, respectively. d) Fit amplitudes corresponding to collapsed, expanded and intermediate structures for each cluster at 300K.

As shown in Fig 6d, S9 Fig, S10 Fig, the collapsed and expanded ensembles are dominated by conformations from cluster BXH and XHB, respectively. Cluster BXH is split comparably between collapsed and intermediate conformations, while cluster XHB contributes almost exclusively to the amplitude of the expanded conformation.

Tertiary networks in locally disordered clusters can be stabilized or destabilized by salt-bridges

Several residues near the mutation (residues E64,E68, and E69) are negatively charged (Fig 1c). Since the accessibility of these charged side-chains are dependent upon cluster, each cluster has a different salt bridging network.

Locally ordered clusters (HHH,BBB) have few persistent salt bridges; such conformations have reduced local degrees of freedom and a limited number of local residues that can simultaneously form salt-bridges. Among all clusters, the ordered cluster HHH has the least number of salt bridges and maximum number of hydrogen bonds per frame for both V66 and M66.

The disordered clusters BXH and XHB display a large number of salt bridges (Fig 5c) in similar networks, consistent with important role of electrostatics in maintaining protein disorder. As shown in Fig 7 and S11 Fig, in BXH conformations a high number of salt bridges is associated with a high number of hydrogen bonds, with

the most salt-bridges and hydrogen bonds leading to the most compact structures. In contrast, the hydrogen-bonding pattern of XHB conformations competes with the natural salt-bridging network, so that the number of hydrogen bonds is anti-correlated with the number of salt bridges. This is consistent with the previous characterization by Levine et *al* [13,31], which demonstrated that increased cooperation between salt bridges and hydrogen bonding resulted in more compact structures even in IDPs.

Fig 7. Effects on radius of gyration of switching from cooperation to competition between salt-bridging and hydrogen-bonding. Binned distributions of the total number of salt bridges, hydrogen-bonds and R_g for V66 and M66 at low temperatures (300K-317K) Probability densities of the total number of salt bridges per frame vs R_g or total number of hydrogen bonds per frame for locally-disordered cluster BXH (V66), replaced by cluster XHB for M66. Representative conformations are shown, colored by secondary structure, with residue 66 in orange.

Conclusion

We have carried out 78μs of fully-atomistic MD simulation of the 90 residue prodomain of brain-derived neurotrophic factor, with and without the disease-associated Val66Met mutation. Through extensive decomposition of the resulting ensembles, we find that the backbone configuration around residue 66 adopts several configurations at 300K, each of which is associated with a particular set of long-range contacts, summarized in Fig 8. Val66Met modifies the distribution of backbone configurations in this region and consequently also affects tertiary contacts, destabilizing a collapsed conformation in favor of an extended one.

Fig 8. Proposed mechanism for modulation of tertiary contacts via local conformation around residue 66. Blue circles and red circles indicate pairing regions in positively and negatively charged regions, respectively; residue 66 ($0'$) is indicated by a hollow black circle. Common long-range contacts are shown as solid lines, with possible sequence-dependent contacts shown as dashed lines (V66 in black, M66 in green). Orange symbols represent helical stretches (rectangles) or forced helical breaks (lines), corresponding to the associated cluster in the right column. Cluster XHB shifts the helical break in the C-terminal direction from its location in cluster BXH, allowing increased backbone pairing within the C-terminal side but decoupling it from the N-terminal side.

In our early efforts to cluster conformations, we intentionally avoided methods that relied too heavily on residue 66, since they seemed likely to artificially exaggerate the effects of the side-chain substitution on long-range interactions. Despite this concern,

clustering based on the peptide backbone at 65,66, and 67 was still eventually selected because it predicted long-range interactions across replicas within a single simulation much more robustly than numerous other methods (and residue combinations). This clustering method also yields cluster-tertiary contacts relationships that are surprisingly insensitive to side-chain at residue 66 (although cluster populations themselves are affected).

These three residues (65-66-67) possess several meaningful properties, beyond including the disease-associated mutation of our original interest. They are a) neutral residues inserted in a stretch of acidic residues constituting the most highly charged region of the protein, b) close to the interface ($\sim$ residue 60) between the positively charged N-terminal side of the sequence and the negatively charged C-terminal side of the sequence, and c) directly adjacent to the sequence midpoint at E68. We have not yet isolated which of these contribute to the critical role of residues 65-67 in determining tertiary interactions, and it is possible that overlap of (a) and (b) with (c) is intrinsic to the protein design.

The cluster containing the most compact conformations and balanced hierarchy of tertiary contacts, BXH, requires a helical break at residue 65, midway between the charge interface and the sequence midpoint. Shifting this break closer to the sequence midpoint and away from the charge interface (as in cluster XHB), decouples the N-terminal and C-terminal sides and results in an expanded structure in which the N-terminal side is largely "unfolded." (Fig 8)

One proposed mechanism underlying the effects of the Val66Met SNP involves M66 binding more stably to SorCS2. Anastasia et al [42] observed differential kinetics for interactions between BDNF prodomain and SorCS2, and the secondary structure from H65 to L71 changed upon interactions of M66 but not V66 with SorCS2. The present simulation results identify a compact semi-folded conformation (cluster BXH) which is highly populated for V66 at 300K but uncommon in the M66 ensemble. The analogous conformation in M66 is cluster XHB, which has a semi-folded C-terminal domain, but an extremely disordered positively-charged N-terminal domain with almost no long-range backbone contacts. Partial folding or other stabilization of this domain by its interaction with SorCS2 could account for stable M66 binding; it would not be expected for V66, which prefers the enthalpically favorable, more compact BXH cluster

even in the unbound state.

Materials and Methods

System setup

To account for differences in starting coil conformation, we included six unique structures to represent residues 23-113 of BDNF prodomain. All structures were built using I-Tasser [44–46], Robetta and Modeller [47], and all were simulated in a water box at 600K for 50 ns at a constant volume. From the six resulting trajectories, 10 structures with correct proline isomers were selected (based on at least 3ps time interval); in total, our study included 60 unique prodomain structures. All structures were cooled to 300K for 1ns, while prolines were restrained in trans-conformation. M66 replicas were generated by substituting Met for Val at residue 66. Each V66 and M66 replica was placed in a dodecahedron water box with 25,000 TIP3P [48] water molecules and a 0.15M salt concentration (NaCl) for a total system size of approximately 75,000 atoms. The same volume for each replica was ensured by fixing the simulation box of each replica to the average box size (10.2 nm).

Molecular Dynamics Simulation

All simulations used the amber99SB-ILDN force field [49] in the GROMACS 5.0.7 simulation package, [50, 51], with a time step of 2 fs. 60 replicas were used with temperatures ranging from 300-420K, with exponential spacing. Energy minimization for each replica was followed by NVT equilibration at 300K for 1 ns and NPT equilibration at 300K and 1atm pressure for 2ns.

Each replica was then simulated using T-REMD [52] with an exchange frequency of 1ps for 650 ns, giving a total simulation time of 78 μs with NVT ensemble. A different random seed was used for the Langevin dynamics of each replica. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method [53], with a 1 nm cutoff and a 0.12 nm grid spacing. Periodic boundary conditions were also used to reduce system size effects. Bonds with H-atoms were constrained using the LINCS (linear constraint solver) algorithm [54]. The time constant for temperature coupling is

1.5ps. The average exchange acceptance probability ranged between 0.13-0.23 for both V66 and M66. For both V66 and M66 groups, about 500 ns for each replica were discarded for equilibration purposes.

Analysis of MD Trajectories

For MD simulations, the secondary structure content was calculated with the STRIDE program incorporated in VMD, [55] which takes into account the combination of backbone dihedral angles and hydrogen bonding. Helix includes α-helix and 3_{10}-helix and β includes β-strand and β-bridge. The hydrogen bonds were calculated with |D-A|distance <= .35 nm and angle D-H-A angle <= 40°. For salt bridges, distance <= .32 nm was used as cutoff between the anionic and cationic atom. The radius of gyration was calculated using the all atoms.

Helix length calculation

The length of helix formed at each residue was calculated by determining the number of consecutive residues in which the dihedral angles satisfied ϕ <0°and -120°<ψ <50°, as in [56, 57].

Clustering

Dihedral angles of residues 65, 66 and 67 at each temperature were used to generate five cluster centroids; these exact values were calculated using K-means clustering algorithm in python. These centroids were used as initial values for k-means algorithm. The ensemble was divided into five clusters at every temperature for both V66 and M66. It was observed that for each cluster, ψ angles were localized within a region noted H (-120°<ψ <50°, enclosing, but not limited to, the $\alpha-$helix domain of a Ramachandran plot) or B (ψ >50°or ψ <-120°, enclosing, but not limited to, the β-sheet region of a Ramachandran plot); the five clusters have been named accordingly, with cluster HHH having residues 65,66, and 67 in the H region for $\sim$ 99% of its frames, cluster BXH having residue 65 in the B region, no constraints on residue 66, and residue 67 in the H region for $\sim$ 99% of the frames, etc.

Curve fitting for radius of gyration (R_g)

The R_g distribution curve was fit at each temperature with the sum of three Gaussian functions; these Gaussian functions had means of $\sim 1.30nm$ (collapsed), $\sim 1.36nm$ (intermediate) and $\sim 1.46nm$ (expanded).

$$y(x) = A_c.e^{\frac{-(x-\mu_c)^2}{2(\sigma_c)^2}} + A_i.e^{\frac{-(x-\mu_i)^2}{2(\sigma_i)^2}} + A_e.e^{\frac{-(x-\mu_e)^2}{2(\sigma_e)^2}} \quad (1)$$

where μ_c,μ_i,μ_e are the means and A_c,A_i,A_e are the amplitudes for collapsed, intermediate and expanded states respectively.

Tertiary contacts network

The contact networks were build using Cytoscape [58] with linear representation of residues. Each protein residue comprises a node in the network, with interactions between residues represented as edges. The strength of individual interactions can be interpreted by the thickness of the edge line on the network diagram. The transparency of an edge increases as it is found at more temperatures. If residue 66 or its neighboring residues (A51-P79) are involved in h-bond formation, its edge is drawn above the node; otherwise, the edge is drawn at the bottom of the node. To focus on significant interactions, interactions showing more than 3% persistence were considered in network visualization.

Interpretation of Chemical Shifts

Prior to the present study, Anastasia et al [42] measured chemical shifts for the BDNF prodomain (residues 21-113) using NMR, and then used backbone NMR chemical shifts to predict secondary structure via TALOS+ [59] and SSP [60]. While TALOS+ has been widely used for structured proteins, neither SSP or TALOS+ differentiates between beta and PPII propensity, which is a significant limitation for disordered proteins. For comparison with simulation data, we reinterpreted the chemical shifts directly from [42], deposited at Biological Magnetic Resonance Bank. Secondary structure was predicted using d2D [61], for both chemical shifts (CA,CB,CO,N,HN) generated from MD trajectories using SPARTA+ [62] and the NMR data (This calculation used did not use HA chemical shifts.)

Identification of Pairing Regions

The regions with peaks in β propensity in d2D predictions from NMR data are marked as "pairing regions" (Fig 2c), with region $0'$ located at the SNP (residue 66), negatively-notated regions located between the N-terminus and residue 66, and positively-notated regions located between residue 66 and the C-terminus.

Acknowledgments

The authors are grateful to Dr. Clay Bracken and Dr. Barbara Hempstead of Weill Cornell Medical Center for helpful discussions. Computational time was provided through XSEDE resources via NSF MCB110149.

References

1. Uversky VN. Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta. 2013;1834(5):932–51. doi:10.1016/j.bbapap.2012.12.008.

2. Panchenko AR, Babu MM. Editorial overview: Linking protein sequence and structural changes to function in the era of next-generation sequencing. Curr Opin Struct Biol. 2015;32:viii–x. doi:10.1016/j.sbi.2015.06.005.

3. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J Mol Biol. 2004;337(3):635–645. doi:10.1016/j.jmb.2004.02.002.

4. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005;6(3):197–208. doi:10.1038/nrm1589.

5. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J Mol Biol. 2004;337(3):635–645. doi:10.1016/j.jmb.2004.02.002.

6. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 2005;272(20):5129–5148. doi:10.1111/j.1742-4658.2005.04948.x.

7. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. Chem Rev. 2014;114(13):6561–6588. doi:10.1021/cr400514h.

8. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol. 2011;21(3):432–40. doi:10.1016/j.sbi.2011.03.011.

9. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the Database of Disordered Proteins. Nucleic Acids Res. 2007;35(Database):D786–D793. doi:10.1093/nar/gkl893.

10. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. 2015;32:102–112. doi:10.1016/j.sbi.2015.03.008.

11. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A. 2013;110(33):13392–7. doi:10.1073/pnas.1304749110.

12. Vacic V, Markwick PRL, Oldfield CJ, Zhao X, Haynes C, Uversky VN, et al. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. PLoS Comput Biol. 2012;8(10):e1002709. doi:10.1371/journal.pcbi.1002709.

13. Larini L, Gessel MM, LaPointe NE, Do TD, Bowers MT, Feinstein SC, et al. Initiation of assembly of tau(273-284) and its ΔK280 mutant: an experimental and computational study. Phys Chem Chem Phys. 2013;15(23):8916. doi:10.1039/c3cp00063j.

14. Ganguly D, Chen J, Dyson H, Wright P, Uversky V, Oldfield C, et al. Modulation of the Disordered Conformational Ensembles of the p53 Transactivation Domain by Cancer-Associated Mutations. PLOS Comput Biol. 2015;11(4):e1004247. doi:10.1371/journal.pcbi.1004247.

15. Viet MH, Nguyen PH, Derreumaux P, Li MS. Effect of the English Familial Disease Mutation (H6R) on the Monomers and Dimers of Aβ40 and Aβ42. ACS Chem Neurosci. 2014;5(8):646–657. doi:10.1021/cn500007j.

16. Viet MH, Nguyen PH, Ngo ST, Li MS, Derreumaux P. Effect of the Tottori Familial Disease Mutation (D7N) on the Monomers and Dimers of Aβ ₄₀ and Aβ ₄₂. ACS Chem Neurosci. 2013;4(11):1446–1457. doi:10.1021/cn400110d.

17. Truong PM, Viet MH, Nguyen PH, Hu CK, Li MS. Effect of Taiwan Mutation (D7H) on Structures of Amyloid-β Peptides: Replica Exchange Molecular Dynamics Study. J Phys Chem B. 2014;118(30):8972–8981. doi:10.1021/jp503652s.

18. Zhan YA, Wu H, Powell AT, Daughdrill GW, Ytreberg FM. Impact of the K24N mutation on the transactivation domain of p53 and its binding to murine double-minute clone 2. Proteins Struct Funct Bioinforma. 2013;81(10):1738–1747. doi:10.1002/prot.24310.

19. Xu L, Shan S, Wang X. Single Point Mutation Alters the Microstate Dynamics of Amyloid β-Protein Aβ42 as Revealed by Dihedral Dynamics Analyses. J Phys Chem B. 2013;117(20):6206–6216. doi:10.1021/jp403288b.

20. Bah A, Forman-Kay JD. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. J Biol Chem. 2016;291(13):6696–6705. doi:10.1074/jbc.R115.695056.

21. He Y, Chen Y, Mooney SM, Rajagopalan K, Bhargava A, Sacho E, et al. Phosphorylation-induced Conformational Ensemble Switching in an Intrinsically Disordered Cancer/Testis Antigen. J Biol Chem. 2015;290(41):25090–25102. doi:10.1074/jbc.M115.658583.

22. Alexander Conicella AE, Zerze GH, Mittal J, Fawzi Correspondence NL, Conicella AE, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Struct Des. 2016;24:1537–1549. doi:10.1016/j.str.2016.07.007.

23. Iešmantavičius V, Jensen MR, Ozenne V, Blackledge M, Poulsen FM, Kjaergaard M. Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered Protein Reveals Long-Range Helix–Helix Interactions. J Am Chem Soc. 2013;135(27):10155–10163. doi:10.1021/ja4045532.

24. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol. 2007;17(1):3–14. doi:10.1016/j.sbi.2007.01.009.

25. Stanley N, Esteban-Martín S, De Fabritiis G. Progress in studying intrinsically disordered proteins with atomistic simulations. Prog Biophys Mol Biol. 2015;119:47–52. doi:10.1016/j.pbiomolbio.2015.03.003.

26. Ithuralde RE, Roitberg AE, Turjanski AG. Structured and Unstructured Binding of an Intrinsically Disordered Protein as Revealed by Atomistic Simulations. J Am Chem Soc. 2016;138(28):8742–8751. doi:10.1021/jacs.6b02016.

27. Knott M, Best RB, Hummer G, de Bakker P, Word J. A Preformed Binding Interface in the Unbound Ensemble of an Intrinsically Disordered Protein: Evidence from Molecular Simulations. PLoS Comput Biol. 2012;8(7):e1002605. doi:10.1371/journal.pcbi.1002605.

28. Invernizzi G, Lambrughi M, Regonesi ME, Tortora P, Papaleo E. The conformational ensemble of the disordered and aggregation-protective 182-291 region of ataxin-3. Biochim Biophys Acta. 2013;1830(11):5236–47. doi:10.1016/j.bbagen.2013.07.007.

29. Abeln S, Frenkel D. Disordered flanks prevent peptide aggregation. PLoS Comput Biol. 2008;4(12):e1000241. doi:10.1371/journal.pcbi.1000241.

30. Yedvabny E, Nerenberg PS, So C, Head-Gordon T. Disordered Structural Ensembles of Vasopressin and Oxytocin and Their Mutants. J Phys Chem B. 2015;119(3):896–905. doi:10.1021/jp505902m.

31. Levine ZA, Larini L, LaPointe NE, Feinstein SC, Shea JE. Regulation and aggregation of intrinsically disordered peptides. Proc Natl Acad Sci U S A. 2015;112(9):2758–63. doi:10.1073/pnas.1418155112.

32. Pappu RV, Wang X, Vitalis A, Crick SL. A polymer physics perspective on driving forces and mechanisms for protein aggregation. Arch Biochem Biophys. 2008;469(1):132–41. doi:10.1016/j.abb.2007.08.033.

33. Notaras M, Hill R, van den Buuse M. The BDNF gene Val66Met polymorphism as a modifier of psychiatric disorder susceptibility: progress and controversy. Mol Psychiatry. 2015;20(8):916–930. doi:10.1038/mp.2015.27.

34. Korte M, Carroll P, Wolf E, Brem G, Thoenen H, Bonhoeffer T. Hippocampal long-term potentiation is impaired in mice lacking brain-derived neurotrophic factor. Proc Natl Acad Sci. 1995;92(19):8856–8860. doi:10.1073/pnas.92.19.8856.

35. Autry AE, Monteggia LM. Brain-derived neurotrophic factor and neuropsychiatric disorders. Pharmacol Rev. 2012;64(2):238–258. doi:10.1124/pr.111.005108.

36. Björkholm C, Monteggia LM. BDNF - a key transducer of antidepressant effects. Neuropharmacology. 2016;102:72–79. doi:10.1016/j.neuropharm.2015.10.034.

37. Autry AE, Adachi M, Nosyreva E, Na ES, Los MF, Cheng Pf, et al. NMDA receptor blockade at rest triggers rapid behavioural antidepressant responses. Nature. 2011;475(7354):91–95. doi:10.1038/nature10130.

38. Soliman F, Glatt CE, Bath KG, Levita L, Jones RM, Pattwell SS, et al. A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. Science. 2010;327(5967):863–6. doi:10.1126/science.1181886.

39. Chen ZY, Bath K, McEwen B, Hempstead B, Lee F. Impact of genetic variant BDNF (Val66Met) on brain structure and function. Novartis Found Symp. 2008;289:180–8; discussion 188–95.

40. Verhagen M, van der Meij A, van Deurzen PAM, Janzing JGE, Arias-Vásquez A, Buitelaar JK, et al. Meta-analysis of the BDNF Val66Met polymorphism in major depressive disorder: effects of gender and ethnicity. Mol Psychiatry. 2010;15(3):260–71. doi:10.1038/mp.2008.109.

41. Feng D, Kim T, Özkan E, Light M, Torkin R, Teng KK, et al. Molecular and Structural Insight into proNGF Engagement of p75NTR and Sortilin. J Mol Biol. 2010;396(4):967–984. doi:10.1016/j.jmb.2009.12.030.

42. Anastasia A, Deinhardt K, Chao MV, Will NE, Irmady K, Lee FS, et al. Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction. Nat Commun. 2013;4:2490. doi:10.1038/ncomms3490.

43. Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu RV. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J. 2017;112(1):16–21. doi:10.1016/j.bpj.2016.11.3200.

44. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2014;12(1):7–8. doi:10.1038/nmeth.3213.

45. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725–738. doi:10.1038/nprot.2010.5.

46. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9(1):40. doi:10.1186/1471-2105-9-40.

47. Šali A, Blundell TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol. 1993;234(3):779–815. doi:10.1006/jmbi.1993.1626.

48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926–935. doi:10.1063/1.445869.

49. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins Struct Funct Bioinforma. 2010; p. NA–NA. doi:10.1002/prot.22711.

50. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1-3):43–56. doi:10.1016/0010-4655(95)00042-E.

51. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2(June 2016):19–25. doi:10.1016/j.softx.2015.06.001.

52. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999;314(1-2):141–151. doi:10.1016/S0009-2614(99)01123-9.

53. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577–8593. doi:10.1063/1.470117.

54. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997;18(12):1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

55. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33–8, 27–8.

56. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. J Am Chem Soc. 2009;131(49):17908–17918. doi:10.1021/ja9069024.

57. Iglesias J, Sanchez-Martínez M, Crehuet R. SS-map. Intrinsically Disord Proteins. 2013;1(1):e25323. doi:10.4161/idp.25323.

58. Ahlstrom LS, Baker JL, Ehrlich K, Campbell ZT, Patel S, Vorontsov II, et al. Network visualization of conformational sampling during molecular dynamics simulation. J Mol Graph Model. 2013;46:140–9. doi:10.1016/j.jmgm.2013.10.003.

59. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009;44(4):213–23. doi:10.1007/s10858-009-9333-z.

60. Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. Protein Sci. 2006;15(12):2795–2804. doi:10.1110/ps.062465306.

61. Sormanni P, Camilloni C, Fariselli P, Vendruscolo M. The s2D Method: Simultaneous Sequence-Based Prediction of the Statistical Populations of

Ordered and Disordered Regions in Proteins. J Mol Biol. 2015;427(4):982–996. doi:10.1016/j.jmb.2014.12.007.

62. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR. 2010;48(1):13–22. doi:10.1007/s10858-010-9433-9.

63. Creamer TP, Rose GD. Side-chain entropy opposes a-helix formation but rationalizes experimentally determined helix-forming propensities (a-helix/protein folding/protein engineering). Nat Sci. 1992;89(250):5937–5941.

Supporting Information

S1 Fig. Mixing of replicas during the simulation. a) Mean square displacement (MSD) of each replica in the temperature range (300K-420K) for both V66 and M66. Replicas visiting 300K are colored green and remaining replicas are colored red. b) Population of each replica at 300K. c) Number of round trips completed by each replica.

S2 Fig. Temperature dependence of helical length around residue 66. a) Helical propensity at residues 50-77, determined from STRIDE, for helix of length 10,11 and 12 containing residue 66, with curves colored according to temperature. M66 folds into a 12 residue helix at residues 62-74 at high temperature. Representative conformations are shown, colored by secondary structure, with residue 66 in stick representation. b) Total number of replicas at which lengths for helices containing residue 66 is observed, for a range of temperatures. The helices of longer lengths are formed in 4 or more replicas at high temperatures.

S3 Fig. Backbone hydrogen bonding partners of residue 66 and length of beta sheet formed at every residue at 300K . a) Population of backbone hydrogen bonding of residue 66 with all other residues in the sequence. Residue 66 forms a weak contact with residue at $+4'$ region b) Length of beta strands formed at each residue. Pairing regions show higher density of beta of length 3 or more for both V66 and M66. The strand length formed at each residue was calculated by determining

the number of consecutive residues in which the dihedral angles satisfied ψ >50°and ϕ <-90°or ψ <-120°and ϕ <-90°.

S4 Fig. Distribution of dihedral angles at residue 65,66 and 67 for each cluster at 300K.

S5 Fig. Distribution of dihedral angles with temperature at residue 65,66 and 67 at each cluster. Dotted lines correspond to -120°<ψ <50°.

S6 Fig. Linear networks of transient tertiary contacts a) Hydrogen-bonding and b) salt-bridging pairs are shown for the entire ensemble and then decomposed by cluster at high temperatures (398K-420K). The backbone tertiary-contact network is made for V66 and M66, with each residue serving as a node in the network, as described in Methods. Residues at the pairing regions are colored black. Backbone interactions serve as edges between individual network nodes; the thickness of the edge corresponds to the strength of the hydrogen bond, and the transparency of the edge increases as its frequency increases. If residue 66 or its nearby residues (A51-P79) are involved in hydrogen-bond formation then the edge is drawn above the node; otherwise, it is drawn at the bottom of the node. Fewer contact pairs are observed at high temperatures. V66 forms stronger backbone hydrogen-bonding for beta bridge formations and M66 forms stronger backbone hydrogen-bonding for helix formation at residue 66. Cluster HHH in M66 simultaneously forms salt-bridge and hydrogen bonds from residues near the SNP.

S7 Fig. Secondary structure propensity, determined using STRIDE from MD trajectories, at each residue, with curves colored according to temperature.

S8 Fig. Length of helix formed at every cluster at 300K. Cluster HHH and Cluster XHB can form long cooperative helices at residue 66. A Smaller helix with 8 or fewer residues forms near residue 66 for clusters BXH,BBB and HBB.

S9 Fig. Distribution of R_g for each cluster. <R_g>with temperature for a) entire ensemble and b) each cluster. The <R_g>for V66 and M66 reverses with increase in temperature. We observe the same trend reversal for cluster HHH. c) Distribution of

R_g for each cluster at every temperature. The line color transitions from blue (cold) to red (hot) with increase in temperature from 300K to 420K.

S10 Fig. Distribution of amplitude of the fitted gaussian curves with temperature. Cluster BXH and XHB shows high amplitude of collapsed and expanded states, respectively at 300K. Cluster HHH shows increase in amplitude for collapsed structure at high temperature.

S11 Fig. Population densities of salt bridges and hydrogen bonds for V66 and M66 for each cluster at low temperatures (300K-317K). Simultaneous salt bridge and hydrogen bonds formation stabilize cluster BXH and only salt bridge formation destabilize cluster XHB. a) Population densities of total number of salt bridges per frame vs R_g (normalized with respect to total number of frames). b) Population densities of total number of hydrogen bonds per frame vs R_g (normalized with respect to total number of frames). c) Population densities of total number of salt bridge per frame vs total number of hydrogen bonds per frame (normalized with respect to total number of frames).